

Fourier Analysis of Numerical Algorithms for the Maxwell Equations*

YEN LIU

NASA Ames Research Center, Moffett Field, California 94035

Received October 25, 1993; revised September 11, 1995

The Fourier method is used to analyze the dispersive, dissipative, and isotropy errors of various spatial and time discretizations applied to the Maxwell equations on multi-dimensional grids. Both Cartesian grids and non-Cartesian grids based on hexagons and tetradecahedra are studied and compared. The numerical errors are quantitatively determined in terms of phase speed, wavenumber, propagation direction, grids spacings, and CFL number. The study shows that centered schemes are more efficient and accurate than upwind schemes and the non-Cartesian grids yield superior isotropy than the Cartesian ones. For the centered schemes, the staggered grids produce less errors than the unstaggered ones. A new unstaggered algorithm which has all the best properties is introduced. Using an optimization technique to determine the nodal weights, the new algorithm provides the highest accuracy among all the schemes discussed. The study also demonstrates that a proper choice of time discretization can reduce the overall numerical errors due to the spatial discretization. © 1996 Academic Press, Inc.

I. INTRODUCTION

The computations of electromagnetics have long been dominated by frequency-domain approaches. In particular, the method of moments (MoM) [1] has become a standard tool in determining the radar cross sections (RCS) for two-dimensional objects. The recent development of advanced stealth aircraft requires accurate and efficient simulations of electromagnetic waves through three-dimensional complex structures with layered material media over a wide frequency range (from 100 MHz to 20 GHz). These requirements cannot be easily obtained by the frequency-domain approaches. This has prompted the development of alternative approaches: direct time-domain solvers for the Maxwell curl equations.

The first time-domain technique in computational electromagnetics (CEM) was developed by Yee [2]. The method, generally referred as the finite-difference time-domain (FD-TD) method [3], is based on staggered central differencing in space and staggered leapfrog integration in time for Cartesian coordinates. Many authors have imple-

mented Yee's method and applied it to various problems [3–7]. In the past few years, this method has been generalized to include the modeling of curved surfaces [8–9] and body oriented grids [10–13]. It is currently one of the most successful techniques in CEM. More recently, the progress in interdisciplinary computational physics has created a new approach that uses numerical algorithms developed for solving the fluid flow equations in computational fluid dynamics (CFD) to solve the time-domain Maxwell equations. Shankar *et al.* were the first to introduce this type of approach and coined it a CFD-based method [14–15]. The method employs an upwind Riemann solver and the Lax–Wendroff time integration in a finite-volume formulation. It is also a successful technique in CEM.

Numerical approximations inevitably introduce errors. The errors associated with time-dependent wave equations involve dissipation, dispersion, and anisotropy. Dissipation causes the attenuation of wave amplitude and dispersion causes incorrect wave propagating speed, and these errors may depend on the direction of wave propagation with respect to the grid. The most troublesome aspect is that these errors are cumulative in nature. After propagating a fairly long distance or time, the solution can be greatly affected or destroyed and sometimes becomes non-physical. Being able to determine the magnitudes and understand the nature of these errors is an essential factor in developing numerical schemes for solving partial differential equations.

The Fourier method has been an indispensable tool in constructing analytical solutions for partial differential equations over the past two centuries. Today it has been further utilized in analyzing their numerical solutions [16–17]. In this paper, we present some systematic studies of various spatial and time discretizations applied to the Maxwell equations on multi-dimensional grids. The purposes of this paper are: (1) to quantitatively determine the dispersive, dissipative, and isotropy errors inherent in certain numerical algorithms; and (2) to examine and compare these errors on Cartesian and non-Cartesian grids. The non-Cartesian grids include grids based on hexagons for two dimensions and tetradecahedra for three dimensions.

* This paper was originally presented at the AIAA 31st Aerospace Sciences Meeting held in Reno, Nevada on Jan. 11–14, 1993.

Both staggered and unstaggered grids are considered. Individual errors introduced by the spatial and time discretizations are analyzed first and the combined errors then follow. Centered schemes, upwind schemes using a Riemann solver and intensity-vector splitting, as well as a new unstaggered scheme are studied and compared. Leapfrog and Runge–Kutta time integration methods are examined. Discussions are first given for the Cartesian grid and then the non-Cartesian grids. For some cases, comparisons will be made between the unstaggered and staggered grids. Analysis is given mainly for two dimensions. Extension to three dimensions is described briefly at the end of paper. Some numerical results are also included.

II. MAXWELL EQUATIONS

The time-domain Maxwell equations for non-conducting dielectrics can be written in a vector form as

$$\frac{\partial \mathbf{F}}{\partial t} + \nabla \times \mathbf{l} = 0, \quad (1)$$

where

$$\mathbf{F} = \begin{bmatrix} \mathbf{D} \\ \mathbf{B} \end{bmatrix} \quad (2a)$$

contains the electric displacement and magnetic induction vectors (sometimes referred as the electric and magnetic flux density vectors) and

$$\mathbf{l} = \begin{bmatrix} -\mathbf{H} \\ \mathbf{E} \end{bmatrix} \quad (2b)$$

contains the magnetic and electric intensity vectors. For simplicity, we assume linear isotropic constitutive relations

$$\begin{aligned} \mathbf{D} &= \varepsilon \mathbf{E} \\ \mathbf{B} &= \mu \mathbf{H}, \end{aligned} \quad (3)$$

where the permittivity ε and the permeability μ of the material are scalar constants. The two properties determine the wave phase speed $c = 1/\sqrt{\varepsilon\mu}$. Since the divergence conditions of the Maxwell equations,

$$\begin{aligned} \nabla \cdot \mathbf{D} &= 0 \\ \nabla \cdot \mathbf{B} &= 0, \end{aligned} \quad (4)$$

are satisfied automatically if they are imposed initially, one can just integrate the time-domain Maxwell equations (1)

to obtain instantaneous or time-history variations of the electric and magnetic fields.

One-dimensional equations are generally used in the Fourier analysis, in which one can examine only the dispersive and dissipative errors. In order to study the isotropy property of a numerical scheme, two- or three-dimensional equations must be used. In two dimensions, the Maxwell equations can be decomposed into two independent sets of equations, the transverse magnetic modes (TM) and the transverse electric modes (TE). The Fourier analysis yields the same results for both sets. In this paper we shall mainly use the two-dimensional TM set,

$$\begin{aligned} \frac{\partial D_z}{\partial t} &= \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} \\ \frac{\partial B_x}{\partial t} &= -\frac{\partial E_z}{\partial y} \\ \frac{\partial B_y}{\partial t} &= \frac{\partial E_z}{\partial x}, \end{aligned} \quad (5)$$

to examine the errors. The three-dimensional set will be used in Sections IV.2 and VIII.

Finite-difference, finite-volume, or finite-surface techniques can be utilized for the spatial discretization. Both finite-volume and finite-surface formulations are based on the integral form of Eq. (1). The former utilizes the Gauss curl theorem over grid cells converting the spatial term to a surface integral, while the latter applies Stokes's theorem over grid faces converting the spatial term to a line integral. In a uniform grid, these formulations can be identical to the finite-difference formulation, depending on how integrands are evaluated and what quadratures are used. We will use the finite-difference formulation for the most part of our discussions.

III. FOURIER ANALYSIS

It is well known that the exact solution of Maxwell's equations for constant properties consists of the superposition of linear, non-dispersive, and non-dissipative harmonic waves

$$\mathbf{F} = \begin{bmatrix} \mathbf{D}_0 \\ \mathbf{B}_0 \end{bmatrix} e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)}, \quad (6)$$

where \mathbf{D}_0 and \mathbf{B}_0 are the amplitudes, \mathbf{k} is the wavenumber, which has a magnitude κ and Cartesian components $(\kappa_x, \kappa_y, \kappa_z)$, and $\omega = \kappa c$ is the frequency. On the other hand, if one assumes periodic boundary conditions, the numerical solution can be decomposed into discrete Fourier series through a similarity transformation. In a matrix

representation, one writes the spatial discretization as the product of a circulant matrix and the vector of unknowns. The numerical solution depends on the properties of both the eigenvalues and eigenvectors of this circulant matrix. It is easy to show that each eigenvector corresponds to a *discrete* harmonic. However, the maximum and minimum points of a continuous harmonic and a discrete harmonic do not always coincide, due to finite resolution in the discrete space. As a result, amplitudes of eigenvectors do vary in general. At high frequencies, shapes of the eigenvectors appear like modulated harmonic waves. This difference between the discrete and continuous harmonics has not been widely treated in numerical analysis. We shall report this difference and its implication to the propagation of parasitic waves [17–18] elsewhere. Here, our purpose is to examine the properties of the numerical errors; it is therefore sufficient to write the eigenvectors as continuous harmonics and the numerical solution as

$$\mathbf{F} = \mathcal{F}(t)e^{i\mathbf{k}\cdot\mathbf{r}}. \quad (7)$$

For a given spatial discretization, the Fourier method reduces Eqs. (1) or (5) to a set of ordinary differential equations at each discretized location of the form

$$\frac{d\mathcal{F}}{dt} = \mathbf{G}_s\mathcal{F}. \quad (8)$$

Here \mathbf{G}_s is called the spatial amplification matrix and is a function of phase speed, wavenumber, and gridspacings. Equation (8) has a solution that varies as $e^{\lambda t}$, where λ are the eigenvalues of \mathbf{G}_s . Note that, from Eq. (6), the exact solution varies as $e^{-i\kappa c t}$. In general, $\lambda = \lambda_r + i\lambda_i$ is a complex number. Its real part determines the dissipative error of the spatial discretization, and its imaginary part determines the dispersive error. Let the superscript * denote the quantity obtained from the discretized equation. The numerical phase speed due to the spatial discretization is defined as

$$c^* \equiv -\frac{\lambda_i}{\kappa}. \quad (9)$$

For a harmonic wave, at a fixed point in space, the ratio of a field variable at two time levels separated by an interval Δt is $e^{-i\kappa c \Delta t}$. The argument $\phi = -\kappa c \Delta t$ is called the phase shift in the time interval Δt . Thus, for the exact time integration one finds

$$\phi^* = \frac{c^*}{c} \phi. \quad (10)$$

We say that the numerical solution experiences a phase lead or lag if c^*/c or ϕ^*/ϕ is greater or less than 1, respectively. The ratio c^*/c is called the normalized numerical phase speed and ϕ^*/ϕ is the normalized numerical phase shift. Since the exact solution is non-dissipative, the numerical dissipative error for the spatial discretization in the time interval Δt is simply $e^{\lambda \Delta t}$. For a uniform gridspacing Δs , the dissipation error can be rewritten as $(e^{\lambda \Delta s/c})^\nu$, where $\nu \equiv c \Delta t / \Delta s$ is the CFL number. The term inside the parentheses is called the normalized numerical amplitude.

Equation (8) can be diagonalized through a similarity transformation; one then obtains a scalar equation

$$\frac{df}{dt} = \lambda f, \quad (11)$$

where f is \mathcal{F} projected on the eigenspace. Equation (11) can be considered as a representative equation. For a given time discretization, the amplification factor σ , defined as the ratio of f at two adjacent time levels f^{n+1}/f^n , can be expressed as a function of $\lambda \Delta t$. Substituting the eigenvalues of \mathbf{G}_s into σ , one obtains the combined errors of spatial and time discretizations. Alternatively, the combined discretization can be expressed as

$$\mathcal{F}^{n+1} = \mathbf{G}\mathcal{F}^n, \quad (12)$$

and the combined errors are determined from the eigenvalues of the total amplification matrix \mathbf{G} . Note that for some schemes, such as the Lax–Wendroff scheme, which cannot be expressed in the form of Eq. (11), the first method is not applicable. The eigenvalues of \mathbf{G} are the amplification factor σ . The modulus of σ determines the dissipative error and stability, and the argument determines the combined phase shift

$$\phi^* \equiv \text{Arg}(\sigma). \quad (13)$$

IV. NUMERICAL ERRORS OF SPATIAL DISCRETIZATION

In this section the numerical errors of spatial discretizations on Cartesian and non-Cartesian grids are analyzed. The effects of data location on staggered or unstaggered grids and data management for colocated or uncolocated components are investigated. For the unstaggered grid, both electric and magnetic vector fields are placed on the same primary grid. On the other hand, for the staggered grid, they are placed on different grids, one on the primary grid and the other on the dual grid which is staggered with respect to the primary one. Since the data are vectors, there are two choices to define their components at each location of discretization on the staggered grid. In the first

one, designated as the collocated staggered grid, all components of one vector field are defined at the vertices of the primary grid and all components of the other vector field at the vertices of the dual grid [19]. In the second one, designated as the uncollocated staggered grid, the normal component of one vector field is defined at the primary face centers (edge centers in two dimensions) and the normal component of the other field at the dual face centers [2]. Our analysis begins with centered schemes. This is followed by an analysis of upwind schemes.

1. Centered Schemes

(a) Cartesian Grids

The simplest grid in numerical computations is the Cartesian grid. Its dual grid is obtained simply by shifting the primary grid a half of grid point in each direction. Shown in Fig. 1 are the three placements of unknowns in two dimensions, where the solid lines represent the primary grid and the dashed lines are the dual grid. For the unstaggered grid, the unknowns D_z , B_x , and B_y are all defined at points (j, k) ; for the collocated staggered grid, D_z are defined at points (j, k) and both B_x and B_y are defined at points $(j + \frac{1}{2}, k + \frac{1}{2})$; and for the uncollocated staggered grid, D_z are defined at points (j, k) and B_x and B_y are defined at points $(j, k + \frac{1}{2})$ and $(j + \frac{1}{2}, k)$, respectively. Except at the boundaries, all configurations contain the same number of unknowns.

Applying second-order central differences to Eq. (5), one obtains

$$\begin{aligned} \frac{dD_{z_j,k}}{dt} &= \frac{H_{y_{j+1,k}} - H_{y_{j-1,k}}}{2\Delta x} - \frac{H_{x_{j,k+1}} - H_{x_{j,k-1}}}{2\Delta y} \\ \frac{dB_{x_j,k}}{dt} &= -\frac{E_{z_{j,k+1}} - E_{z_{j,k-1}}}{2\Delta y} \\ \frac{dB_{y_j,k}}{dt} &= \frac{E_{z_{j+1,k}} - E_{z_{j-1,k}}}{2\Delta x} \end{aligned} \quad (14a)$$

for the unstaggered grid,

$$\begin{aligned} \frac{dD_{z_j,k}}{dt} &= \frac{H_{y_{j+1/2,k+1/2}} + H_{y_{j+1/2,k-1/2}} - H_{y_{j-1/2,k+1/2}} - H_{y_{j-1/2,k-1/2}}}{2\Delta x} \\ &\quad - \frac{H_{x_{j+1/2,k+1/2}} + H_{x_{j-1/2,k+1/2}} - H_{x_{j+1/2,k-1/2}} - H_{x_{j-1/2,k-1/2}}}{2\Delta y} \\ \frac{dB_{x_{j+1/2,k+1/2}}}{dt} &= -\frac{E_{z_{j+1,k+1}} + E_{z_{j,k+1}} - E_{z_{j+1,k}} - E_{z_{j,k}}}{2\Delta y} \\ \frac{dB_{y_{j+1/2,k+1/2}}}{dt} &= \frac{E_{z_{j+1,k+1}} + E_{z_{j+1,k}} - E_{z_{j,k+1}} - E_{z_{j,k}}}{2\Delta x} \end{aligned} \quad (14b)$$

for the collocated staggered grid, and

$$\begin{aligned} \frac{dD_{z_j,k}}{dt} &= \frac{H_{y_{j+1/2,k}} - H_{y_{j-1/2,k}}}{\Delta x} - \frac{H_{x_{j,k+1/2}} - H_{x_{j,k-1/2}}}{\Delta y} \\ \frac{dB_{x_{j,k+1/2}}}{dt} &= -\frac{E_{z_{j,k+1}} - E_{z_{j,k}}}{\Delta y} \\ \frac{dB_{x_{j+1/2,k}}}{dt} &= \frac{E_{z_{j+1,k}} - E_{z_{j,k}}}{\Delta x} \end{aligned} \quad (14c)$$

for the uncollocated staggered grid. Here Δx and Δy are the grid spacings. The schemes for the unstaggered grid (14a) and the uncollocated staggered grid (14c) involve the same number of operations, while the scheme for the collocated staggered grid (14b) requires about twice the number of operations. One should also notice that the unstaggered scheme (14a) divides the system into two independent sets of unknowns, which sometimes can lead to undesirable numerical oscillations. This is commonly referred to as the odd-even decoupling or the checkerboard

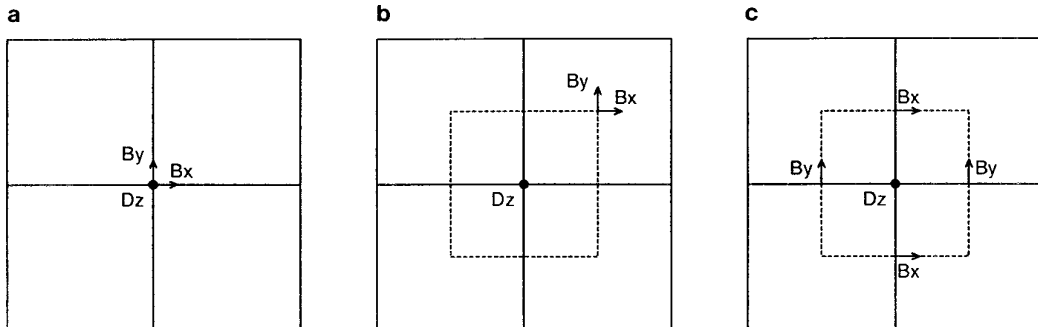


FIG. 1. Placement of unknowns on two-dimensional Cartesian grids: (a) unstaggered grid; (b) collocated staggered grid; (c) uncollocated staggered grid.

decoupling. There is no decoupling for the staggered schemes (14b) and (14c). The Fourier analysis can be performed by substituting Eq. (7) into Eqs. (14). The eigenvalues of each corresponding matrix \mathbf{G}_s are pure imaginary or zero, implying that all the schemes are non-dissipative, but dispersive. The normalized numerical phase speed can be easily obtained as

$$\frac{c^*}{c} = \begin{cases} \left[\frac{\sin^2 \xi}{\kappa^2 \Delta x^2} + \frac{\sin^2 \eta}{\kappa^2 \Delta y^2} \right]^{1/2} & \text{unstaggered;} \\ 2 \left[\frac{\cos^2(\eta/2) \sin^2(\xi/2)}{\kappa^2 \Delta x^2} + \frac{\cos^2(\xi/2) \sin^2(\eta/2)}{\kappa^2 \Delta y^2} \right]^{1/2} & \text{colocated staggered;} \\ 2 \left[\frac{\sin^2(\xi/2)}{\kappa^2 \Delta x^2} + \frac{\sin^2(\eta/2)}{\kappa^2 \Delta y^2} \right]^{1/2} & \text{unlococated staggered.} \end{cases} \quad (15)$$

Here the non-dimensional wavenumber components ξ and η are defined as $\kappa_x \Delta x$ and $\kappa_y \Delta y$, respectively. Let $\theta = \tan^{-1}(\kappa_x/\kappa_y)$ be the direction of wave propagation. One immediately sees that the numerical phase speed has a dependence on the direction of wave propagation and therefore is not isotropic. If one assumes a uniform grid-spacing, i.e., $\Delta x = \Delta y = \Delta s$, and defines the number of points per wavelength to be $N = 2\pi/\kappa \Delta s$, Eq. (15) can be rewritten as

$$\frac{c^*}{c} = \begin{cases} \frac{1}{2} \frac{N}{\pi} \left[\sin^2 \left(\frac{2\pi \cos \theta}{N} \right) + \sin^2 \left(\frac{2\pi \sin \theta}{N} \right) \right]^{1/2} & \text{unstaggered;} \\ \frac{N}{\pi} \left[\cos^2 \left(\frac{\pi \sin \theta}{N} \right) \sin^2 \left(\frac{\pi \cos \theta}{N} \right) + \cos^2 \left(\frac{\pi \cos \theta}{N} \right) \sin^2 \left(\frac{\pi \sin \theta}{N} \right) \right]^{1/2} & \text{colocated staggered;} \\ \frac{N}{\pi} \left[\sin^2 \left(\frac{\pi \cos \theta}{N} \right) + \sin^2 \left(\frac{\pi \sin \theta}{N} \right) \right]^{1/2} & \text{unlococated staggered.} \end{cases} \quad (16)$$

From the above equations, one notes that the unstaggered grid requires twice the number of points per wavelength in each direction (i.e., four times for two-dimensional computations) in order to have the same numerical phase speed as the uncolocated staggered grid. Shown in Fig. 2 are the polar diagrams of the normalized numerical phase speed

c^*/c for $N = 2, 4, 8$, and 16 for the three grids. A comparison of the three normalized numerical phase speeds for $N = 20$ is given in Fig. 3. The numerical phase speed in all grids experiences a phase lag. The lag decreases as N increases. In other words, for a given gridspacing Δs , the error $(1 - c^*/c)$ for the high frequency modes is greater than that for the low frequency modes. For a fixed N , the error on the colocated staggered grid interlaces between the errors on the other two grids. The two figures also demonstrate the anisotropy inherent in the discretizations, Eqs. (14). One observes that the error is greatest along the axes ($\theta = 0, \pi/2, \pi$, and $3\pi/2$) and least along the diagonals ($\theta = \pi/4, 3\pi/4, 5\pi/4$, and $7\pi/4$) for both the unstaggered and uncolocated staggered grids, and contrarily, the error is greatest along the diagonals and least along the axes for the colocated staggered grid. An important quantity to measure is the isotropy error, defined as the difference between the maximum and minimum values of the normalized numerical phase speed. Although a proper choice of time discretization, as will be discussed later, can improve the overall dispersive error, it is unable to reduce the anisotropy caused by the space discretizations. For $N = 20$, as shown in Fig. 3, the isotropy errors are 0.8% on the unstaggered grid, 0.4% on the colocated staggered grid, and 0.2% on the uncolocated staggered grid. In order to have the isotropy error less than 0.1%, one requires 58 points per wavelength for the unstaggered grid, 41 points for the colocated staggered grid, and 29 points for the uncolocated staggered grid, respectively. Expanding Eq. (16), one obtains the error

$$1 - \frac{c^*}{c} = \begin{cases} \left(\frac{1}{2} + \frac{1}{6} \cos 4\theta \right) \frac{\pi^2}{N^2} + \dots & \text{unstaggered;} \\ \left(\frac{1}{4} - \frac{1}{12} \cos 4\theta \right) \frac{\pi^2}{N^2} + \dots & \text{colocated staggered;} \\ \left(\frac{1}{8} + \frac{1}{24} \cos 4\theta \right) \frac{\pi^2}{N^2} + \dots & \text{unlococated staggered.} \end{cases} \quad (17)$$

The above equation shows that the leading dispersive errors are all inversely proportional to N^2 . Therefore, when N becomes relatively large, the error reduces to a quarter each time that the gridspacing is halved. It also shows that the leading dispersive error is a function of the wave propagation direction θ on each grid, with a period of $\pi/2$, and the maximum value being twice the minimum. Comparing the three discretizations, the leading maximum error and isotropy error on the uncolocated staggered grid are a quarter and a half of those on the unstaggered and colocated staggered grids, respectively. Taking also the

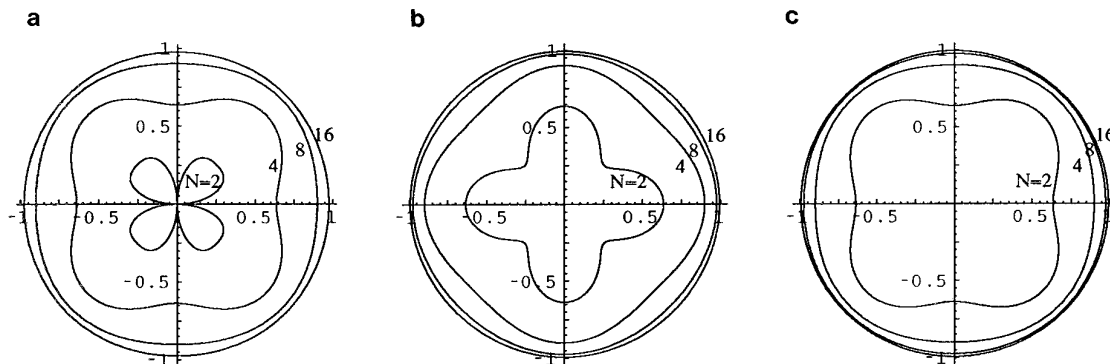


FIG. 2. Polar diagrams of the normalized numerical phase speed for Cartesian grids: (a) unstaggered grid; (b) colocated staggered grid; (c) uncolocated staggered grid.

total operations into account, one finds that the uncolocated staggered grid is four times more efficient than the other two grids.

(b) Hexagonal Grids

The major deficiency of conventional schemes, even for high-order ones, comes from their one-dimensional approach in which each spatial operator is approximated by employing data only along one coordinate line. Since only a 5-point Cartesian stencil is involved in each discretization, it is not surprising that all three schemes which we have just discussed exhibit some anisotropy. In order to reduce the isotropy error, data not located on the axes must also be utilized. In Ref. [15], a 9-point square stencil on an unstaggered Cartesian grid was used to study the errors for the scalar wave equation. While the leading error can be isotropic, the overall dispersive error may be increased. Here we present more efficient and accurate

alternatives which are based on 7-point stencils on regular hexagonal or triangular grids. The primary grid is composed of equilateral hexagons with edge of length Δs . Each hexagon can be considered as a union of six equilateral triangles. The dual of the triangles is also a regular hexagon. One can consider all three placements of the unknowns as discussed for the Cartesian grids. However, the colocated staggered grid will encounter a 7-point stencil for one field and a 4-point stencil for the other. Therefore, only the unstaggered grid and uncolocated staggered grid (referred to hereafter as the staggered grid) are considered. Figure 4a shows the unstaggered grid in which all unknowns are defined at the vertices of triangles. Figure 4b shows the staggered grid and its dual (dashed lines) in which D_z are defined at the face centers of dual hexagons (i.e., vertices of triangles) and three normal components of \mathbf{B} , B_1 , B_2 , and B_3 , are defined at the edge centers of primary triangles.

A general procedure for constructing high-order

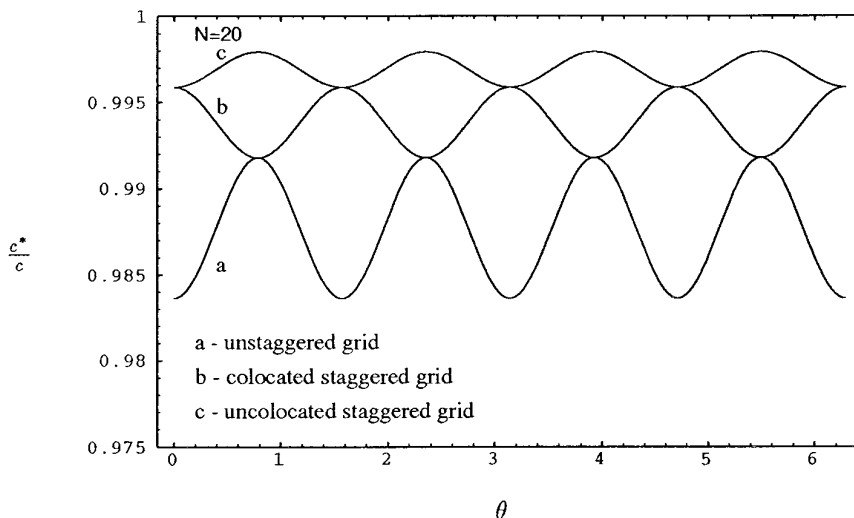


FIG. 3. Comparison of the normalized numerical phase speeds for Cartesian grids.

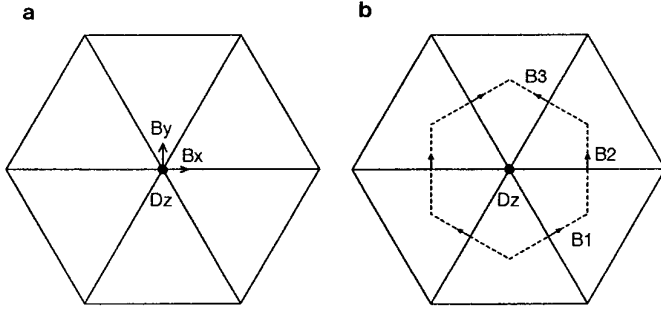


FIG. 4. Placement of unknowns on two-dimensional hexagonal grids: (a) unstaggered grid; (b) staggered grid.

schemes for arbitrary multi-dimensional grids is given in Ref. [20]. Second-order accurate centered schemes for the hexagonal grids can be obtained by using only the data at six nearest neighbors. One thus obtains

$$\frac{dD_{z_j,k}}{dt} = \frac{\begin{pmatrix} 2H_{y_{j+1,k}} - 2H_{y_{j-1,k}} + H_{y_{j+1/2,k+\sqrt{3}/2}} \\ -H_{y_{j-1/2,k+\sqrt{3}/2}} + H_{y_{j+1/2,k-\sqrt{3}/2}} - H_{y_{j-1/2,k-\sqrt{3}/2}} \end{pmatrix}}{6\Delta s} - \frac{\sqrt{3} \begin{pmatrix} H_{x_{j+1/2,k+\sqrt{3}/2}} - H_{x_{j+1/2,k-\sqrt{3}/2}} \\ + H_{x_{j-1/2,k+\sqrt{3}/2}} - H_{x_{j-1/2,k-\sqrt{3}/2}} \end{pmatrix}}{6\Delta s} \quad (18a)$$

$$\frac{dB_{x_j,k}}{dt} = -\frac{\sqrt{3} \begin{pmatrix} E_{z_{j+1/2,k+\sqrt{3}/2}} - E_{z_{j+1/2,k-\sqrt{3}/2}} \\ + E_{z_{j-1/2,k+\sqrt{3}/2}} - E_{z_{j-1/2,k-\sqrt{3}/2}} \end{pmatrix}}{6\Delta s}$$

$$\frac{dB_{y_j,k}}{dt} = \frac{\begin{pmatrix} 2E_{z_{j+1,k}} - 2E_{z_{j-1,k}} + E_{z_{j+1/2,k+\sqrt{3}/2}} \\ -E_{z_{j-1/2,k+\sqrt{3}/2}} + E_{z_{j+1/2,k-\sqrt{3}/2}} - E_{z_{j-1/2,k-\sqrt{3}/2}} \end{pmatrix}}{6\Delta s}$$

for the unstaggered grid and

$$\frac{dD_{z_j,k}}{dt} = \frac{2 \begin{pmatrix} H_{1_{j+1/4,k-\sqrt{3}/4}} - H_{1_{j-1/4,k+\sqrt{3}/4}} \\ + H_{2_{j+1/2,k}} - H_{2_{j-1/2,k}} \\ + H_{3_{j+1/4,k+\sqrt{3}/4}} - H_{3_{j-1/4,k-\sqrt{3}/4}} \end{pmatrix}}{3\Delta s}$$

$$\frac{dB_{1_{j+1/4,k-\sqrt{3}/4}}}{dt} = \frac{E_{z_{j+1/2,k-\sqrt{3}/2}} - E_{z_{j,k}}}{\Delta s}$$

$$\frac{dB_{2_{j+1/2,k}}}{dt} = \frac{E_{z_{j+1,k}} - E_{z_{j,k}}}{\Delta s}$$

$$\frac{dB_{3_{j+1/4,k+\sqrt{3}/4}}}{dt} = \frac{E_{z_{j+1/2,k+\sqrt{3}/2}} - E_{z_{j,k}}}{\Delta s} \quad (18b)$$

for the staggered grid. One notices that the total number of unknowns for the staggered grid are one-third more than those for the unstaggered grid, but the discretization is simpler and the total number of operations is only about half as large. Again, the eigenvalues of each corresponding matrix \mathbf{G}_s are pure imaginary or zero, implying both schemes are non-dissipative but dispersive. The normalized numerical phase speeds can be written as

$$\frac{c^*}{c} = \begin{cases} \frac{2}{3} \frac{1}{\kappa \Delta s} \left[3 \cos^2 \frac{\xi}{2} \sin^2 \frac{\sqrt{3}\eta}{2} + \left(\sin \xi + \sin \frac{\xi}{2} \cos \frac{\sqrt{3}\eta}{2} \right)^2 \right]^{1/2} & \text{unstaggered;} \\ \sqrt{8/3} \frac{1}{\kappa \Delta s} \left[\sin^2 \frac{\xi}{2} + \sin^2 \left(\frac{\xi}{4} - \frac{\sqrt{3}\eta}{4} \right) + \sin^2 \left(\frac{\xi}{4} + \frac{\sqrt{3}\eta}{4} \right) \right]^{1/2} & \text{staggered.} \end{cases} \quad (19)$$

Here ξ and η are defined as $\kappa_x \Delta s$ and $\kappa_y \Delta s$, respectively. Figure 5 presents the polar diagrams of the normalized numerical phase speeds for $N = 2, 4, 8$, and 16 for the two grids. The numerical phase speed also experiences a phase lag, but is almost isotropic. Expanding Eq. (19), the error can be expressed as

$$1 - \frac{c^*}{c} = \begin{cases} \frac{1}{2} \frac{\pi^2}{N^2} - \left(\frac{1}{12} + \frac{1}{120} \cos 6\theta \right) \frac{\pi^4}{N^4} + \dots & \text{unstaggered;} \\ \frac{1}{8} \frac{\pi^2}{N^2} - \left(\frac{7}{1152} + \frac{1}{720} \cos 6\theta \right) \frac{\pi^4}{N^4} + \dots & \text{staggered.} \end{cases} \quad (20)$$

The leading term is still inversely proportional to N^2 , but is now independent of the direction of propagation in both grids. The value of this term is exactly the average value of its Cartesian counterpart. Comparing the two discretizations, this value for the unstaggered grid is four times that for the staggered grid, also indicating that the former requires about twice the number of points per wavelength in order to have the same error as the latter. Taking also the total operations into account, the staggered grid is now eight times more efficient than the unstaggered grid. The anisotropy first appears in the fourth-order term, which is two orders of magnitude smaller than its Cartesian counterpart for both grids. While the periods are all $\pi/3$, the isotropy error for the unstaggered grid is six times as large as that for the staggered grid. This corresponds to 0.001% for the former and 0.00017% for the latter when $N = 20$, which are 800 $((20/\pi^2)N^2)$ and 1200 $((30/\pi^2)N^2)$ times smaller

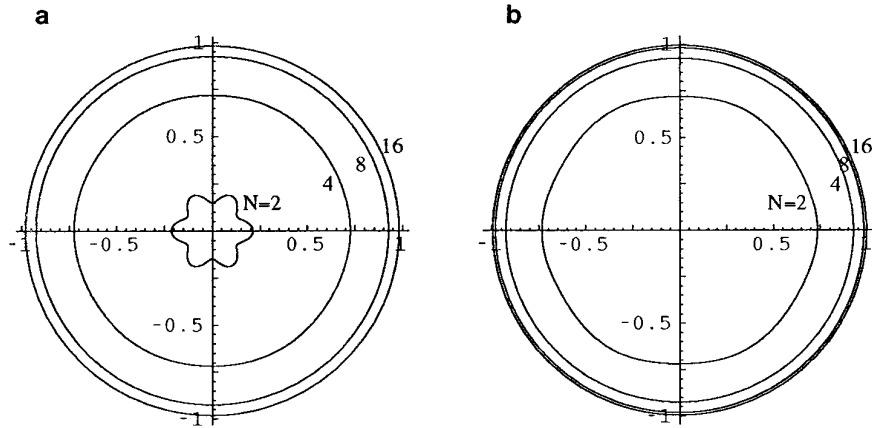


FIG. 5. Polar diagrams of the normalized numerical phase speed for hexagonal grids: (a) unstaggered grid; (b) staggered grid.

than their Cartesian counterparts, respectively. A comparison of the phase speeds for the two hexagonal grids and their Cartesian counterparts for $N = 20$ is given in Fig. 6. It clearly shows less dependence on the direction of propagation for the hexagonal grids and the superiority of the staggered grids.

2. Upwind Schemes

Upwind schemes have long been successfully used in CFD. They were designed to capture discontinuities, such as shocks and contact surfaces which are often found in fluid flows. There are basically two types of upwind algorithms in CFD, one associated with flux-vector splittings and the other with Riemann solvers. Similar algorithms have not been applied in solving the Maxwell equations

until recently when a Riemann solver was employed in a finite-volume procedure by Shankar *et al.* [14–15]. Extensions of these CFD-based algorithms to CEM are straightforward. Here we present some fundamental elements of these extensions. In general, both vector fields are required at locations of discretization for these upwind schemes. Therefore, numerical errors will be examined on an unstaggered grid, and for simplicity, only the regular Cartesian grid is considered. However, formulations are given in a coordinate-free, vector form.

(a) Intensity-Vector Splitting

Let \mathbf{n} be the unit normal vector in a positive direction to a coordinate surface in a finite-difference grid, or a cell

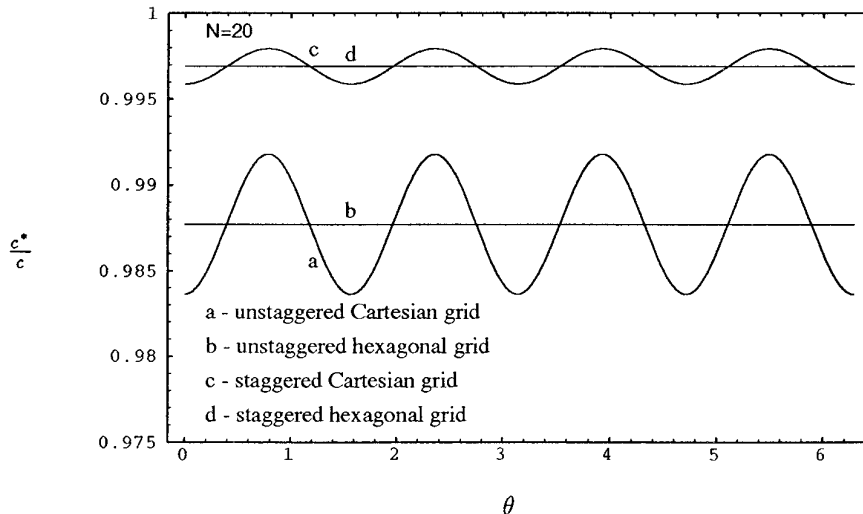


FIG. 6. Comparison of the normalized numerical phase speeds for Cartesian and hexagonal grids.

surface in a finite-volume surface. The spatial term in the time-domain Maxwell equations is associated with the tangential intensity vector

$$\mathbf{l}_t \equiv \mathbf{n} \times \mathbf{l}. \quad (21)$$

The Jacobian matrix, defined as

$$\mathbf{A} \equiv \frac{\partial \mathbf{l}_t}{\partial \mathbf{F}}, \quad (22)$$

plays a central role in the splitting algorithm. It has three distinct eigenvalues, 0, c , and $-c$, corresponding to characteristic speeds of a static field, a positively directed wave, and a negatively directed wave, respectively. According to the signs of the eigenvalues, the matrix \mathbf{A} can be split into two parts through a similarity transformation using its eigenvector matrix \mathbf{R}

$$\mathbf{A} = \mathbf{A}^+ + \mathbf{A}^- = \mathbf{R}(\Lambda^+ + \Lambda^-)\mathbf{R}^{-1}, \quad (23)$$

where Λ^+ and Λ^- are the diagonal eigenvalue matrices containing only the positive and negative eigenvalues, respectively. One can easily verify the homogeneity property $\mathbf{l}_t = \mathbf{A}\mathbf{F}$. Thus, the tangential intensity vector \mathbf{l}_t can also be split into two parts

$$\mathbf{l}_t = \mathbf{l}_t^+ + \mathbf{l}_t^-, \quad (24)$$

where

$$\mathbf{l}_t^+ \equiv \mathbf{A}^+\mathbf{F} = \frac{1}{2}\mathbf{n} \times \begin{bmatrix} -\mathbf{H} - \varepsilon\mathbf{c}\mathbf{n} \times \mathbf{E} \\ \mathbf{E} - \mu\mathbf{c}\mathbf{n} \times \mathbf{H} \end{bmatrix} \quad (25a)$$

and

$$\mathbf{l}_t^- \equiv \mathbf{A}^-\mathbf{F} = \frac{1}{2}\mathbf{n} \times \begin{bmatrix} -\mathbf{H} + \varepsilon\mathbf{c}\mathbf{n} \times \mathbf{E} \\ \mathbf{E} + \mu\mathbf{c}\mathbf{n} \times \mathbf{H} \end{bmatrix}. \quad (25b)$$

By expressing \mathbf{n} in terms of the unit vectors \mathbf{i} , \mathbf{j} , and \mathbf{k} , the split vectors on a Cartesian grid can be easily obtained.

While the scheme may be more useful at material interfaces where the fields can be discontinuous, here we just examine its properties in a single material for the purpose of comparison. To ensure stability, the positive part is generally discretized with backward differences (∇) and the negative part with forward differences (Δ). For two-

dimensional TM waves, Eq. (5) thus becomes

$$\begin{aligned} \frac{dD_z}{dt} &= \frac{1}{2}(\Delta_x + \nabla_x)H_y - \frac{1}{2}(\Delta_y + \nabla_y)H_x \\ &\quad + \frac{1}{2}\varepsilon c(\Delta_x - \nabla_x + \Delta_y - \nabla_y)E_z \\ \frac{dB_x}{dt} &= -\frac{1}{2}(\Delta_y + \nabla_y)E_z \\ &\quad + \frac{1}{2}\mu c(\Delta_y - \nabla_y)H_x \\ \frac{dB_y}{dt} &= \frac{1}{2}(\Delta_x + \nabla_x)E_z \\ &\quad + \frac{1}{2}\mu c(\Delta_x - \nabla_x)H_y. \end{aligned} \quad (26)$$

The above intensity-vector splitting scheme can be considered as the central scheme plus an artificial damping (or smoothing) which corresponds to the last term in each of the above equations. If the first-order, two-point one-sided differences are used, the above scheme is first-order accurate and the damping term is equivalent to a second derivative. Similarly, if the second-order, three-point one-sided differences are used, the scheme is second-order accurate and the damping term is equivalent to a fourth derivative. Fourier analysis leads to

$$\mathbf{G}_s = \begin{bmatrix} c(c_x + c_y) & -\frac{1}{\mu}s_y & \frac{1}{\mu}s_x \\ -\frac{1}{\varepsilon}s_y & cc_y & 0 \\ \frac{1}{\varepsilon}s_x & 0 & cc_x \end{bmatrix}, \quad (27)$$

where

$$\begin{aligned} c_x &= \frac{-1 + \cos \xi}{\Delta x}, & s_x &= i \frac{\sin \xi}{\Delta x}, \\ c_y &= \frac{-1 + \cos \eta}{\Delta y}, & s_y &= i \frac{\sin \eta}{\Delta y}, \end{aligned} \quad (28a)$$

for the first-order scheme and

$$\begin{aligned} c_x &= \frac{-3 + 4 \cos \xi - \cos 2\xi}{2\Delta x}, & s_x &= i \frac{4 \sin \xi - \sin 2\xi}{2\Delta x}, \\ c_y &= \frac{-3 + 4 \cos \eta - \cos 2\eta}{2\Delta y}, & s_y &= i \frac{4 \sin \eta - \sin 2\eta}{2\Delta y}, \end{aligned} \quad (28b)$$

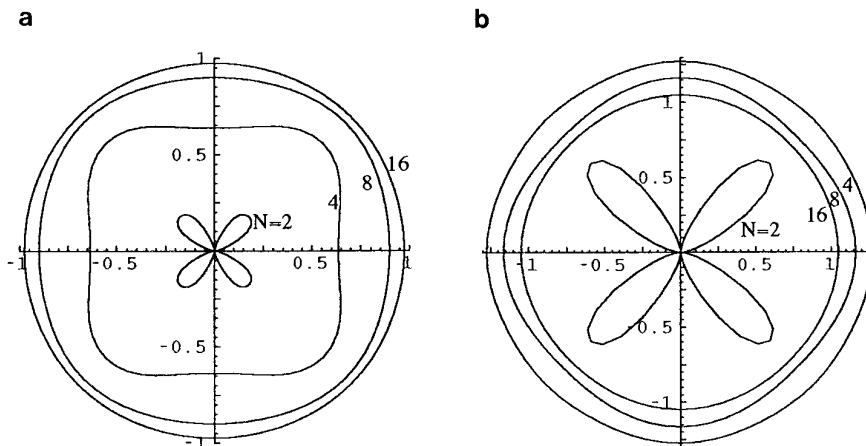


FIG. 7. Polar diagrams of the normalized numerical phase speed for the upwind schemes: (a) first order; (b) second order.

for the second-order scheme. Two eigenvalues of \mathbf{G}_s , which correspond to the $\pm c$ directed waves, are complex, implying the scheme is both dispersive and dissipative. The third eigenvalue is negative, indicating the static field will decay. Figures 7 and 8 show the normalized numerical phase speeds and amplitudes for the scheme on an equally spaced, unstaggered Cartesian grid. For the first-order scheme, the numerical phase speed experiences a phase lag and the phase error is comparable to that of the unstaggered centered scheme, but the scheme is extremely dissipative. By contrast, for the second-order scheme, the phase speed experiences a phase lead except when $N = 2$, and the phase error is relatively larger and the amplitude error is relatively smaller than for the first-order scheme.

(b) Riemann Solver

The Riemann solver is based on the exact solution of the one-dimensional Maxwell equations. Consider a plane

wave propagating in the direction of \mathbf{n} , and let ζ be the distance from the origin to a given plane. Equations (1) thus become

$$\frac{\partial \mathbf{F}}{\partial t} + \mathbf{n} \times \frac{\partial \mathbf{I}}{\partial \zeta} = 0. \tag{29}$$

Multiplying by \mathbf{R}^{-1} , the above equations can be written in the decoupled characteristic form and one finds the Riemann invariants

$$\begin{aligned} \mathbf{w}^+ &= \mathbf{n} \times (\mathbf{H} + \varepsilon c \mathbf{n} \times \mathbf{E}) \\ \mathbf{w}^- &= \mathbf{n} \times (\mathbf{H} - \varepsilon c \mathbf{n} \times \mathbf{E}) \end{aligned} \tag{30}$$

are constants along the $+c$ and $-c$ characteristics, respectively. Let the symbol $\bar{}$ denote the field state at a material interface and the subscripts L and R denote the field states

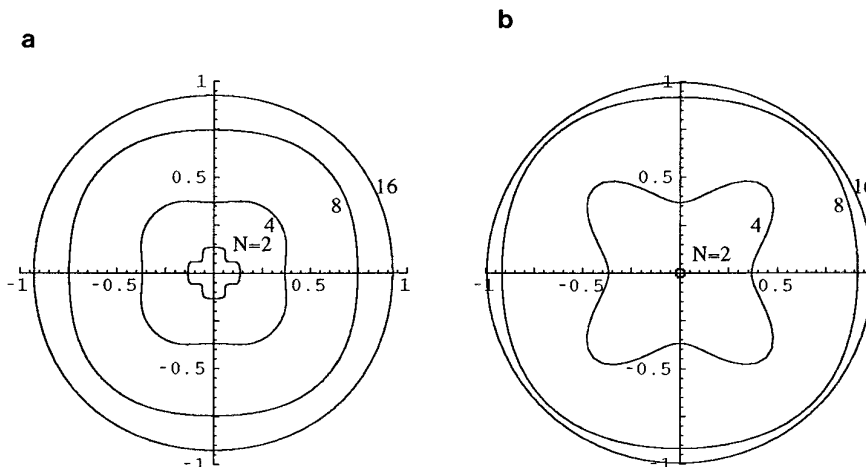


FIG. 8. Polar diagrams of the normalized numerical amplitude for the upwind schemes: (a) first order; (b) second order.

on the left and right of the interface, respectively. At the material interface, the Maxwell equations require the tangential components of the electric and magnetic fields be continuous; the intensity vector $\bar{\mathbf{I}}_t$ can then be derived as

$$\begin{aligned} \bar{\mathbf{I}}_t &= \mathbf{n} \times \begin{bmatrix} -\bar{\mathbf{H}} \\ \bar{\mathbf{E}} \end{bmatrix} \\ &= \mathbf{n} \times \begin{bmatrix} \frac{(\mu c)_R \mathbf{H}_R + (\mu c)_L \mathbf{H}_L - \mathbf{n} \times (\mathbf{E}_R - \mathbf{E}_L)}{(\mu c)_L + (\mu c)_R} \\ \frac{(\varepsilon c)_R \mathbf{E}_R + (\varepsilon c)_L \mathbf{E}_L + \mathbf{n} \times (\mathbf{H}_R - \mathbf{H}_L)}{(\varepsilon c)_L + (\varepsilon c)_R} \end{bmatrix}. \end{aligned} \quad (31)$$

The left and right states are evaluated by extrapolation and interpolation from neighboring states. The simplest ones are the zeroth-order extrapolation

$$\begin{aligned} \mathbf{I}_{Rj+1/2} &= \mathbf{I}_{j+1}, \\ \mathbf{I}_{Lj+1/2} &= \mathbf{I}_j \end{aligned} \quad (32a)$$

and the linear extrapolation

$$\begin{aligned} \mathbf{I}_{Rj+1/2} &= \frac{1}{2}(3\mathbf{I}_{j+1} - \mathbf{I}_{j+2}), \\ \mathbf{I}_{Lj+1/2} &= \frac{1}{2}(3\mathbf{I}_j - \mathbf{I}_{j-1}). \end{aligned} \quad (32b)$$

In a material of constant properties, the finite-volume formulation using above the Riemann solver (31) is identical to the intensity-vector splitting scheme. The formulation is first order if (32a) is used and second order if (32b) is used. We will not repeat the Fourier analysis here. Inspecting Eqs. (25) and (31), we see that the two upwind schemes require significantly more operations than the centered schemes.

V. NUMERICAL ERRORS OF TIME DISCRETIZATION

Time discretization, like spatial discretization, can also introduce errors. However, these errors are only associated with dispersion and dissipation, and they are isotropic. In this section, we briefly review several explicit time differencing schemes including the staggered leapfrog (LF1), unstaggered leapfrog (LF2), and third- and fourth-order Runge–Kutta methods (RK3, RK4). Both leapfrog methods are second order. In the staggered one the electric and magnetic fields are staggered in time by an interval $\Delta t/2$, while they are defined at the same time level for the unstaggered one. For a function f , the former can simply be represented by

$$f^{n+1} = f^n + \Delta t f'^{n+1/2} \quad (33)$$

and the latter by

$$f^{n+1} = f^{n-1} + 2\Delta t f'^n, \quad (34)$$

where the symbol ' denotes the first time derivative. (Equation (33) is no longer valid if the equation contains any source or damping terms. One should then use the matrix form representation.) The third- and fourth-order Runge–Kutta methods are well known and can be written as

$$\begin{aligned} f^{n+1/3} &= f^n + \frac{1}{3}\Delta t f'^n \\ f^{n+2/3} &= f^n + \frac{2}{3}\Delta t f'^{n+1/3} \\ f^{n+1} &= f^n + \frac{1}{4}\Delta t (3f'^{n+2/3} + f'^n) \end{aligned} \quad (35)$$

and

$$\begin{aligned} \tilde{f}^{n+1/2} &= f^n + \frac{1}{2}\Delta t f'^n \\ \hat{f}^{n+1/2} &= f^n + \frac{1}{2}\Delta t \tilde{f}'^{n+1/2} \\ \bar{f}^{n+1} &= f^n + \Delta t \hat{f}'^{n+1/2} \\ f^{n+1} &= f^n + \frac{1}{6}\Delta t (f'^n + 2\tilde{f}'^{n+1/2} + 2\hat{f}'^{n+1/2} + \bar{f}'^{n+1}), \end{aligned} \quad (36)$$

respectively. Using Eq. (11) as the representative equation, one can determine the amplification factor

$$\sigma = \begin{cases} 1 + \frac{1}{2}(\lambda\Delta t)^2 \pm \sqrt{\left[1 + \frac{1}{2}(\lambda\Delta t)^2\right]^2 - 1} & \text{staggered leapfrog LF1;} \\ \lambda\Delta t \pm \sqrt{(\lambda\Delta t)^2 + 1} & \text{unstaggered leapfrog LF2;} \\ 1 + \lambda\Delta t + \frac{1}{2}(\lambda\Delta t)^2 + \frac{1}{6}(\lambda\Delta t)^3 & \text{third-order Runge–Kutta RK3;} \\ 1 + \lambda\Delta t + \frac{1}{2}(\lambda\Delta t)^2 + \frac{1}{6}(\lambda\Delta t)^3 + \frac{1}{24}(\lambda\Delta t)^4 & \text{fourth-order Runge–Kutta RK4.} \end{cases} \quad (37)$$

The stability bound for each method is well known and can be easily determined from the above equation. All methods are dispersive. For a pure imaginary λ , the leapfrog methods are not dissipative, but the Runge–Kutta methods are. The normalized numerical phase shifts due to time discretization alone for the four methods are plotted in Fig. 9. Notice that LF1, LF2, and RK3 have a phase lead within their stability ranges, and RK4 has a phase lag for $-\lambda_i\Delta t < 1.88$ and a phase lead otherwise. As a final note, if currents or any physical or artificial damping terms

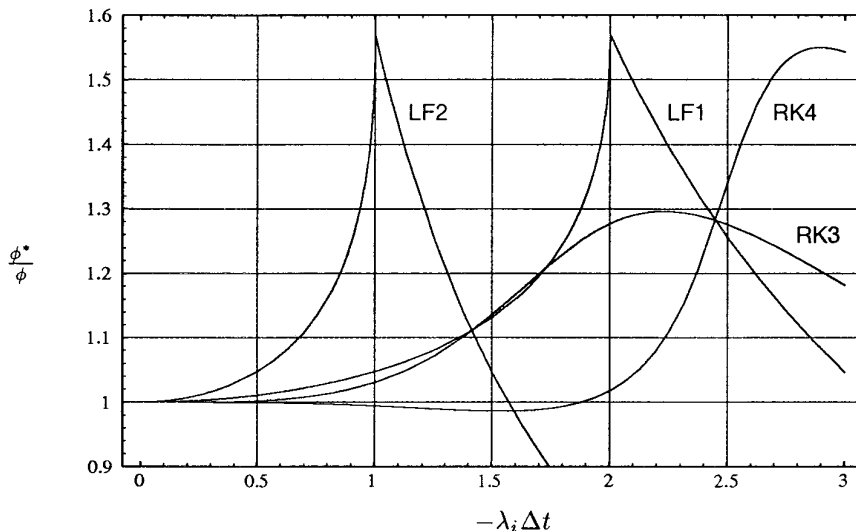


FIG. 9. Normalized numerical phase shift of time differencing schemes.

are present, LF2 is unconditionally unstable and LF1, RK3, and RK4 are conditionally stable.

VI. NUMERICAL ERRORS OF SPATIAL AND TIME DISCRETIZATION

We have discussed many spatial and time schemes individually in the previous two sections. In this section, we will investigate the numerical errors from the combined discretizations. These errors can be determined from the eigenvalues of the amplification matrix \mathbf{G} directly as described in Section III. For some cases, the procedure can be simplified by substituting the eigenvalues of \mathbf{G}_s given in Eqs. (15), (19), or (27) into Eq. (37). Since the procedure is straightforward, we will analyze the results for a few representative cases.

When combining the two discretizations, it is important to realize that the errors due to one can partially cancel the errors introduced by the other. One might think that the fourth-order Runge–Kutta method is more accurate than the second-order leapfrog method. It is true for time discretization alone, but not when used with spatially centered schemes. This is illustrated in Fig. 10, in which the uncolocated staggered scheme (14c) with $N = 10$ is used. The middle curve labeled “Exact” is the normalized phase shift using the exact time integration. The difference between the exact curve and the value 1 is the spatial phase error, which has a phase lag. The top and bottom curves are also the normalized phase shift but using the staggered leapfrog and fourth-order Runge–Kutta methods with $\nu = \sqrt{2}/2$, respectively. The difference between each curve and the exact curve is an additional phase error due to the time discretization. Although this additional phase error

is small for RK4, it actually moves the exact curve further away from 1 because it also experiences a phase lag. LF1 has a larger error but experiences a phase lead, and therefore it moves the exact curve toward 1 and makes the total error less than that for RK4.

It can be shown that the staggered leapfrog in time, like the staggered central difference in space, is two times more efficient than the unstaggered one. Perhaps the best choice on regular grids for second-order schemes is the combination of uncolocated staggered centered schemes with the staggered leapfrog method. This combination produces no dissipative error and the least dispersive error, which decreases as the CFL number increases until reaching the stability limit. This combination for the Cartesian grid constitutes Yee’s method [2]. Shown in Fig. 11a is the polar diagram of the normalized numerical phase shift of Yee’s method (equal spacing) at the maximum CFL number $\nu = \sqrt{2}/2$. Comparing this figure with Fig. 2c, one sees that the phase errors of the central difference are reduced by incorporating the leapfrog method for all N . Along the diagonals one has the perfect shift (no errors) for all frequency (i.e., independent of N). (The perfect shift actually can be obtained for any wave propagating in a direction θ by choosing the grid aspect ratio $\Delta x/\Delta y = \tan \theta$ and the time interval $c\Delta t = \Delta x \cos \theta = \Delta y \sin \theta$. Using Snell’s law one can also determine the grid spacings in dielectrics to preserve the perfect shift for all incident, reflected, and transmitted waves.) However, along the axes, the error is relatively large, which is 0.84% when $N = 10$ or 0.21% when $N = 20$, for example. For a propagation distance of one wavelength, this would develop a phase lag of 3.0° and 0.75° , respectively. This error is anisotropic and cumula-

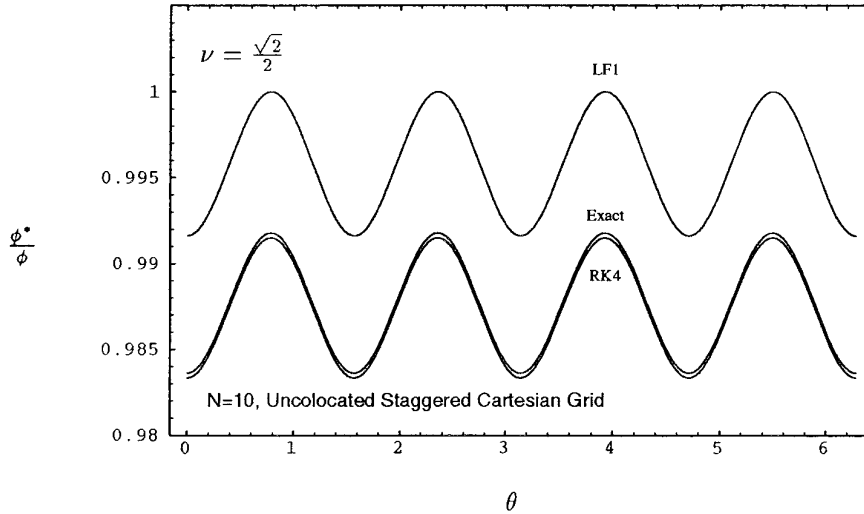


FIG. 10. Comparison of the normalized numerical phase shifts for time discretizations.

tive and may be troublesome for the class of scattering problems. By contrast, the phase shifts are almost isotropic for the hexagonal grid, as shown in Fig. 11b at the maximum CFL number $\nu = \sqrt{6}/3$. Although a perfect shift cannot be obtained at this maximum CFL number, the maximum error is much smaller than that on the Cartesian grid. For example, it is 0.14% when $N = 10$ or 0.034% when $N = 20$. This is equivalent to a 0.51° or a 0.12° phase lag per wavelength, respectively, which is six times as small as the Cartesian counterpart. For comparison, although we did not discuss the Lax–Wendroff scheme, Shankar’s method would give maximum errors of 1.58% in phase and 0.24% in amplitude when $N = 10$ and 0.41% in phase and 0.015% in amplitude when $N = 20$ at $\nu = 1$.

VII. A NEW UNSTAGGERED SCHEME

Among the schemes (second order) we have discussed, the centered schemes on uncollocated staggered grids are superior, in term of accuracy, to the others on regular grids. However, when extended to higher orders where larger stencils are to be used, they gradually lose their advantage. Furthermore, there exists a problem in forming the discretized equations on general grids. For a given control surface the spatial term requires the tangential components of the vector fields along the boundary edges ($\oint \mathbf{I} \cdot d\mathbf{l}$). However, the unknowns are given as the surface integrals of the normal components of the fields over the dual faces ($\int \mathbf{F} \cdot d\mathbf{S}$). Unless the primary and dual faces are regular and orthogonal, the two components may not be parallel

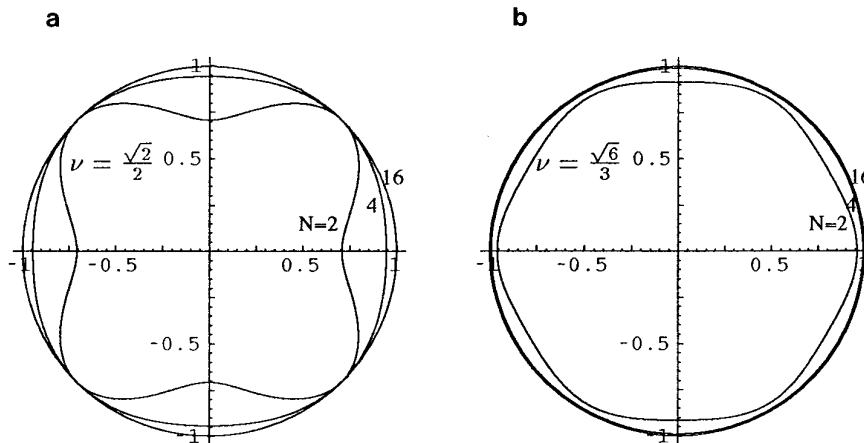


FIG. 11. Polar diagrams of the normalized numerical phase shift for the staggered leapfrog method: (a) staggered Cartesian grid; (b) staggered hexagonal grid.

or at the same locations. Therefore, the unknowns cannot be directly used in evaluating the spatial terms. In order to preserve the accuracy, the reconstruction of the complete vector fields from the neighboring components becomes necessary. The use of unstaggered grids can avoid such a tedious and time consuming process. There thus seems to exist a paradoxical situation that one prefers the staggered grids for their superior accuracy, on the one hand, and the unstaggered grids for their convenience, on the other.

Recently, we have introduced a high-accuracy unstaggered scheme for solving wave equations [21–23], in which an upwind or upwind-biased differencing is used for discretizing one field and its anti-symmetric counterpart is used for the other. There were two approaches used in determining the nodal weights. One is based on the standard Taylor expansions and the other is derived from non-linear optimizations of numerical errors. As we shall discuss below, the new scheme retains the desired properties of both the unstaggered and staggered grids that:

- (1) all components of electric and magnetic field vectors are collocated;
- (2) the system is non-dissipative, although each differencing for the first derivatives may be dissipative; and
- (3) all even and odd point unknowns are coupled.

In the remainder of this section, we first show that the new unstaggered scheme, when using two-point, first-order backward and forward differencings, is equivalent to the second-order staggered Yee scheme. We then extend this to general M -point differencings and show that the new scheme is capable of archiving an overall $2(M - 1)$ order of accuracy in space. Following the standard differencings, we also present an analysis for the optimized differencings. By varying the nodal weights using a non-linear optimization technique, we show that the new scheme is capable of producing accurate solutions for very high frequency waves. The implementation of the new scheme to practical problems using generalized curvilinear coordinates is being reported in a separate paper [24].

It is well known that the Maxwell curl equations (1) actually represent two vector wave equations. Eliminating the magnetic field in Eqs. (14c), one finds that the Laplacian of D_z on the uncolocated staggered grid is approximated by the second-order three-point central difference (δ^2),

$$(\delta_x^2 + \delta_y^2)D_{z_j,k} = \frac{D_{z_{j+1,k}} - 2D_{z_{j,k}} + D_{z_{j-1,k}}}{\Delta x^2} + \frac{D_{z_{j,k+1}} - 2D_{z_{j,k}} + D_{z_{j,k-1}}}{\Delta y^2}. \quad (38)$$

Similar approximations can be found for B_x and B_y if the divergence conditions (4) are satisfied exactly. The above

equation can also be obtained by a two-point first-order backward difference (∇) followed by a forward difference (Δ) in each direction or *vice versa* (i.e., $\delta^2 = \Delta\nabla = \nabla\Delta$). With this in mind, let us apply the backward difference to $\nabla \times \mathbf{H}$ and the forward difference to $\nabla \times \mathbf{E}$; Eq. (5) thus becomes

$$\begin{aligned} \frac{dD_{z_j,k}}{dt} &= \frac{H_{y_{j,k}} - H_{y_{j-1,k}}}{\Delta x} - \frac{H_{x_{j,k}} - H_{x_{j,k-1}}}{\Delta y} \\ \frac{dB_{x_{j,k}}}{dt} &= -\frac{E_{z_{j,k+1}} - E_{z_{j,k}}}{\Delta y} \\ \frac{dB_{y_{j,k}}}{dt} &= \frac{E_{z_{j+1,k}} - E_{z_{j,k}}}{\Delta x}. \end{aligned} \quad (39)$$

It is easy to verify that the normalized numerical phase speed for the above scheme is given as

$$\frac{c^*}{c} = 2 \left[\frac{\sin^2(\xi/2)}{\kappa^2 \Delta x^2} + \frac{\sin^2(\eta/2)}{\kappa^2 \Delta y^2} \right]^{1/2}, \quad (40)$$

which is exactly the same as that for the uncolocated staggered centered scheme (Yee scheme). The dissipative errors from the two first-order differencings cancel each other. As a result, the system becomes non-dissipative and second-order accurate. Comparing the two unstaggered schemes, Eqs. (39) and (14a), we see that the former is not only more compact than the latter, but also it couples all the even and odd point unknowns. It is not surprising that using the two first-order one-sided differencings for the spatial terms would give a more accurate result than using the second-order central differencing.

The above scheme can be extended to higher order. Let us consider a general stencil consisting of $(l + r + 1)$ points, in which there are l points to the left of point $x = x_j$ and r points to the right. Assuming $l \geq r$, the first x derivative of a function u (any components of \mathbf{E} or \mathbf{H}) can be approximated by an upwind or upwind-biased differencing

$$\nabla_x u_j = \frac{1}{\Delta x} \sum_{m=-l}^r a_m u_{j+m}. \quad (41a)$$

Here a_m are the nodal weights. The corresponding anti-symmetric counterpart can be written in the form

$$\Delta_x u_j = -\frac{1}{\Delta x} \sum_{m=-r}^l a_{-m} u_{j+m}, \quad (41b)$$

and its stencil consists of an opposite number of points. The y and z derivatives can be approximated in the same fashion. Traditionally, a_m are determined by matching the

coefficients of the Taylor series of the above equations. Since the difference equations need to be consistent with the differential equations as $\Delta x \rightarrow 0$, Eqs. (41) have to be at least first-order accurate, i.e., a_m must satisfy

$$\sum_{m=-l}^r a_m = 0 \quad (42a)$$

and

$$\sum_{m=-l}^r m a_m = 1. \quad (42b)$$

The order of accuracy for Eqs. (41) can be increased up to $l + r$, for which a_m must also satisfy

$$\sum_{m=-l}^r m^n a_m = 0 \quad \text{for } n = 2, 3, \dots, l + r. \quad (42c)$$

As we have seen above, in order to obtain a higher accuracy in space for the system, it is more important to check the accuracy of

$$\Delta_x \nabla_x u_j = -\frac{1}{\Delta x^2} \left[b_0 u_j + \sum_{m=1}^{l+r} b_m (u_{j+m} + u_{j-m}) \right], \quad (43)$$

where

$$b_m = \sum_{n=-l}^{r-m} a_n a_{n+m}. \quad (44)$$

One immediately sees that Eq. (43) is symmetric and thus non-dissipative. The maximum order of accuracy for the system in space could be doubled to $2(l + r)$ if b_m satisfy

$$\sum_{m=1}^{l+r} m^n b_m = \begin{cases} -\frac{b_0}{2} & n = 0; \\ -1 & n = 2; \\ 0 & n = 4, 6, \dots, 2(l + r). \end{cases} \quad (45)$$

Notice that Eqs. (45) are nonlinear in a_m , and thus not independent of Eqs. (42). Therefore, there could be more than one set of solutions or no solutions, dependent on what the desired orders of Eqs. (42) and (45) are.

In analogy to Eqs. (39), we now apply Eq. (41a) to $\nabla \times \mathbf{H}$ and the anti-symmetric counterpart equation (41b) to $\nabla \times \mathbf{E}$, or *vice versa*. The normalized numerical phase

speed for the new scheme can be derived as

$$\frac{c^*}{c} = \left[\frac{\lambda_x^2}{\kappa^2 \Delta x^2} + \frac{\lambda_y^2}{\kappa^2 \Delta y^2} \right]^{1/2}, \quad (46)$$

where

$$\lambda_x^2 = \left| \sum_{m=-l}^r a_m e^{im\xi} \right|^2 = b_0 + 2 \sum_{m=1}^{l+r} b_m \cos m\xi \quad (47a)$$

and

$$\lambda_y^2 = \left| \sum_{m=-l}^r a_m e^{im\eta} \right|^2 = b_0 + 2 \sum_{m=1}^{l+r} b_m \cos m\eta. \quad (47b)$$

Hereafter, for simplicity, we assume $\Delta x = \Delta y = \Delta s$. The combined space-time phase shift for the new scheme can be determined by substituting Eqs. (47) into Eqs. (37). For example, if the staggered leapfrog LF1 method is used for the time integration, the combined phase shift is

$$\phi^* = -\cos^{-1} \left[1 - \frac{1}{2} \nu^2 (\lambda_x^2 + \lambda_y^2) \right]. \quad (48)$$

For illustrative purposes, we use a two-point ($l = 1$ and $r = 0$) and two four-point ($l = 2$ and $r = 1$) differencings as examples. In all examples we require the differencing operators (41a) and (41b) to be at least first-order accurate, i.e., a_m must satisfy Eqs. (42a) and (42b). One can easily verify that the first two relations in Eqs. (45) are automatically satisfied, implying the system is at least second-order accurate. In the two-point case, there is only one set of solutions for a_m , i.e.,

$$\text{NS1: } a_{-1} = -1, \quad a_0 = 1.$$

In the four-point cases, additional relations from Eq. (42c) or (45) can be used to determine a_m , which give

$$\text{NS2: } \begin{aligned} a_{-2} &= \frac{1}{6}, & a_{-1} &= -1, \\ a_0 &= \frac{1}{2}, & a_1 &= \frac{1}{3}, \end{aligned}$$

and

$$\begin{aligned} a_{-2} &= -0.0100591535644260312, \\ a_{-1} &= 0.1246954507172851852, \\ \text{NS3: } a_0 &= -1.2192134407412922768, \\ a_1 &= 1.1045771435884331228, \end{aligned}$$

respectively. For the former, the spatial accuracy is third order for the differencing operators and fourth order for

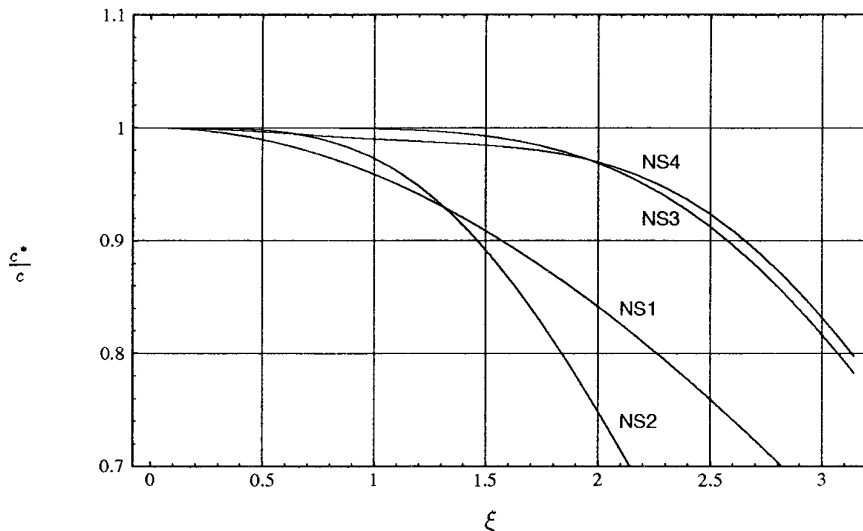


FIG. 12. Normalized numerical phase speed of the new unstaggered algorithm.

the system. However, for the latter, the differencing operators are only first-order accurate but the system is sixth-order accurate. Figure 12 shows their normalized numerical phase speeds at $\theta = 0$. The improvement of the four-point ones can be clearly observed, especially the scheme NS3. Here we see again that lower-order differencings for individual derivatives could result in a higher accuracy for the system due to error cancelations. Figure 12 also includes an optimized scheme NS4 which is discussed below.

So far we have discussed only those schemes for which the nodal weights are determined from the standard Taylor expansions. One problem with this type of differencing is that the scheme quickly becomes inaccurate for very high-frequency wave propagations (large κ) or very coarse-grid discretizations (large Δx or large Δy), as shown in previous figures. The reason is simply that the truncated Taylor series is no longer valid when ξ or η becomes large. To overcome this problem, Miranker [25], in 1971, presented a variety of difference schemes derived from constrained minimizations of the L_2 norms of the local truncation errors for a class of hyperbolic differential equations. Due to its complexity in solving the minimization analytically, the technique apparently went relatively unnoticed until it was recently rediscovered by many authors [21, 23, 26–28]. The minimization involved simultaneous treatments of spatial and time discretizations. To avoid this complexity, most authors separated the treatments of space and time and simplified the procedure to one-dimensional operators. This actually does not lead to a true minimization for the system of equations. The major difference between our approach and that of others is that we constructed the

object function for minimization from the full discretization of the equations. A comparison of several optimized schemes for acoustic equations was given in Ref. [23].

The most important element in the minimization problem is defining the object function. It is generally constructed from some norm of numerical errors subject to some equality and inequality constraints for designing a finite-difference scheme. The equality constraints may involve some requirements that the scheme be of a certain order of accuracy, and the inequality constraints must contain a condition that the scheme be stable. In order to accommodate the situation that incident and scattering waves may contain a spectrum of frequencies and travel in many possible directions, we can seek the minimum of an integrated error over a range of wavenumbers $[\zeta_l, \zeta_u]$, directions $[\theta_l, \theta_u]$, and CFL numbers $[\nu_l, \nu_u]$. Here the non-dimensional wavenumber, $\zeta \equiv \kappa \Delta s$, is the product of the physical wavenumber and the gridspacing. In the past, we constructed the object function from the integrated amplification factor error

$$\text{objfn} = \int_{\zeta_l}^{\zeta_u} \int_{\theta_l}^{\theta_u} \int_{\nu_l}^{\nu_u} \|\sigma - e^{-i\nu\zeta}\| \zeta d\zeta d\theta d\nu. \quad (49a)$$

If the system is non-dissipative, we also used the integrated phase error

$$\text{objfn} = \int_{\zeta_l}^{\zeta_u} \int_{\theta_l}^{\theta_u} \int_{\nu_l}^{\nu_u} (\phi^* - \phi)^2 \zeta d\zeta d\theta d\nu. \quad (49b)$$

Some authors preferred using the error of the solution itself. Depending on how we measure the errors, other

types of object functions can also be constructed. In the above equations, Gaussian quadratures were used in evaluating the integrations if analytical formulas were not available. The minimization problem was then solved either numerically by standard optimization techniques [29] or analytically using Mathematica [30] or other symbolic tools. In the following, we illustrate the procedure using the staggered leapfrog time integration together with the new unstaggered scheme (41). If we require the scheme be at least first-order accurate in space and time, the design variables just involve nodal weights a_m from the spatial discretization. Thus, the problem reduces to minimize the object function (49b), subject to the equality constraints Eqs. (42a) and (42b) and the inequality constraint $|1 - \frac{1}{2}\nu^2(\lambda_x^2 + \lambda_y^2)| \leq 1$. The inequality constraint actually sets the upper bound of the CFL number in terms of the nodal weights. The integration range of the CFL number in Eq. (49b) should always be less than or equal to this upper bound. From previous discussions, the numerical phase shift ϕ^* is given in Eq. (48) and the exact phase shift ϕ is equal to $-\nu\zeta$. The solution depends on the size of the stencil and the ranges of the parameters in Eq. (49b). If

we use a four-point stencil and choose the ranges from 0 to $\pi/2$ for ζ , 0 to 2π for θ , and 0 to 0.5 for ν , the minimization leads to

$$\begin{aligned} a_{-2} &= -0.020662670233282, \\ a_{-1} &= 0.147059221094637, \\ \text{NS4: } a_0 &= -1.232130431529782, \\ a_1 &= 1.105733880640202. \end{aligned}$$

In Fig. 13, we plot the normalized numerical phase shift for the new scheme at $\nu = 0.5$ using the four sets of coefficients NS1, NS2, NS3, and NS4. These schemes are second-order in time and second-order, fourth-order, sixth-order, and second-order accurate in space for the system, respectively. When N is large, the accuracy follows the orders of the schemes. However, when N becomes very small, order loses its meaning, as we can see from the figures that for $N = 4$ the second-order NS4 is the most accurate one and the fourth-order NS2 is the most inaccurate one. Notice that here the time discretization LF1 uses only a two-point

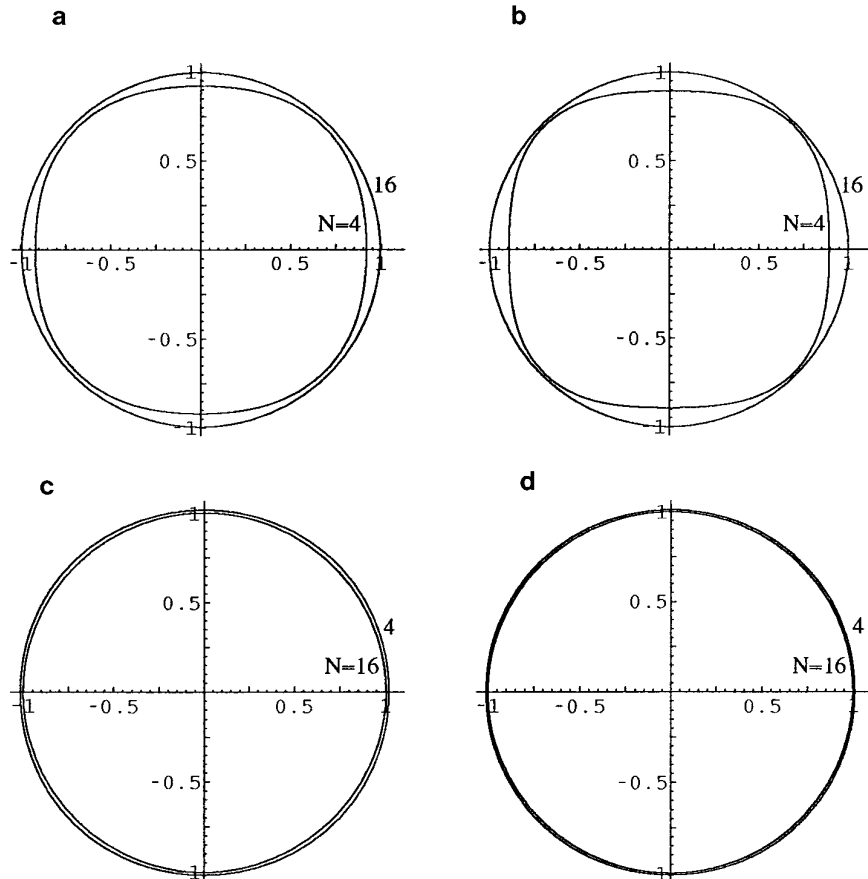


FIG. 13. Polar diagrams of the normalized numerical phase shift for the new unstaggered scheme at $\nu = 0.5$: (a) NS1; (b) NS2; (c) NS3; (d) NS4.

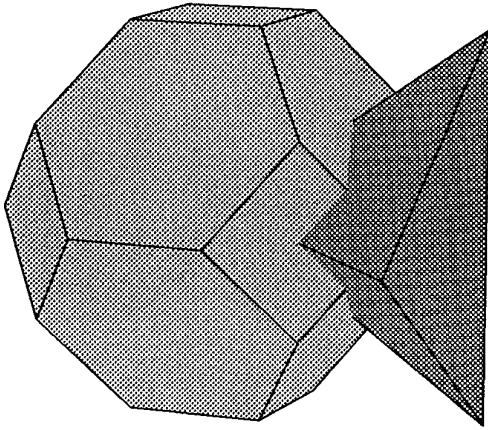


FIG. 14. Tetradekahedron and dual tetrahedron.

stencil and no minimization. Utilizing a larger stencil or multistep methods and employing the minimization in time can further improve the accuracy. Nevertheless, NS4 does produce a reasonable accurate solution for a TE wave incident on a conducting cylinder in a grid consisting of only four to five points per wavelength [24]. As a final note, we like to emphasize that higher order schemes, strictly speaking, are only superior asymptotically in the process of refining the gridspacing as $\Delta s \rightarrow 0$ or reducing the frequency as $\kappa \rightarrow 0$. For given ranges of wavenumbers (physical wavenumbers and gridspacings), directions, and CFL numbers, one can construct a scheme which is formally lower order but is more accurate than the standard ones within those ranges. The minimization can be considered as a “trade-off” process that we lower the accuracy

of a scheme in one region in exchange a better resolution in the other.

VIII. THREE-DIMENSIONAL ANALYSIS

There are only a few regular three-dimensional shapes known to crystallographers which are space-filling. Those include the hexahedron, tetradekahedron (truncated octahedron), hexagonal prism, rhombic dodekahedron, and elongated rhombic dodekahedron [31]. Among them, the Cartesian grid consisting of ordered arrays of hexahedra is the simplest. The one corresponding to the hexagonal grid in three dimensions is the tetradekahedral grid. As shown in Fig. 14, a tetradekahedron has 6 square faces and 8 regular hexagonal faces, and its dual is a tetrahedron whose faces are isosceles triangles with sides in the ratio of $\sqrt{3}$ to 2.

Extension of the analyses for the discretizations discussed in Sections IV and VI to the three-dimensional Cartesian grid is straightforward. All the properties for the two-dimensional schemes are also valid for the three-dimensional schemes. However, efficient and accurate numerical schemes for the other grids currently are not available. Here we show only a preliminary study for an extension of Yee’s method to the staggered tetradekahedral grid, a compact centered scheme consisting of 19 independent unknown field components, 12 defined on the edges of tetradekahedra, and 7 on the edges of the dual tetrahedra. A detailed analysis shall be reported in the future. Shown in Fig. 15 are the three-dimensional polar diagrams of the normalized numerical phase shifts of Yee’s method at $N = 2$ and $\nu = 0.5$ for the Cartesian and tetradekahedral grids. Again, the error is greatest along the axes and least

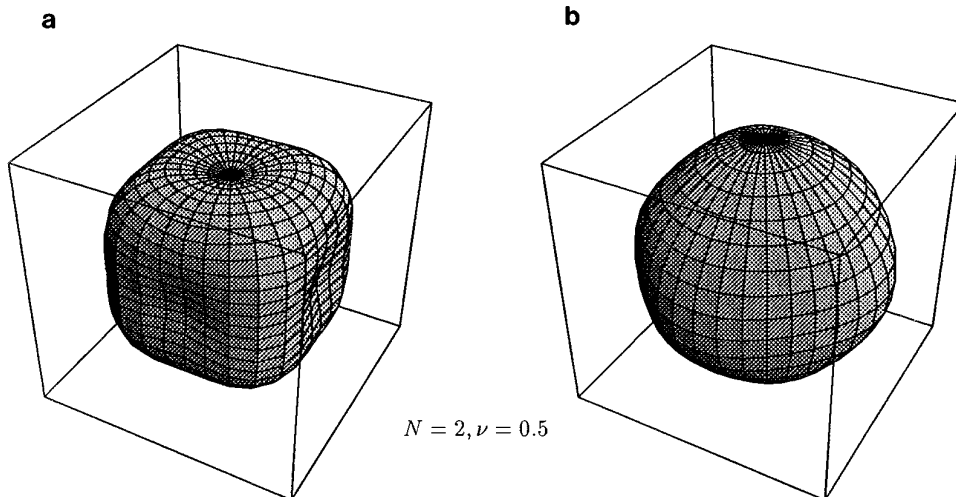


FIG. 15. Polar diagrams of the normalized numerical phase shifts of Yee’s method ($N = 2$, $\nu = 0.5$): (a) staggered Cartesian grid; (b) staggered tetradekahedral grid.

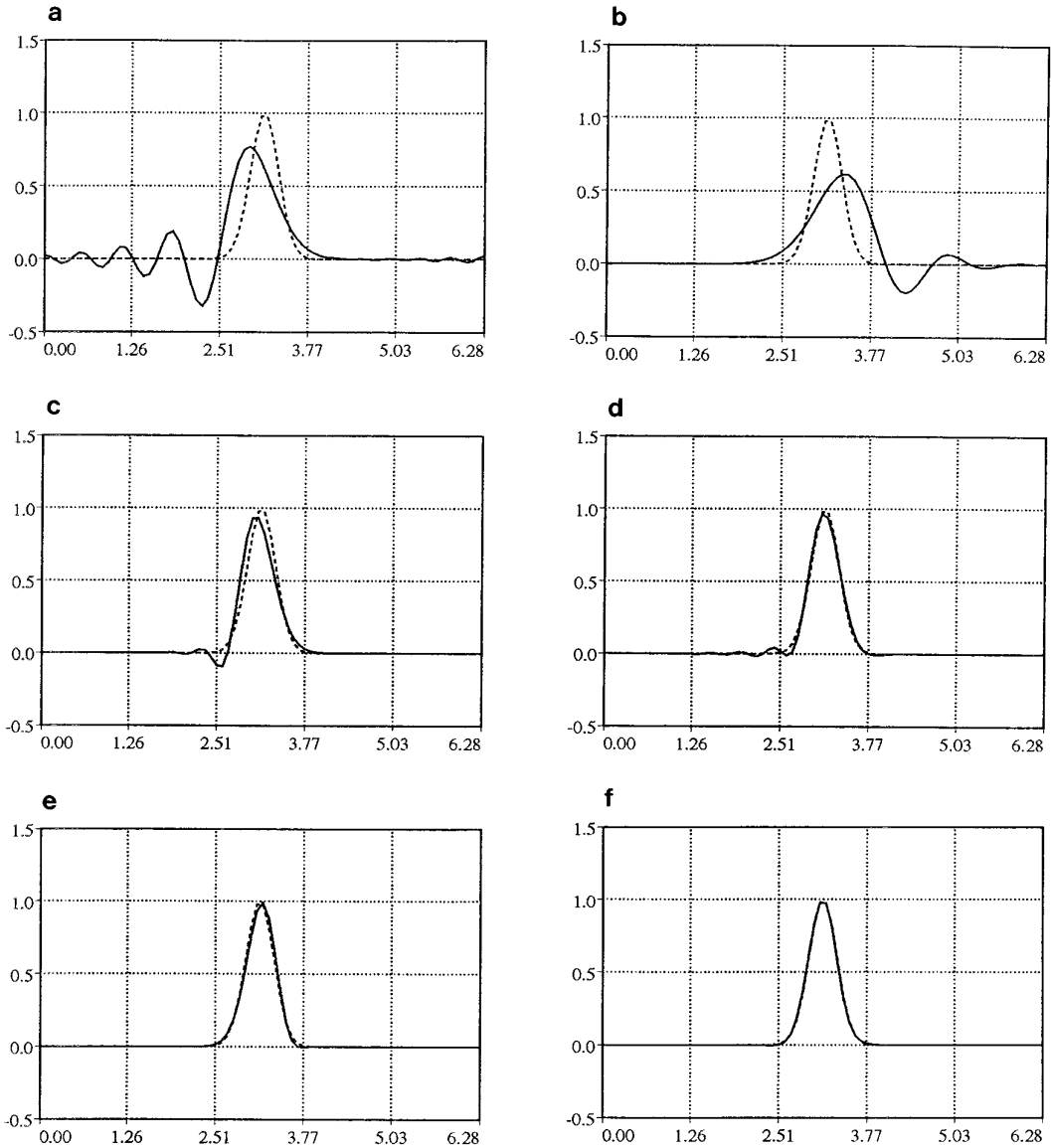


FIG. 16. Numerical solutions of a pulse propagating in free space: (a) CN2; (b) UP2; (c) NS1; (d) NS2; (e) NS3; (f) NS4.

along the diagonals for the Cartesian grid and is nearly isotropic for the tetracaedrahedral grid.

IX. NUMERICAL RESULTS

The Fourier method predicts precisely what a numerical scheme would produce for a plane wave propagating in free space when uniform grids and periodic boundary conditions are used. It is easy to conduct a numerical experiment and verify that the numerical solution and the analytical solution are identical under this situation. Some comparisons of various schemes for a plane wave incident on a perfectly conducting cylinder and sphere have been presented recently in Ref. [32]. Here we present another

experiment for a pulse propagating in free space using the schemes we have analyzed. The exact solution for the pulse takes the form $e^{-(1/2)(kx-\omega t)^2}$. The computation domain consists of a 75-point equally spaced Cartesian grid in the region between 0 and 2π . Periodic conditions were used. The numerical solution was started when the pulse had its maximum value at $x = \pi$. Six schemes were used to compare the numerical solutions for $k = 15$. These include the unstaggered, second-order central (CN2) and upwind (UP2) schemes, and the new schemes using the four set of coefficients NS1 to NS4. Notice that NS1 is equivalent to the Yee scheme. The time integration was carried out by the Runge–Kutta RK3 for UP2 and the staggered leapfrog LF1 for the other five using the CFL number $\nu = 0.5$.

From the Fourier transfer, we notice that the pulse consists of a fair amount of harmonic waves between $\kappa = 1$ and $\kappa = 15$, corresponding to $N = 75$ and $N = 5$. Since each component of the harmonic waves travels at a different numerical phase speed, the numerical solution of the pulse loses its shape as the time evolves. Figure 16 clearly manifests this phenomenon. Shown here are the solutions for each scheme after one cycle when the pulse returns to its original position, in which the solid lines are the numerical solutions and the dashed lines are the exact solution. We have discussed the phase and modulus of each component for the six schemes. Here the solutions show the composite effects from all the components. This results in an oscillation train in the solutions behind the pulse for CN2 and ahead the pulse for UP2, as observed in Figs. 16a and b. Notice that the scheme UP2 is dissipative, which depresses the high-frequency oscillations. The solution thus looks smoother, but the amplitude has a larger error than that of CN2. Both schemes produce very poor results. Using NS1 or the Yee scheme greatly improves the accuracy as demonstrated in Fig. 16c. However, it is still unable to resolve the high frequency waves for $N \leq 15$. As shown in Fig. 16d, employing the four-point NS2 only makes the solution a little better than that of the two-point NS1, although the former is formally two orders of magnitude more accurate than the latter. A better improvement can be obtained from NS3 and NS4, as shown in Figs. 16e and f. The difference between NS2 and NS3 or NS4 is only the nodal weights. Those coefficients were determined by seeking the maximum order of accuracy for the operators, for the system, and the minimum phase error, respectively. Among the six schemes, the optimized NS4 produces the most accurate solution, which is almost identical to the exact solution.

X. CONCLUDING REMARKS

In this paper, the Fourier method has been used to analyze the dispersive, dissipative, and isotropy errors of various spatial and time discretizations applied to the Maxwell equations on multi-dimensional grids. Both Cartesian grids and non-Cartesian grids based on hexagons for two dimensions and tetradecahedra for three dimensions were studied and compared. Effects of data location on staggered or unstaggered grids and data management for collocated or uncollocated components were investigated. The numerical errors were quantitatively determined in terms of phase speed, wavenumber, propagation direction, grid-spacings, and CFL number. The study shows that the upwind intensity–vector splitting scheme and the Riemann solver generate dissipation and higher dispersion and require more operations than the centered schemes. Furthermore, the main purpose of upwind schemes, which is to capture discontinuities automatically, seems to provide no

advantage for computational electromagnetics, since the discontinuities are generally known at material interfaces. Therefore, the centered schemes are more efficient and accurate than the upwind schemes. For the centered schemes, the staggered grids produce less errors than the unstaggered ones and the non-Cartesian grids yield superior isotropy than the Cartesian ones. We have pointed out that a proper choice of time discretization can reduce the overall numerical errors due to the spatial discretization. The study shows that while the fourth-order Runge–Kutta method is more accurate than the second-order leap-frog method timewise, the former is actually less accurate than the latter when combined with spatially centered schemes. A preliminary study demonstrates the superiority of the three-dimensional tetradecahedral grid over the Cartesian grid. We also introduced a new unstaggered algorithm which eliminates the inconvenience of the uncollocated staggered scheme. Using an optimization technique to determine the nodal weights, the new scheme provides the highest accuracy among all the schemes discussed.

REFERENCES

1. R. F. Harrington, *Field Computation by Moments Methods* (Macmillan Co., New York, 1968).
2. K. S. Yee, *IEEE Trans. Antennas Propagat.* **14**, 302 (1966).
3. A. Taflove, *IEEE Trans. Electromagn. Compat.* **22**, 191 (1980).
4. A. Taflove and M. E. Brodwin, *IEEE Trans. Microwave Theory Tech.* **23**, 623 (1975).
5. A. Taflove and K. R. Umashankar, *Proc. IEEE* **77**, 682 (1989).
6. G. Mur, *IEEE Trans. Electromagn. Compat.* **23**, 377 (1981).
7. T. G. Moore, J. G. Blaschak, A. Taflove, and G. A. Kriegsmann, *IEEE Trans. Antennas Propagat.* **36**, 1797 (1988).
8. T. G. Jurgens, A. Taflove, K. R. Umashankar, and T. G. Moore, *IEEE Trans. Antennas Propagat.* **40**, 357 (1992).
9. K. S. Yee, J. S. Chen, and A. H. Chang, *IEEE Trans. Antennas Propagat.* **40**, 1068 (1992).
10. M. Fusco, *IEEE Trans. Electromagn. Compat.* **38**, 76 (1990).
11. M. Fusco, M. V. Smith, and L. W. Gordon, *IEEE Trans. Electromagn. Compat.* **39**, 1463 (1991).
12. R. Holland, *IEEE Trans. Nuclear Sci.* **NS-30**, 4589 (1993).
13. M. Vinokur and M. Yarrow, AIAA Paper 93-0463, 1993 (unpublished).
14. V. Shankar, W. F. Hall, and A. H. Mohammadian, AIAA Paper 89-1987, 1989 (unpublished).
15. V. Shankar, A. H. Mohammadian, and W. F. Hall, *Electromagnetics* **10**, 127 (1990).
16. E. Turkel, *J. Comput. Phys.* **15**, 226 (1974).
17. R. Vichnevetsky and J. B. Bowles, *Fourier Analysis of Numerical Approximations of Hyperbolic Equations* (SIAM, Philadelphia, 1982).
18. L. N. Trefethen, *SIAM Rev.* **23**, 113 (1982).
19. R. W. Noack and D. A. Anderson, AIAA Paper 92-0451, 1992 (unpublished).
20. Y. Liu, "A Generalized Finite-Volume Algorithm for Solving the Maxwell Equations on Arbitrary Grids," in *10th Annual Review of*

- Progress in Applied Computational Electromagnetics*, the Applied Computational Electromagnetics Society, (1994), p. 487.
21. Y. Liu, NASA Ames Research Center, 1992 (unpublished).
 22. Y. Liu, AIAA Paper 93-0368, 1993 (unpublished).
 23. C. L. Chen and Y. Liu, AIAA Paper 93-4324, 1993 (unpublished).
 24. R. Janaswamy and Y. Liu, in preparation.
 25. W. L. Miranker, *Numer. Math.* **17**, 124 (1971).
 26. S. K. Lele, *J. Comput. Phys.* **103**, 16 (1992).
 27. C. K. W. Tam and J. C. Webb, *J. Comput. Phys.* **107**, 262 (1993).
 28. Z. Haras and S. Ta'asan, *J. Comput. Phys.* **114**, 265 (1995).
 29. G. N. Vanderplaats, *Numerical Optimization Techniques for Engineering Design: With Applications* (McGraw-Hill, New York, 1984).
 30. *Mathematica: A System for Doing Mathematics by Computer* (Wolfram Research, Champaign, 1988–93).
 31. J. V. Smith, *Geometrical and Structural Crystallography* (Wiley, New York, 1982), p. 77.
 32. Y. Liu, "A 3D High-Order Unstructured Finite-Volume Algorithm for Solving Maxwell's Equations," in *1995 USNC/URSI Radio Science Meeting, Newport Beach, CA, 1995*.