ON THE NUMERICAL SOLUTION OF THE 2D WAVE EQUATION WITH COMPACT FDTD SCHEMES

Maarten van Walstijn
Sonic Arts Research Centre,
Queen's Univeristy Belfast, United Kingdom
m.vanwalstijn@qub.ac.uk

*Konrad Kowalczyk,**
Sonic Arts Research Centre,
Queen's Univeristy Belfast, United Kingdom
kkowalczyk01@qub.ac.uk

ABSTRACT

This paper discusses compact-stencil finite difference time domain (FDTD) schemes for approximating the 2D wave equation in the context of digital audio. Stability, accuracy, and efficiency are investigated and new ways of viewing and interpreting the results are discussed. It is shown that if a tight accuracy constraint is applied, implicit schemes outperform explicit schemes. The paper also discusses the relevance to digital waveguide mesh modelling, and highlights the optimally efficient explicit scheme.

1. INTRODUCTION

Numerically solving the wave equation is of interest in a range of audio-related contexts, including physics-based sound synthesis and virtual room auralisation. The FDTD method has been known for decades as one of the available numerical tools [1, 2] and has been applied regularly as such (see, for example, [3, 4, 5]). However in the 'audio community', 2D and 3D acoustic systems are more often modelled using the strongly related digital waveguide mesh (DWM) approach. Given that DWM techniques as well as the related wave-variable approach of multi dimensional wave digital filters [6] can be seen as sub-classes of the wider domain of finite difference methods [7], it can be argued that the most comprehensive understanding of any of these methods can be obtained by at least also viewing them from a FDTD perspective. This notion is perhaps best examplified by the fact that the "digital waveguide mesh" was at its inception in [8] immediately analysed using classic FDTD methods. Further support for this reasoning is provided by the fact that there is a vast body of existing FDTD literature to draw from, most of which was developed over the past four decades in the context of electrodynamics.

The aim of this paper is to provide a better insight in the FDTD approximation of the 2D wave equation, which is applicable to 2D room modelling as well as to the simulation of non-stiff 2D resonators (i.e. membranes). As such we restrict ourselves to compact schemes on a rectilinear mesh (i.e. schemes with stencils not wider than 3 nodes in the discrete space-time domain), since higher-order spatial and temporal approximation can massively complicate the treatment of boundaries. This analysis covers a wide range of well known methods, including the standard leapfrog scheme [1], the standard rectilinear DWM [8], the interpolated DWM [9], as well as implicit schemes [10]. The latter are not as well known and far less used, probably because the computational costs of a straight-forward implementation using matrixes is extremely expensive. However, a general family of compact schemes can be defined that allows fast implementation using alternating direction implicit (ADI) methods, which, depending on the relevant critera, can lead to a significantly improved efficiency in comparison with standard explicit schemes, making implicit schemes an important option for high sample rate audio applications [10].

The aforementioned family of compact schemes is investigated in terms of stability, order of accuracy, and dispersion error, with the focus mainly on the latter. Regarding dispersion, much of the attention in the DWM literature has been on finding isotropic methods. Although some consideration to this concept is given here, a larger emphasis is placed on minimising the general dispersion error rather than minimising its dependence on direction. This choice stems from the authors' long-term objective of modelling 2D and 3D acoustic systems (in particular acoustic spaces) with moving sources and receivers, which excludes the use of off-line frequency warping methods, such as those applied in [9, 11].

The analysis presented here includes a novel way of displaying the dispersion as a function of frequency and direction rather than as a function of spatial frequencies (wavenumbers), which allows a more intuitive and direct interpretation of the data. The main results and conclusions of the paper are centered around how particular cases within the family compare to each other in terms of computational efficiency.

2. COMPACT FDTD SCHEMES

2.1. General Formulation

The general objective is the numerical solution of the wave equation in a two-dimensional (x, y) coordinate system:

$$\frac{\partial^2 p}{\partial t^2} = c^2 \left(\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} \right), \tag{1}$$

where p is the variable (that could for example be acoustic pressure or membrane displacement), and c is the wave velocity. In the ensuing formulations, the following discrete space-time domain notations are used:

$$p_{l,m}^n \equiv p(x,y,t)\Big|_{x=lX, y=mX, t=nT}, \tag{2}$$

$$\delta_t^2 p_{l,m}^n \equiv p_{l,m}^{n+1} - 2p_{l,m}^n + p_{l,m}^{n-1}, \tag{3}$$

$$\delta_x^2 p_{l,m}^n \equiv p_{l+1,m}^n - 2p_{l,m}^n + p_{l-1,m}^n, \tag{4}$$

$$\delta_y^2 p_{l,m}^n \equiv p_{l,m+1}^n - 2p_{l,m}^n + p_{l,m-1}^n. \tag{5}$$

Using the centered finite difference operators in Eqs. (3), (4), and (5), a family of compact FD schemes approximating the wave

* This work was supported by the European Social Fund.

equation can be formulated as follows [12]

$$[1 + a(\delta_x^2 + \delta_y^2) + a^2\delta_x^2\delta_y^2]\delta_t^2 p_{l,m}^n = \lambda^2[(\delta_x^2 + \delta_y^2) + b\delta_x^2\delta_y^2]p_{l,m}^n, \quad (6)$$

where $\lambda = cT/X$ is the Courant number, and a and b are free numerical parameters. This formulation is consistent with the wave equation for any real (a, b) (this is easily verified using equations (42), (43), and (44)), and the resulting scheme is implicit except for $a = 0$. While a more generalised family of compact schemes is possible [5], it is useful to restrict the analysis of possible implicit schemes to those that can be factorised into two one-dimensional implicit equations. In that case, an alternating direction implicit (ADI) formulation exists, which allows a fast implementation using the Thomas algorithm for matrix inversion [1, 2]. In the light of that constraint, the formulation in (6) is particularly useful since it allows splitting into the following ADI form, which is more efficient than alternative splitting formulae [12]

$$(1 + a\delta_x^2)p_{l,m}^{n+1*} = \frac{\lambda^2}{a}[-1 + (a - b)\delta_y^2]p_{l,m}^n, \quad (7)$$

$$(1 + a\delta_y^2)\delta_t^2 p_{l,m}^n = p_{l,m}^{n+1*} + \frac{\lambda^2}{a}(1 + b\delta_y^2)p_{l,m}^n, \quad (8)$$

where $p_{l,m}^n$ is the update variable and $p_{l,m}^{n+1*}$ is an intermediate value.

2.2. Single-Frequency Plane-Wave Solutions

Many standard methods for analysis of FDTD schemes are based on examining single-frequency plane-wave solutions. Consider such a wave travelling at position x' along an axis that cuts the Cartesian x-axis at an angle θ. In the continuous space-time domain, solutions of this kind can be written

$$p(x', t) = p_0\, e^{st} e^{-jkx'}, \quad (9)$$

where p_0 is a real-valued amplitude value, $s = \sigma + j\omega$ denotes complex frequency, and k is the wavenumber. Using the coordinate rotation $x' = x\cos(\theta) + y\sin(\theta)$, this becomes

$$p(x, y, t) = p_0\, e^{st} e^{-jk_x x} e^{-jk_y y}, \quad (10)$$

where $k_x = k\cos\theta$ and $k_y = k\sin\theta$. In the discrete space-time domain, the single-frequency plane-wave solution is:

$$p_{l,m}^n = p_0\, e^{snT} e^{-j\hat{k}_x lX} e^{-j\hat{k}_y mX}, \quad (11)$$

where $\hat{k}_x$ and $\hat{k}_y$ denote the effective numerical wavenumbers, which differ from the continuous-domain wavenumbers and are related to a wavenumber $\hat{k}$ associated with the propagation direction along the x'-axis by

$$\hat{k}^2 = \hat{k}_x^2 + \hat{k}_x^2. \quad (12)$$

For solutions of this form, the approximation in (3) can be written

$$\delta_t^2 p_{l,m}^n = \left(z - 2 + z^{-1}\right) p_{l,m}^n, = -4\sin^2(\omega T/2)\, p_{l,m}^n, \quad (13)$$

where $z = e^{sT}$, which represents the classic relationship between the s- and the z-domain found in DSP literature. Hence by the z-transform shift theorem, we have

$$p_{l,m}^{n+1} = z\, p_{l,m}^n, \quad (14)$$

which is easily verified by substituting both sides of Eq. (14) with the form of Eq. (11). Similarly, under the assumption of single-frequency plane-wave solutions the spatial approximations in (4) and (5) can be written

$$\delta_x^2 p_{l,m}^n = -4\sin^2(\hat{k}_x X/2)\, p_{l,m}^n, \quad (15)$$

$$\delta_y^2 p_{l,m}^n = -4\sin^2(\hat{k}_y X/2)\, p_{l,m}^n. \quad (16)$$

2.3. Von Neuman Stability Analysis

Classic von Neumann stability analysis investigates an FDTD scheme for solutions of the form of Eq. (11), and seeks to establish a bound on λ such that no growing solutions exist [1, 2]. From (14) it is clear that any scheme is unstable for $|z| > 1$, hence the necessary stability condition can be expressed as $|z| \leq 1$. By substituting (13), (15), and (16) into (6), the following equation in z can be obtained:

$$z + 2B(s_x, s_y) + z^{-1} = 0, \quad (17)$$

where following [5] we introduce the new variables

$$s_x = \sin^2(\hat{k}_x X/2), \quad (18)$$

$$s_y = \sin^2(\hat{k}_y X/2), \quad (19)$$

and where

$$B(s_x, s_y) = 2\lambda^2 F(s_x, s_y) - 1, \quad (20)$$

with

$$F(s_x, s_y) = \frac{s_x + s_y - 4bs_x s_y}{1 - 4a(s_x + s_y) + 16a^2 s_x s_y}. \quad (21)$$

In FDTD literature, Eq. (17) is known as the amplification equation or the amplification polynomial. The moduli of its two solutions have to be smaller than or equal to unity for any combination (s_x, s_y) and thus any combination $(\hat{k}_x, \hat{k}_y)$. Since s_x and s_y are periodic with π, it is sufficient to consider only real-valued wavenumbers in the range $-\pi/X \leq \hat{k}_x, \hat{k}_y \leq \pi/X$. We note that $\hat{k}_x$ and $\hat{k}_y$ can also become complex-valued [13], but in that case only the real part has to be taken into account with regard to stability analysis. From (17), it can be shown that $|z| \leq 1$ when

$$B^2(s_x, s_y) \leq 1 \quad (22)$$

which yields the bound on λ:

$$\lambda^2 \leq \frac{1}{F_{\max}(s_x, s_y)} \quad (23)$$

For any (a, b) the function $F(s_x, s_y)$ reaches its maximum at one of its extrema, where $s_x, s_y \in [0, 1]$, thus

$$F_{\max} = \max\left(\frac{1}{1 - 4a}, \frac{2 - 4b}{1 - 8a + 16a^2}\right) \quad (24)$$

Therefore the necessary stability condition for the scheme in (6) is

$$\lambda^2 \leq \min\left(1 - 4a, \frac{1 - 8a + 16a^2}{2 - 4b}\right). \quad (25)$$

Since λ^2 must be positive, it follows from (25) that we have the auxiliary constraints:

$$a \leq \frac{1}{4}, \quad b \leq \frac{1}{2}. \quad (26)$$

2.4. Numerical Dispersion

If we re-arrange (17) to

$$z - 2 + z^{-1} = -4\lambda^2 F(s_x, s_y), \tag{27}$$

and, using (13), substitute $z - 2 + z^{-1} = -4\sin(\omega T/2)$, then the dispersion relation results:

$$\sin^2(\omega T/2) = \lambda^2 F(s_x, s_y). \tag{28}$$

The frequency for (s_x, s_y) is therefore

$$\omega = \frac{2}{T}\arcsin\left(\lambda\sqrt{F(s_x, s_y)}\right). \tag{29}$$

The relative phase velocity, defined as the ratio of the effective numerical wave speed $\hat{c} = \omega/\hat{k}$ over the real wave speed c, is a useful measure of numerical dispersion. Using (29) and (12), one obtains

$$v(\hat{k}_x, \hat{k}_y) = \frac{\omega}{\hat{k}\,c} = \frac{2\arcsin\left(\lambda\sqrt{F(s_x, s_y)}\right)}{\lambda\sqrt{(\hat{k}_x X)^2 + (\hat{k}_y X)^2}}. \tag{30}$$

2.5. Axial Phase Velocity

Consider now wave propagation in one of the four axial directions, where $\hat{k}_a$ denotes the wavenumber. Then we have either $\left\{\hat{k}_x^2 = \hat{k}_a^2, \hat{k}_y^2 = 0\right\}$ or $\left\{\hat{k}_x^2 = 0, \hat{k}_y^2 = \hat{k}_a^2\right\}$, which in both cases allows writing the variable F as a function of the axial wavenumber:

$$F(\hat{k}_a) = \frac{\sin^2(\hat{k}_a X/2)}{1 - 4a\sin^2(\hat{k}_a X/2)}. \tag{31}$$

Then from Eq. (28), the axial wavenumber can be written as a unique function of frequency:

$$\hat{k}_a(\omega) = \frac{2}{X}\arcsin\sqrt{\frac{\sin^2(\omega T/2)}{\lambda^2 + 4a\sin^2(\omega T/2)}}, \tag{32}$$

and finally the relative phase velocity in axial directions can now also be written as a function of frequency

$$v_a(\omega) = \frac{\omega}{\hat{k}_a(\omega)\,c} = \frac{(\omega T/2)}{\lambda\arcsin\sqrt{\frac{\sin^2(\omega T/2)}{\lambda^2 + 4a\sin^2(\omega T/2)}}}. \tag{33}$$

2.6. Diagonal Phase Velocity

In diagonal directions, we have $\hat{k}_x^2 = \hat{k}_y^2 = \hat{k}_d^2/2$ (subscript d denoting diagonal directions), which allows writing F as a function of the diagonal wavenumber:

$$F(\hat{k}_d) = \frac{2G(\hat{k}_d) - 4bG^2(\hat{k}_d)}{1 - 8aG(\hat{k}_d) + 16a^2G^2(\hat{k}_d)}, \tag{34}$$

where $G(\hat{k}_d) = \sin^2(\sqrt{2}\hat{k}_d X/4)$. Forcing the dispersion relation in (28) and writing G as a function of frequency yields

$$G(\omega) = \frac{4as_t - \lambda^2 - \sqrt{(8a - 4b)\lambda^2 s_t + \lambda^4)}}{16a^2 s_t + 4b\lambda^2}, \tag{35}$$

where $s_t = \sin^2(\omega T/2)$. Using the definition of G, we may then write the diagonal wavenumber as a function of frequency:

$$\hat{k}_d(\omega) = \frac{4}{\sqrt{2}X\arcsin\sqrt{G(\omega)}}. \tag{36}$$

Finally, the relative wave velocity in diagonal directions as a function of frequency becomes

$$v_d(\omega) = \frac{\omega}{\hat{k}_d(\omega)\,c} = \frac{\sqrt{\frac{1}{2}}(\omega T/2)}{\lambda\arcsin\sqrt{G(\omega)}}. \tag{37}$$

2.7. Cut-Off Frequencies

The function $f(x) = \arcsin(x)$ becomes complex-valued when $x > 1$. Hence from (32), the axial wavenumber becomes complex-valued if

$$\frac{\sin^2(\omega T/2)}{\lambda^2 + 4a\sin^2(\omega T/2)} > 1, \tag{38}$$

which can be simplified to

$$\omega T > 2\arcsin\left(\frac{\lambda}{\sqrt{1 - 4a}}\right). \tag{39}$$

For frequencies above this 'cut-off' frequency, waves are evanescent, and resonances are not sustained [13]. Such a cut-off frequency in fact exists for all propagation directions, and can be found by considering for which frequencies the dispersion relation in (28) can be satisfied only for complex-valued wavenumbers, which yields:

$$\omega T > 2\arcsin\left(\lambda\sqrt{F_{mr}(\theta)}\right). \tag{40}$$

where $F_{mr}(\theta)$ is the maximum value of $F(s_x, s_y)$ for a pair of real-valued wavenumbers for which the angle of propagation is θ:

$$F_{mr}(\theta) = \begin{cases} F(1, \sin^2(\pi/2\tan\theta), & |\tan\theta| \le 1 \\ F(\sin^2(\pi/2\tan\theta), 1), & |\tan\theta| > 1 \end{cases} \tag{41}$$

2.8. Order of Acuracy

Truncating to fourth order the Taylor expansions about the point (x, y, t) of the difference operators in Eqs. (3), (4), and (5) yields:

$$\delta_t^2 = T^2\frac{\partial^2}{\partial t^2} + \frac{T^4}{12}\frac{\partial^4}{\partial t^4} + O(T^4), \tag{42}$$

$$\delta_x^2 = X^2\frac{\partial^2}{\partial x^2} + \frac{X^4}{12}\frac{\partial^4}{\partial x^4} + O(X^4), \tag{43}$$

$$\delta_y^2 = X^2\frac{\partial^2}{\partial y^2} + \frac{X^4}{12}\frac{\partial^4}{\partial y^4} + O(X^4). \tag{44}$$

Substitution of the above equations into the general scheme of (6) and applying the required successive substitutions of derivatives according to the modified equation method [14, 5] then allows expressing the overall numerical error or modified equation error as

$$\begin{aligned} E(p, a, b, \lambda, X) \;=\; & -c^2\left[\left(a + \frac{\lambda^2 - 1}{12}\right)\frac{\partial^4 p}{\partial x^4}\right. \\ & +\; \left.\left(b - \frac{1}{6}\right)\frac{\partial^4 p}{\partial x^2 \partial y^2}\right]X^2 + O(X^4). \end{aligned} \tag{45}$$

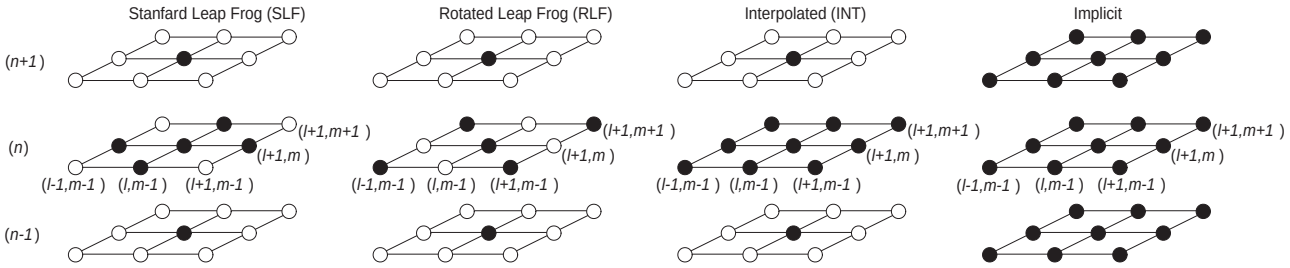

Figure 1: *Compact FDTD Stencils. For each type of stencil, the middle node at time* $(n+1)$ *is updated using the remaining black-coloured nodes.*

3. ANALYSIS OF SPECIAL CASES

Table 1 lists seven special cases of compact schemes. The standard leapfrog (SLF) scheme is mathematically equivalent to the rectilinear digital waveguide mesh [8], and has exactly the same dispersion relation as Yee's classic FDTD scheme [15, 2]. As explained in [5], the rotated leap fog (RLF) scheme can be interpreted as applying standard centered finite operators along the diagonals rather than the horizontal and vertical axes. One could say that the RLF stencil is 45^o rotated in comparison to the SLF scheme; both use a "6-point stencil" in the space-time grid to update the new value at time $(n+1)$ (see Fig. 1). Linear combinations of the SLF and the RLF also exist, which result in a 10-point update stencil (see Fig. 1). Such schemes, for which $a = 0$ and $0 < b < \frac{1}{2}$, have been referred to as 'interpolated schemes' [5]. We have listed two special cases of interpolated schemes in Table 1. As a reference to other studies, it is useful to point out that the interpolated scheme with $b = \frac{1}{6}$ and $\lambda = \sqrt{\frac{1}{2}}$ is nearly equivalent to the interpolated digital waveguide mesh [10], the latter having $b = 0.1879$.

Table 1: *Special cases of compact FDTD schemes.*

scheme	a	b	stability bound
SLF	0	0	$\lambda \leq \sqrt{\frac{1}{2}}$
RLF	0	$\frac{1}{2}$	$\lambda \leq 1$
INT(1/4)	0	$\frac{1}{4}$	$\lambda \leq 1$
INT(1/6)	0	$\frac{1}{6}$	$\lambda \leq \sqrt{\frac{3}{4}}$
MFI	$\frac{1}{4} - \frac{1}{2\sqrt{3}}$	$\frac{1}{6}$	$\lambda \leq 1$
FOA	$\frac{1-\lambda^2}{12}$	$\frac{1}{6}$	$\lambda \leq \sqrt{3} - 1$
OPT	0.0492	0.228	$\lambda \leq 0.77$

The remaining special cases in Table 1 are implicit schemes, which use a 26-point update stencil (see Fig. 1). The maximally flat isotropic (MFI) scheme is the scheme for which the difference between axial and diagonal relative phase velocity is maximally flat. Such a scheme is particularly useful when the aim is to apply (off-line) pre- and post-warping techniques in order to remove as much as possible any direction-independent numerical dispersion, such as those presented in [9].

The implicit scheme with $a = \frac{1-\lambda^2}{12}$ and $b = \frac{1}{6}$, originally identified in [16], defines a subgroup of schemes within the familiy of Eq. (6) of fourth-order accuracy (FOA), as can be seen directly from Eq. (45).

The OPT scheme is an optimisation example, that is computationally most efficient under the criterion that the relative phase velocity error should not exceed 1% within any specified bandwidth (see Sec. 3.2).

3.1. Relative Phase Velocity

In Fig. 2, the relative phase velocity, as calculated with (30), is plotted as a function of the wavenumbers. What can be immediately observed from these plots is how 'isotropic' a particular scheme is. For example, the INT(1/6) scheme (results plotted for both $\lambda^2 = 0.75$, which is the stability bound, and $\lambda^2 = 0.5$, which coincides with the interpolated digital waveguide mesh) and the MFI scheme are far more 'round' than the other special cases. Note that the MFI scheme is the most isotropic of the three.

What is difficult - if not impossible - to determine from these plots is the numerical dispersion at a certain frequency. In some studies [11, 17], a circle is drawn on plots of this type with the intention of indicating the "highest normalised temporal frequency". Unfortunately this does not work, since the wavenumbers (spatial frequencies) are not proportional to the temporal frequencies (this follows straight from the dispersion relation).

It is however possible to convert a set of wavenumbers $(\hat{k}_x, \hat{k}_y)$ to frequency ω (using Eq. (29)) and angle of propagation θ. This way 'polar plots' of the relative phase velocity can be generated, where the polar radius corresponds to (normalised) frequency and the polar angle indicates the direction (see Fig. 3). Now we may draw circles on the plot in order to indicate particular frequencies. These plots also immediately reveal the cut-off frequency for any direction (beyond which waves in the discrete system are evanescent and dispersion is of highly reduced relevance). As can be seen, only the INT(1/4) scheme fills the complete bandwidth up to Nyquist for all directions, while the SLF, the RLF, and the INT(1/6) with $\lambda^2 = 0.5$ have the most 'severe' cut-off frequencies.

It can also be observed from Fig. 3 that the FOA and the OPT scheme have the largest bandwidth in which the numerical dispersion remains relatively small. Amongst the explicit schemes, the INT(1/6) (with $\lambda^2 = 0.75$) and the INT(1/4) scheme perform relatively well in that regard.

A general observation that can be made is that the highest numerical dispersion consistently occurs in either the axial or the diagonal direction. Hence perhaps the most useful and informative is to plot the relative phase velocity only for these directions, as depicted in Fig. 4. This allows better to observe the exact value of

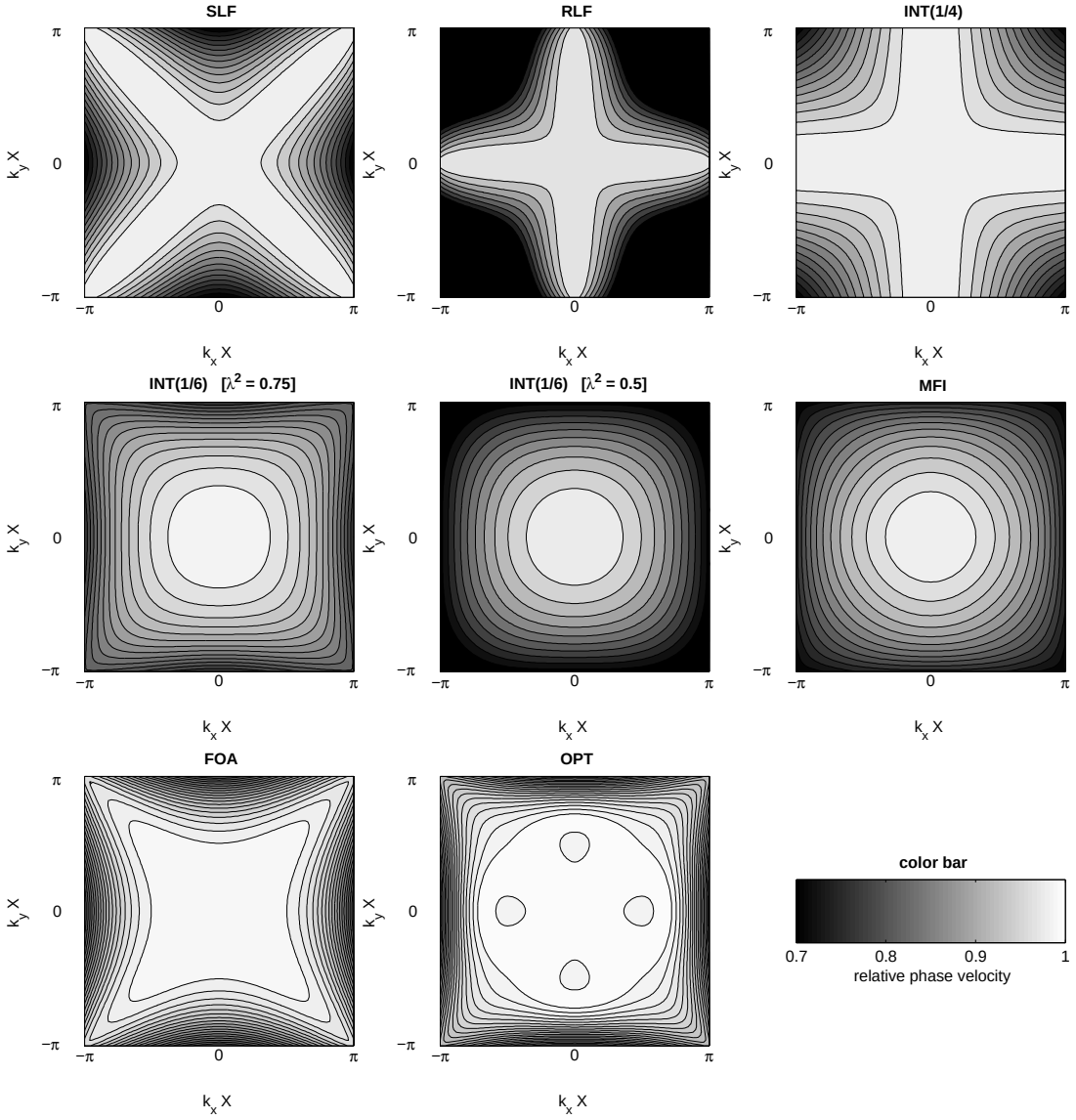

Figure 2: *Relative phase velocity as a function of x- and y-direction wavenumbers. The darkness in the plots indicates the relative phase velocity error, where white indicates a zero error, and any error larger then or equal to* 0.3 *is represented with black.*

the relative phase velocity, and for example reveals more clearly that v_a exceeds unity for the OPT scheme.

3.2. Computational Efficiency

While the relative phase velocity is a useful indicator of the characteristics of a scheme, the plots in Figs. 2, 3, and 4 do not allow a direct interpretation of how efficient they are in comparison. This is mainly because they use different Courant numbers. While some insight can be gained from comparing schemes for equal Courant numbers (see, for example, [5]), it seems more informative to restrict any comparisons to schemes at their highest possible value of the Courant number (provided that $\lambda \leq 1$), since that is when the numerical dispersion is generally the lowest. In order to investigate the dependence of computational efficiency on the free parameters (a, b) (with λ chosen at the stability bound), the effect of different λ values has to be taken into account.

In an audio context, one often requires precise results over a specified bandwidth. For example, in room acoustics applications the dispersion error should ideally be kept low up to the highest available frequency, as otherwise compact wavefronts are smeared out over time, which is likely to affect the capability of the listener to localise sources. Thus a suitable metric for computational costs defined as the amount of computation required in order to obtain a certain accuracy over a certain bandwidth[1]. For the purpose of

[1] For membrane synthesis, it might make more sense to use a different metric that allows the error to increase with frequency, since dispersion at high frequencies are not that relevant in that scenario. However, the conclusions regarding explicit schemes are actually generally applicable,

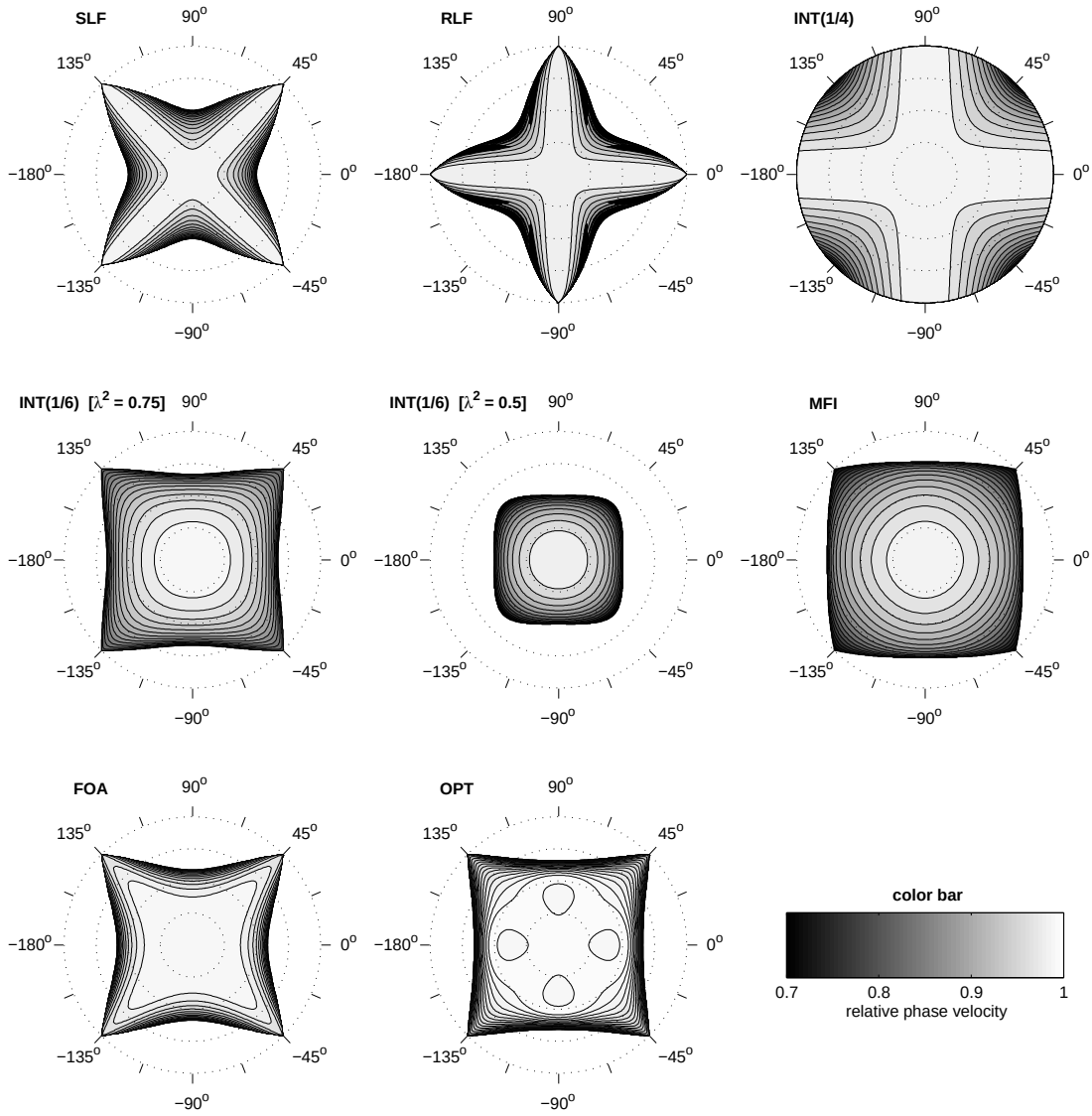

Figure 3: *Relative phase velocity as a function of frequency (polar plot radius) and propagation angle (polar plot angle). In each plot, starting from the most inner circle, the dotted-line circles indicate* $f = (\frac{1}{8}, \frac{1}{4}, \frac{3}{8}, \frac{1}{2})f_s$.

comparing the schemes in that sense, let's specify this as

> *the computational density required such that the deviation of the relative phase velocity from its ideal unity value is not larger than a critical error* e_c *up to a critical frequency* ω_c.

As a first indication, we may define the computational density simply as the number of nodal updates per square meter per second. This enables direct comparison between schemes that use exactly the same nodal update computations, such as all implicit schemes. Using this definition of computational density, we can define the following metric of *relative efficiency*:

$$\varepsilon(a, b, e_c) = \frac{\rho_{nu}(0, 0, e_c, \omega_c)}{\rho_{nu}(a, b, e_c, \omega_c)}, \tag{46}$$

where $\rho_{nu}(a, b, e_c, \omega_c)$ denotes the computational density for scheme (a, b) that meets the criterion (e_c, ω_c) and $\rho_{nu}(0, 0, e_c, \omega_c)$ is the reference (SLF) scheme that we normalise by. Note that while ρ_{nu} depends on the critical frequency, ε does not.

The value of ρ_{nu} is calculated by first determining the sampling frequency required to meet the accuracy criterion, for which Eqs. (33) and (37) can be used. This involves first determining the (normalised) frequencies $\omega_a T$ and $\omega_d T$ at which $|1 - v_a|$ and $|1 - v_d|$ are equal to e_c. Because (33) and (37) are not directly invertable, this has to be done using optimisation methods. Once these critical frequencies have been found, the required sample rate

and although we have not explictly addressed alternative metrics, some tentative conclusions regarding implicit schemes could also be made for the separate case of membrane synthesis on basis of the analysis presented.

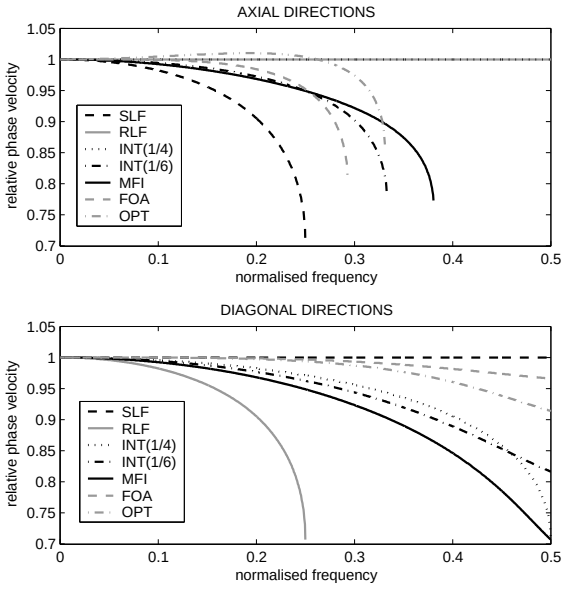

Figure 4: *Relative phase velocty for axial directions (top), and diagonal directions (bottom).*

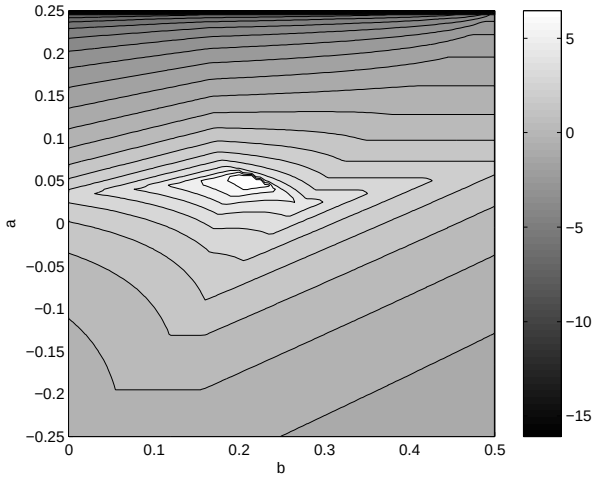

Figure 5: *Relative efficiency (plotted as* $\log_2[\varepsilon(a,b,e_c)]$*) for* $e_c = 0.01$*. The optimal scheme lies in the white area. For any combination of the free parameters* (a,b)*, the Courant number was chosen at the stability bound.*

is calculated as $f_s = \frac{\omega_c}{\min(\omega_a T, \omega_d T)}$. The required spatial step X is then obtained directly from $\lambda = (cT)/X$. The required nodal density is thus $\rho_n = X^{-2}$ nodes per m^2, and the required computational density (as in the number of nodal updates per m^2 per second) is $\rho_{nu} = \frac{f_s}{(X)^2}$.

Calculations of $\varepsilon(a,b,e_c)$ have been performed for a large set (a,b) and a range of e_c values. An example result ($e_c = 1\%$) is plotted in Fig. 5.

Table 2: *Relative Efficiency* $\varepsilon(e_c)$.

scheme	$\varepsilon(10\%)$	$\varepsilon(1\%)$	$\varepsilon(0.1\%)$	$\varepsilon(0.01\%)$
SLF	1.00	1.00	1.00	1.00
RLF	0.50	0.50	0.50	0.50
INT(1/4)	4.00	4.00	4.00	4.00
INT(1/6)	2.16	3.23	3.44	3.46
MFI	2.08	1.65	1.59	1.58
FOA	2.42	12.3	71.4	408
OPT	3.50	41.9	5.10	4.91

Table 2 shows the relative efficiency of the special cases at their top Courant number for $e_c = 10\%$, 1%, 0.1%, and 0.01%. Note that one cannot directly compare values of different columns. Direct comparison is also not possible between schemes that use different amounts of nodes in their stencil, since this is not taken into account in the calculations. However, some important conclusions regarding such comparisons can still be drawn without getting into the messy business of counting FLOPS:

- A nodal update with an interpolated scheme requires only a fraction of additional arithmetics in comparison with the SLF or the RLF scheme [10]. Hence it follows that the INT(1/4) is nearly 4 times more efficient than the SLF for any possible value of e_c.
- Inspection of results along the line $a = 0$ reveals the remarkable notion that amongst the explicit schemes, the INT(1/4) is always the optimally efficient scheme, regardless of the choice of e_c.
- Regarding implicit schemes (which all require exactly the same computational costs per nodal update for a given ADI implementation), the FOA scheme is the most efficient option, except when the cricital error is high, in which case a significant better scheme can be found via optimisation.
- As for comparing explicit schemes with implicit schemes, the computational costs of the ADI method per nodal update is roughly twice as much as that of explicit schemes [10]. Therefore it can be concluded that an implicit scheme that is more efficient than any explicit scheme can almost always be found, except if the critical error is chosen very high ($e_c > 2\%$).

Additional insight can be gained by marking the positions of the various schemes on a map of the (a,b) domain (see Fig. 6). The solid line indicates the positions of optimised schemes when varying the critical error from $e_c = 10\%$ to $e_c = 0.01\%$. The FOA scheme can be seen as the optimal scheme for $e_c \to 0$, and has the 'maximally flat' phase velocity error. Another interesting point is that the FOA scheme with λ at its top value uses $a = -\frac{1}{4} + \frac{1}{2\sqrt{3}}$, thus is the 'mirror-image' of the MFI scheme in the (a,b) domain.

4. CONCLUSIONS

Compact FDTD schemes describe a useful group of finite difference methods and related techniques for simulation of acoustic systems governed by the wave equation. While considerable

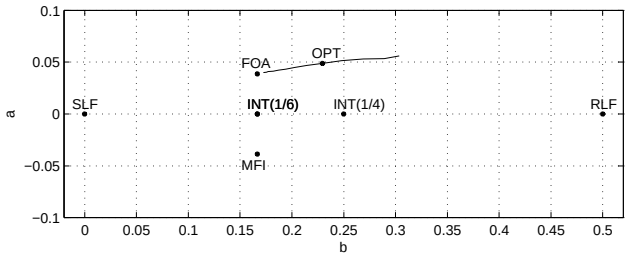

Figure 6: *Positions of schemes in the (a, b) domain. The solid line indicates the positions of schemes found by optimising efficiency under the criterion of a specified bandwidth in which the phase velocity error does not exceed a crital error.*

ground on this topic had been covered previously (in particular in [5, 18]), this paper has added and highlighted the following insights:

A useful subgroup of all possible compact schemes was chosen, that allows a particularly efficient ADI implementation (which appears to fall outside the 'ADI region' defined in [5]).

New ways of formulating and plotting the relative phase velocity were presented, such that its dependence on frequency can be interpreted immediately; these methods also allow direct assessment of the 'cut-off' frequency for any direction, which is particularly important in a digital audio context.

A new scheme (MFI) was identified that is more isotropic than the interpolated waveguide mesh, thus providing a useful alternative for applications that involve frequency warping.

A way of comparing schemes in terms of computational efficiency was formulated, using nodal update density as the main metric, and taking into account the variation of λ amongst the schemes. Using this efficiency formulation, and applying a specific accuracy criterion that is logical in a digital audio context, it has been shown that implicit schemes are generally significantly more efficient than explicit schemes, except when a weak accuracy criterion is applied.

It was also shown that amongst explicit schemes, the interpolated scheme with $b = \frac{1}{4}$ is the optimally efficient scheme, and about four times more efficient than the standard leapfrog scheme (which is equivalent to the rectilinear digital waveguide mesh). It follows that, unless the aim is to apply frequency warping techniques in simulation scenarios with static sources and receivers, this should always be the choice of scheme if explicitness is necessary or desirable.

5. REFERENCES

[1] J.C. Strikwerda, *Finite difference schemes and partial differential equations*, Wadsworth & Brooks, Pacific Grove, CA, 1989.

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

[3] L. Hiller and P. Ruiz, "Synthesizing musical sounds by solving the wave equation for vibrating objects: Parts I and II," *J. Audio Eng. Soc.*, vol. 19, no. 7, pp. 542–551, 1971.

[4] D. Botteldooren, "Finite-difference time-domain simulation of low-frequency room acoustic problems," *J. Acoustical Society America*, vol. 98, no. 6, pp. 3302–3308, 1995.

[5] S. Bilbao, "Parameterized families of finite difference schemes for the wave equation," *Numerical Methods for Partial Differential Equations*, vol. 20, no. 3, pp. 463–480, 2004, Wiley Periodicals, Inc.

[6] A. Fettweiss, "Multidimensional wave digital filters for discrete-time modelling of maxwell's equations," *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, vol. 5, no. 3, pp. 183–201, 1992.

[7] S. Bilbao, *Wave and scattering methods for numerical simulation*, John Wiley & Sons, London, 2004.

[8] S.van Duyne and J.O. Smith III, "The 2-D digital waveguide mesh," *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, October 1993, Mohonk, NY.

[9] L. Savioja and V. Välimäki, "Reducing the dispersion error in the digital waveguide mesh using interpolation and frequency warping techniques," *IEEE Trans. Speech and Audio Processing*, vol. 8, pp. 184–194, March 2000.

[10] K. Kowalczyk and M. van Walstijn, "On-line simulation of 2D resonators with reduced dispersion error using compact implicit finite difference methods," *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, vol. 1, pp. 285–288, April 2007, Honolulu, Hawaii.

[11] V. Välimäki and L. Savioja, "Interpolated and warped 2D digital waveguide mesh algorithms," *Proc. of the COST G-6 Conference on Digital Audio Effects (DAFx-00), Verona, Italy*, pp. 201–206, December 2000.

[12] A.R. Gourlay and A.R. Mitchell, "A classification of split methods for hyperbolic equations in several space dimentions," *SIAM Journal on Numerical Analysis*, vol. 6, no. 1, pp. 62–71, March 1969.

[13] J. B. Schneider and R. J. Kruhlak, "Dispersion of homogeneous and inhomogeneous waves in the yee finite-difference time-domain grid," *IEEE Trans. Microwave Theory and Techniques*, vol. 49, no. 2, pp. 280–287, 2001.

[14] R.F. Warming and B.J. Hyett, "The modified equation approach to the stability and accuracy analysis of finite-difference methods," *J. Comput. Phys*, , no. 14, pp. 159–179, 1974.

[15] K. S. Yee, "Numerical solution of initial boundary value problems involving maxwells equations in isotropic media," *IEEE Trans. Antennas Propagat.*, vol. 14, pp. 302–307, 1966.

[16] G. Fairweather and A.R. Mitchell, "A high accuracy alternating direction implicit method for the wave equation," *J. Inst. Math. Appl.*, vol. 1, pp. 309–316, May 1965.

[17] F. Fontana and D. Rocchesso, "Signal-theoretic characterization of waveguide mesh geometries for models of two-dimensional wave propagation in elastic media," *IEEE Trans. on Speech and Audio Processing*, vol. 9, no. 2, pp. 152–161, 2001.

[18] S. Bilbao, "Finite difference schemes and digital waveguide networks for the wave equation: Stability, passivity, and numerical dispersion," *IEEE Trans. on Speech and Audio Processing*, vol. 11, no. 3, pp. 255–266, 2003.