Journal of Computational Physics **165,** 189–215 (2000)
doi:10.1006/jcph.2000.6610, available online at http://www.idealibrary.com on IDEAL®

Application of the Space–Time Conservation Element and Solution Element Method to One-Dimensional Convection–Diffusion Problems

Sin-Chung Chang,* Xiao-Yen Wang,† and Wai-Ming To‡

*NASA Glenn Research Center, Cleveland, Ohio 44135; †Taitech Inc., at NASA Glenn Research Center, Cleveland, Ohio 44135; and ‡AP Solution, Inc., at NASA Glenn Research Center, Cleveland, Ohio 44135
E-mail: sin-chung.chang@grc.nasa.gov, wangxy@turbot.grc.nasa.gov, WaiMing.To@grc.nasa.gov
Web site: http://www.grc.nasa.gov/www/microbus

Received March 2, 1999; revised August 10, 2000

In the space–time conservation element and solution element (CE/SE) method, the independent marching variables used comprise not only the mesh values of the physical dependent variables but also, in contrast to a typical numerical method, the mesh values of the spatial derivatives of these physical variables. The use of the extra marching variables results from the need to construct the two-level, explicit and nondissipative schemes which are at the core of the CE/SE development. It also results from the need to minimize the stencil while maintaining accuracy. In this paper, using the 1D a–μ scheme as an example, the effect of this added complication on consistency, accuracy, and operation count is assessed. As part of this effort, an equivalent yet more efficient form of the a–μ scheme in which the independent marching variables are the local fluxes tied to each mesh point is introduced. Also, the intriguing relations that exist among the a–μ. Leapfrog, and DuFort–Frankel schemes are further explored. In addition, the redundancy of the Leapfrog, DuFort–Frankel, and Lax schemes and the remedy for this redundancy are discussed. This paper is concluded with the construction and evaluation of a CE/SE solver for the inviscid Burgers equation.

Key Words: space–time; flux conservation; conservation element; solution element; shocks.

1. INTRODUCTION

The space–time conservation element and solution element (CE/SE) method is a new high-resolution, genuinely multidimensional, and unstructured-mesh compatible numerical method for solving conservation laws [1–21]. Since its inception in 1991 [1], the CE/SE

0021-9991/00 $35.00

method has been used to obtain highly accurate numerical solutions for 1D, 2D, and 3D inviscid and viscous flow problems involving shocks, contact discontinuities, vortices, acoustic waves, boundary layers, chemical reactions, and hydraulic jump. Without the aid of preconditioning or other special techniques, the method can be applied to both steady and unsteady flows with speeds ranging from Mach number = 0.00288 to 10 [16].

Development of the CE/SE method is motivated by a desire to build a general and coherent numerical framework that avoids the limitations and complications of the traditional methods. As a result, the CE/SE method was built from ground zero using a set of design principles [2, 3] that facilitate simplicity, robustness, and accuracy. They include: (i) enforcing both local and global flux conservation in space and time, with flux evaluation at an interface being an integral part of the solution procedure and requiring no interpolation or extrapolation; (ii) unifying space and time and treating them as a single entity; (iii) requiring that a numerical scheme be built from a nondissipative core scheme such that the numerical dissipation can be effectively controlled and, as a result, will not overwhelm the physical dissipation; (iv) considering the mesh values of the physical dependent variables and their spatial derivatives as independent marching variables, to be solved for simultaneously; (v) defining conservation elements and solution elements such that the simplest stencil will result; (vi) excluding the use of characteristics-based techniques (such as Riemann solvers); and (vii) avoiding the use of ad hoc techniques as much as possible.

Note that the a–μ scheme and the a–ϵ scheme, which are, respectively, the CE/SE solvers of a convection–diffusion equation (see Eq. (2.1)) and its pure convection version, were described in [2] along with their Euler and Navier–Stokes extensions. However, because of the need to minimize the length of the manuscript, several important topics such as the consistency, accuracy, and operation count of the a–μ scheme were not addressed in Chang [2]. As will be shown, because of the fact that the mesh values of both the dependent variables and its spatial derivatives are treated as independent marching variables in the CE/SE development, the concept of consistency for the a–μ scheme is by no means trivial. In addition to addressing several topics left untreated in Chang [2], in this paper we will describe an equivalent yet more efficient form of the a–μ scheme in which the independent marching variables are the local fluxes tied to each mesh point. Furthermore, a CE/SE solver for the inviscid Burgers equation will also be introduced and evaluated.

As a side trip, the intriguing relations [2] that exist among the a–μ, Leapfrog, and DuFort–Frankel schemes will be further explored in this paper. In addition, the redundancy of the classical Leapfrog, DuFort–Frankel, and Lax schemes, and the remedy to this redundancy, will also be discussed.

2. NUMERICAL SCHEMES

In this section, we shall (i) briefly review the 1D CE/SE a–μ scheme described in [2] and then recast it in a numerically more efficient form; and (ii) describe a new CE/SE scheme for solving the inviscid Burgers equation.

2.1. *The a–μ Scheme*

Consider a dimensionless form of the 1-D convection–diffusion equation, i.e.,

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} - \mu\frac{\partial^2 u}{\partial x^2} = 0, \tag{2.1}$$

where a and μ (≥ 0) are constants. Let $x_1 = x$ and $x_2 = t$ be considered the coordinates of a two-dimensional Euclidean space E_2. By using Gauss' divergence theorem in the space–time E_2, it can be shown that Eq. (2.1) is the differential form of the integral conservation law

$$\oint_{S(V)} \mathbf{h} \cdot d\mathbf{s} = 0. \tag{2.2}$$

Here (i) $S(V)$ is the boundary of an arbitrary space–time region V in E_2, (ii) $\mathbf{h} = (au - \mu\partial u/\partial x, u)$ is a current density vector in E_2, and (iii) $d\mathbf{s} = d\sigma\mathbf{n}$ with $d\sigma$ and $\mathbf{n}$, respectively, is the area and the outward unit normal of a surface element on $S(V)$. Note that (i) $\mathbf{h} \cdot d\mathbf{s}$ is the space–time flux of $\mathbf{h}$ leaving the region V through the surface element $d\mathbf{s}$, and (ii) all mathematical operations can be carried out as though E_2 were an ordinary two-dimensional Euclidean space.

Let Ψ denote the set of all mesh points in E_2 (dots in Fig. 1a). There is a solution element (SE) associated with each $(j, n) \in \Psi$. Let the solution element SE(j, n) be the *interior* of the *space–time* region bounded by a dashed curve depicted in Fig. 1b. It includes a horizontal line segment, a vertical line segment, and their immediate neighborhood.

For any $(x, t) \in$ SE(j, n), $u(x, t)$ and $\mathbf{h}(x, t)$ are approximated by $\mathbf{u}^*(x, t; j, n)$ and $\mathbf{h}^*(x, t; j, n)$, respectively. Here

$$u^*(x, t; j, n) = u_j^n + (u_x)_j^n(x - x_j) + (u_t)_j^n(t - t^n) \tag{2.3}$$

and

$$\mathbf{h}^*(x, t; j, n) = (au^*(x, t; j, n) - \mu\partial u^*(x, t; j, n)/\partial x,\ u^*(x, t; j, n)). \tag{2.4}$$

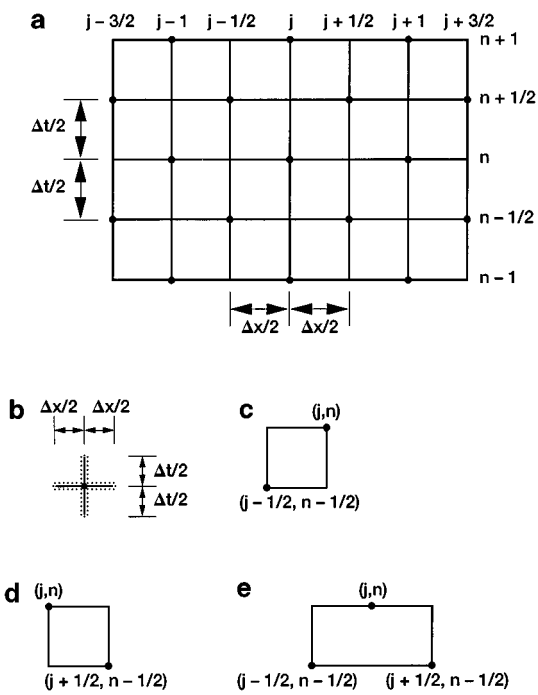

FIG. 1. The space–time mesh, CEs, and SEs used in the CE/SE method. (a) The staggered space–time mesh. (b) SE (j, n). (c) CE$_-(j, n)$. (d) CE$_+(j, n)$. (e) CE(j, n) at an interior mesh point (j, n).

Note that here that (i) u_j^n, $(u_x)_j^n$, and $(u_t)_j^n$ are constants in SE(j, n), (ii) (x_j, t^n) are the coordinates of the mesh point (j, n), and (iii) Eq. (2.4) is the numerical analogue of the definition $\mathbf{h} = (au - \mu\partial u/\partial x, u)$.

Let $u = u^*(x, t; j, n)$ satisfy Eq. (2.1) (i.e., $\nabla \cdot \mathbf{h}^* = 0$) within SE$(j, n)$. Then one has

$$(u_t)_j^n = -a(u_x)_j^n. \tag{2.5}$$

Note that because Eq. (2.3) is a first-order Taylor's expansion, the diffusion term in Eq. (2.1) has no counterpart in Eq. (2.5). As a result, the diffusion term has no impact on how $u^*(x, t; j, n)$ varies with time *within* SE(j, n). However, as will be shown shortly, through its role in the numerical analogue of Eq. (2.2), it does influence time-dependence of numerical solutions. Moreover, the legitimacy of Eq. (2.5) is supported by the results of stability and consistency study of the a–μ scheme given in [1, 2] and Section 3 of this paper.

Combining Eqs. (2.3) and (2.5), one has

$$u^*(x, t; j, n) = u_j^n + (u_x)_j^n \left[(x - x_j) - a(t - t^n)\right], \quad (x, t) \in \text{SE}(j, n). \tag{2.6}$$

Thus, u_j^n and $(u_x)_j^n$ are the only independent marching variables associated with the mesh point (j, n).

Let E_2 be divided into nonoverlapping rectangular regions (see Fig. 1a) referred to as conservation elements (CEs). As depicted in Figs. 1c and 1d, two nonoverlapping CEs, i.e., CE$_-(j, n)$ and CE$_+(j, n)$, are associated with each interior mesh point $(j, n) \in \Psi$. On the other hand, a single CE, i.e., CE$_-(j, n)$ (CE$_+(j, n)$), is associated with a mesh point $(j, n) \in \Psi$ on the right (left) spatial boundary. The conservation element CE(j, n) (see Fig. 1e), which will be used only in Section 2.2, is the union of CE$_-(j, n)$ and CE$_+(j, n)$. Obviously, the boundary of CE$_-(j, n)$ is formed by subsets of SE(j, n) and SE$(j - 1/2, n - 1/2)$, while that of CE$_+(j, n)$ is formed by subsets of SE(j, n) and SE$(j + 1/2, n - 1/2)$. By assuming

$$\oint_{S(V)} \mathbf{h}^* \cdot d\mathbf{s} = 0 \tag{2.7}$$

with $V = \text{CE}_+(j, n)$ and $V = \text{CE}_-(j, n)$, respectively, one obtains two conservation conditions at each mesh point $(j, n) \in \Psi$. Using these two conditions along with Eqs. (2.4) and (2.6), the two independent marching variables u_j^n and $(u_x)_j^n$ can be expressed as the functions of the independent marching variables at the mesh points $(j \pm 1/2, n - 1/2)$ [2, pp. 299–300]; i.e.,

$$u_j^n = \frac{1}{2}\left\{(1+\nu)u_{j-1/2}^{n-1/2} + (1-\nu)u_{j+1/2}^{n-1/2} + (1-\nu^2-\xi)\left[(u_x^+)_{j-1/2}^{n-1/2} - (u_x^+)_{j+1/2}^{n-1/2}\right]\right\} \tag{2.8}$$

and

$$\begin{aligned}(u_x^+)_j^n = \frac{-1}{2(1-\nu^2+\xi)}\Big\{&(1-\nu^2)\left(u_{j-1/2}^{n-1/2} - u_{j+1/2}^{n-1/2}\right) \\ &+ (1-\nu^2-\xi)\left[(1-\nu)(u_x^+)_{j-1/2}^{n-1/2} + (1+\nu)(u_x^+)_{j+1/2}^{n-1/2}\right]\Big\}.\end{aligned} \tag{2.9}$$

Here

$$1 - \nu^2 + \xi \neq 0 \tag{2.10}$$

and

$$\nu \stackrel{\text{def}}{=} \frac{a\Delta t}{\Delta x}, \quad \xi \stackrel{\text{def}}{=} \frac{4\mu\Delta t}{(\Delta x)^2} \quad \text{and} \quad (u_x^+)_j^n \stackrel{\text{def}}{=} \frac{\Delta x}{4}(u_x)_j^n. \tag{2.11}$$

The a–μ scheme is formed by Eqs. (2.8) and (2.9). Note that it is explained in [2] that the local conservation conditions used to construct the a–μ scheme lead to a global conservation relation (i.e., the total flux leaving the boundary of any space–time region that is the union of any combination of CEs will also vanish).

Also note that the expression on the right side of Eq. (2.8) can be written as a linear combination of the four marching variables $u_{j\pm1/2}^{n-1/2}$ and $(u_x^+)_{j\pm1/2}^{n-1/2}$. Each combination coefficient is a constant which can be evaluated once and used repeatedly in the marching procedure. Because a pair of these coefficients differ only in sign, one concludes that it requires three multiplications, two additions, and one subtraction to evaluate u_j^n. Applying the same argument to Eq. (2.9), one concludes that it requires six multiplications, four additions, and two subtractions to evaluate both u_j^n and $(u_x^+)_j^n$ for each $(j, n) \in \Psi$.

In [2], it is shown that Eqs. (2.8) and (2.9) can be derived from a perspective different from that shown above. In the following, an equivalent but numerically more efficient and physically more revealing form of the a–μ scheme will be derived from the new perspective. Note that this "new" form is really the original form of the a–μ scheme given in [1].

In the new derivation, the locations of mesh points (dots in Fig. 2a) are identical to those shown in Fig. 1a. However, the solution element (denoted by $\text{SE}'(j, n)$) associated with any $(j, n) \in \Psi$ is defined to be the *interior* of a rhombus centered at (j, n) (see Fig. 2b). On the other hand, the conservation element (denoted by $\text{CE}'(j, n)$) associated with (j, n) is defined to be the union of $\text{SE}'(j, n)$ and its boundary (see Fig. 2c). Note that a side of the rhombus is in general *not* a characteristic line of Eq. (2.1). It is simply a line segment joining two points of intersection (not marked by dots) of horizontal and vertical mesh lines. For any $(x, t) \in \text{SE}'(j, n)$, $u(x, t)$ and $\mathbf{h}(x, t)$, respectively, again are approximated by $u^*(x, t; j, n)$ and $\mathbf{h}^*(x, t; j, n)$, which are defined by Eqs. (2.3) and (2.4) respectively.

Furthermore, Eq. (2.2) is approximated by

$$\oint_{S(V^*)} \mathbf{h}^* \cdot d\mathbf{s} = 0, \tag{2.12}$$

where V^* is the union of any combination of CEs. Because an SE is the interior of a CE, $\mathbf{h}^*$ is not defined on $S(V^*)$, the boundary of V^*. As a result, the above surface integration is to be carried out over a surface that is in the interior of V^* and immediately adjacent to $S(V^*)$. A necessary condition of Eq. (2.12) is that, for any (j, n),

$$\oint_{S(\text{CE}'(j,n))} \mathbf{h}^* \cdot d\mathbf{s} = 0; \tag{2.13}$$

i.e., the total flux leaving any conservation element is zero. It is shown in [2] that, given Eqs. (2.3) and (2.4), Eq. (2.13) is equivalent to Eq. (2.5). As a result, Eqs. (2.5) and (2.6) can be assumed in the following derivation.

By applying Eq. (2.12) separately to two neighboring CEs and then to their union, it is seen that Eq. (2.12) also requires that, at the interface separating any two neighboring CEs, the total flux entering the interface from one side must be equal to that leaving it from another side (Note: As a result of the definitions of the current CEs and SEs, $\mathbf{h}^*$ at the two

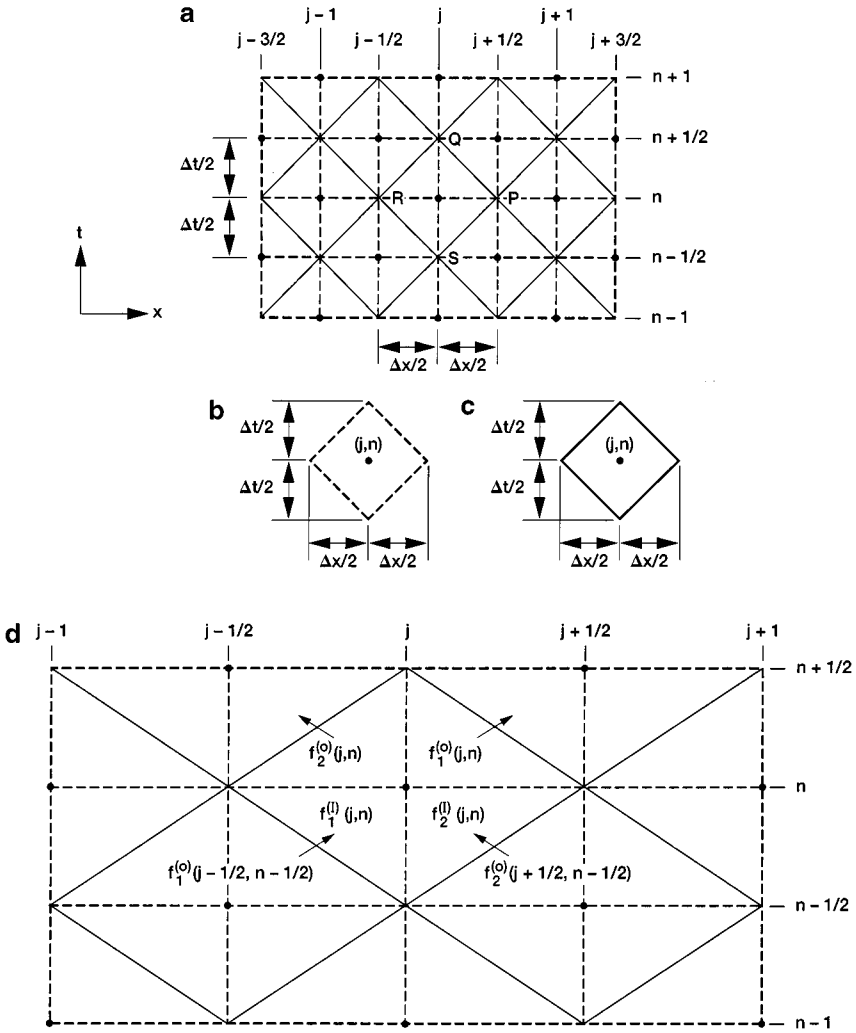

FIG. 2. The alternative SEs and CEs. (a) The relative positions of SEs and CEs. (b) $\mathrm{SE}'(j, n)$. (c) $\mathrm{CE}'(j, n)$. (d) Interface flux conservation relations.

sides of the interface are evaluated using information from two different SEs.) Obviously, the local flux conservation conditions at all interfaces and within all CEs (i.e., Eq. (2.13)) are equivalent to the global conservation condition Eq. (2.12).

To study the interface conditions, consider any $(j, n) \in \Psi$. Let $\diamond PQRS$ be the parallelogram depicted in Fig. 2a. Let

$$J(\overline{PQ}) \stackrel{\text{def}}{=} \int_{\overline{PQ}} \mathbf{h}^* \cdot d\mathbf{s},$$

where (i) $d\mathbf{s}$ points in the direction away from the interior of $\diamond PQRS$ and (ii) the integration is carried out over a line segment that is in the interior of $\diamond PQRS$ and immediately adjacent to $\overline{PQ}$. We define $J(\overline{QR})$, $J(\overline{RS})$, and $J(\overline{SP})$ similarly. With the aid of Eqs. (2.4) and (2.6),

it can be shown that

$$J(\overline{PQ}) = \frac{\Delta x}{2}\left[(1+\nu)u_j^n + (1-\nu^2-\xi)(u_x^+)_j^n\right] \tag{2.14}$$

$$J(\overline{QR}) = \frac{\Delta x}{2}\left[(1-\nu)u_j^n - (1-\nu^2-\xi)(u_x^+)_j^n\right] \tag{2.15}$$

$$J(\overline{RS}) = -\frac{\Delta x}{2}\left[(1+\nu)u_j^n - (1-\nu^2+\xi)(u_x^+)_j^n\right] \tag{2.16}$$

and

$$J(\overline{SP}) = -\frac{\Delta x}{2}\left[(1-\nu)u_j^n + (1-\nu^2+\xi)(u_x^+)_j^n\right]. \tag{2.17}$$

Note that Eqs. (2.14)–(2.17) are consistent with Eq. (2.13); i.e.,

$$J(\overline{PQ}) + J(\overline{QR}) + J(\overline{RS}) + J(\overline{SP}) = 0. \tag{2.18}$$

To proceed, let

$$f_1^{(\mathrm{O})}(j,n) \stackrel{\text{def}}{=} \frac{2}{\Delta x} J(\overline{PQ}), \quad f_2^{(\mathrm{O})}(j,n) \stackrel{\text{def}}{=} \frac{2}{\Delta x} J(\overline{QR}) \tag{2.19}$$

and

$$f_1^{(\mathrm{I})}(j,n) \stackrel{\text{def}}{=} -\frac{2}{\Delta x} J(\overline{RS}), \quad f_2^{(\mathrm{I})}(j,n) \stackrel{\text{def}}{=} -\frac{2}{\Delta x} J(\overline{SP}). \tag{2.20}$$

In other words, $f_1^{(\mathrm{O})}(j,n)$ and $f_2^{(\mathrm{O})}(j,n)$, respectively, are the *normalized* fluxes *leaving* $\mathrm{CE}'(j,n)$ through its "future right" and "future left" edges. Similarly, $f_1^{(\mathrm{I})}(j,n)$ and $f_2^{(\mathrm{I})}(j,n)$, respectively, are the *normalized* fluxes *entering* $\mathrm{CE}'(j,n)$ through its "past left" and "past right" edges. For simplicity, a normalized flux will be referred to simply as a flux. Thus, the two fluxes defined in Eq. (2.19) may be referred to as the outgoing fluxes while the two fluxes defined in Eq. (2.20) may be referred to as the incoming fluxes. Note that the interface flux conservation conditions referred to earlier can now be expressed as (see Fig. 2d): For any $(j,n) \in \Psi$,

$$f_1^{(\mathrm{I})}(j,n) = f_1^{(\mathrm{O})}(j-1/2, n-1/2) \quad \text{and} \quad f_2^{(\mathrm{I})}(j,n) = f_2^{(\mathrm{O})}(j+1/2, n-1/2). \tag{2.21}$$

Because of the above relations, in Fig. 2d, a single arrow is drawn across an interface to represent both the flux entering and the flux leaving this interface.

At this juncture, note that, with the aid of Eqs. (2.14)–(2.17), (2.19), and (2.20), Eqs. (2.8) and (2.9) can also be obtained using Eq. (2.21). In the following, Eq. (2.21) will be used to construct an alternative scheme in which $f_1^{(\mathrm{O})}(j,n)$ and $f_2^{(\mathrm{O})}(j,n)$ are the independent marching variables.

To proceed, let

$$\mathbf{f}^{(\mathrm{O})}(j,n) \stackrel{\text{def}}{=} \begin{pmatrix} f_1^{(\mathrm{O})}(j,n) \\ f_2^{(\mathrm{O})}(j,n) \end{pmatrix}, \quad \mathbf{f}^{(\mathrm{I})}(j,n) \stackrel{\text{def}}{=} \begin{pmatrix} f_1^{(\mathrm{I})}(j,n) \\ f_2^{(\mathrm{I})}(j,n) \end{pmatrix} \tag{2.22}$$

$$\mathbf{q}(j,n) \stackrel{\text{def}}{=} \begin{pmatrix} u_j^n \\ (u_x^+)_j^n \end{pmatrix} \tag{2.23}$$

$$\Lambda^{(\mathrm{O})} \stackrel{\text{def}}{=} \begin{pmatrix} 1+\nu & 1-\nu^2-\xi \\ 1-\nu & -(1-\nu^2-\xi) \end{pmatrix} \tag{2.24}$$

and

$$\Lambda^{(\mathrm{I})} \stackrel{\text{def}}{=} \begin{pmatrix} 1+\nu & -(1-\nu^2+\xi) \\ 1-\nu & 1-\nu^2+\xi \end{pmatrix}. \tag{2.25}$$

Then, with the aid of the above definitions and Eqs. (2.19) and (2.20), Eqs. (2.14)–(2.17) can be expressed as

$$\mathbf{f}^{(\mathrm{O})}(j,n) = \Lambda^{(\mathrm{O})}\mathbf{q}(j,n) \tag{2.26}$$

and

$$\mathbf{f}^{(\mathrm{I})}(j,n) = \Lambda^{(\mathrm{I})}\mathbf{q}(j,n). \tag{2.27}$$

Note that, as a result of Eq. (2.10), the inverse of $\Lambda^{(\mathrm{I})}$ exists; i.e.,

$$\left[\Lambda^{(\mathrm{I})}\right]^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -\frac{1-\nu}{1-\nu^2+\xi} & \frac{1+\nu}{1-\nu^2+\xi} \end{pmatrix}. \tag{2.28}$$

It follows from Eq. (2.27) that

$$\mathbf{q}(j,n) = \left[\Lambda^{(\mathrm{I})}\right]^{-1}\mathbf{f}^{(\mathrm{I})}(j,n). \tag{2.29}$$

Because the elements of the matrices $\Lambda^{(\mathrm{I})}$ and $[\Lambda^{(\mathrm{I})}]^{-1}$ are constant, Eqs. (2.27) and (2.29) imply that $f_1^{(\mathrm{I})}(j,n)$ and $f_2^{(\mathrm{I})}(j,n)$ can be uniquely determined in terms of u_j^n and $(u_x^+)_j^n$, and vice versa.

Substituting Eq. (2.29) into Eq. (2.26), one has

$$\mathbf{f}^{(\mathrm{O})}(j,n) = \Omega\mathbf{f}^{(\mathrm{I})}(j,n). \tag{2.30}$$

where

$$\Omega \stackrel{\text{def}}{=} \Lambda^{(\mathrm{O})}\left[\Lambda^{(\mathrm{I})}\right]^{-1}. \tag{2.31}$$

Equation (2.30) can be rewritten as

$$f_\ell^{(\mathrm{O})}(j,n) = \sum_{m=1}^{2} \omega_{\ell m} f_m^{(\mathrm{I})}(j,n), \quad \ell = 1,2. \tag{2.32}$$

Here $\omega_{\ell m}$, $\ell, m = 1, 2$ are the elements of the matrix Ω. By using Eqs. (2.24), (2.28), and (2.31), one has

$$\omega_{11} = \frac{\nu(1-\nu^2)+\xi}{1-\nu^2+\xi}, \quad \omega_{12} = \frac{(1+\nu)(1-\nu^2)}{1-\nu^2+\xi} \tag{2.33a}$$

and

$$\omega_{21} = \frac{(1-\nu)(1-\nu^2)}{1-\nu^2+\xi}, \quad \omega_{22} = \frac{-\nu(1-\nu^2)+\xi}{1-\nu^2+\xi}. \tag{2.33b}$$

A result of Eqs. (2.33a) and (2.33b) is

$$\sum_{\ell=1}^{2} \omega_{\ell m} = 1, \quad m = 1, 2. \tag{2.34}$$

It follows from Eqs. (2.32) and (2.34) that

$$f_1^{(\mathrm{O})}(j,n) + f_2^{(\mathrm{O})}(j,n) = f_1^{(\mathrm{I})}(j,n) + f_2^{(\mathrm{I})}(j,n), \tag{2.35}$$

i.e., the sum of the outgoing fluxes is equal to that of the incoming fluxes. From Eqs. (2.19) and (2.20), it is easy to see that Eq. (2.35) is equivalent to Eq. (2.18).

Next, by combining Eqs. (2.21) and (2.32), one obtains

$$f_1^{(\mathrm{O})}(j,n) = \omega_{11}\, f_1^{(\mathrm{O})}(j-1/2, n-1/2) + \omega_{12}\, f_2^{(\mathrm{O})}(j+1/2, n-1/2) \tag{2.36}$$

and

$$f_2^{(\mathrm{O})}(j,n) = \omega_{21}\, f_1^{(\mathrm{O})}(j-1/2, n-1/2) + \omega_{22}\, f_2^{(\mathrm{O})}(j+1/2, n-1/2). \tag{2.37}$$

Equations (2.36) and (2.37) form a marching scheme in which the outgoing fluxes $f_1^{(\mathrm{O})}(j,n)$ and $f_2^{(\mathrm{O})}(j,n)$ are evaluated in terms of the outgoing fluxes $f_1^{(\mathrm{O})}(j-1/2, n-1/2)$ and $f_2^{(\mathrm{O})}(j+1/2, n-1/2)$. This evaluation requires four multiplications and two additions. However, the operation count is reduced to two multiplications, two additions, and one subtraction if Eq. (2.37) in the above scheme is replaced by

$$f_2^{(\mathrm{O})}(j,n) = f_1^{(\mathrm{O})}(j-1/2, n-1/2) + f_2^{(\mathrm{O})}(j+1/2, n-1/2) - f_1^{(\mathrm{O})}(j,n), \tag{2.38}$$

which is a direct result of Eqs. (2.21) and (2.35). Note that, ignoring additions and subtractions (which can be performed much faster than multiplications), the current operation count (i.e., two multiplications) is only one third of that (i.e., six multiplications) associated with the scheme formed by Eqs. (2.8) and (2.9).

Given the above main marching scheme, a complete marching procedure can be defined using the following information:

(a) For any $(j,0) \in \Psi$, the outgoing fluxes $f_1^{(\mathrm{O})}(j,0)$ and $f_2^{(\mathrm{O})}(j,0)$ can be evaluated using Eq. (2.26) if the the initial data u_j^0 and $(u_x^+)_j^0$ are given.

(b) Let the initial data be periodic; i.e., for any mesh point $(j,0)$,

$$f_\ell^{(\mathrm{O})}(j+K,0) = f_\ell^{(\mathrm{O})}(j,0), \quad \ell = 1, 2 \tag{2.39}$$

where K is a given integer ≥ 1. Then, for any $(j,n) \in \Psi$ with $n \geq 0$, $f_\ell^{(\mathrm{O})}(j,n)$ can be determined in terms of the initial data by using the main marching scheme. Furthermore, by induction, it can be shown that the solution is periodic; i.e., for any $(j,n) \in \Psi$ with $n \geq 0$,

$$f_\ell^{(\mathrm{O})}(j+K,n) = f_\ell^{(\mathrm{O})}(j,n). \tag{2.40}$$

(c) By eliminating $(u_x^+)_j^n$ from Eqs. (2.14) and (2.17) and then using Eqs. (2.19)–(2.21), one has

$$f_1^{(\mathrm{O})}(j,n) = (1+\nu)u_j^n + \frac{1-\nu^2-\xi}{1-\nu^2+\xi}\left[f_2^{(\mathrm{O})}(j+1/2,n-1/2) - (1-\nu)u_j^n\right] \quad (2.41)$$

Similarly, Eqs. (2.15), (2.16), and (2.19)–(2.21) can be used to show that

$$f_2^{(\mathrm{O})}(j,n) = (1-\nu)u_j^n + \frac{1-\nu^2-\xi}{1-\nu^2+\xi}\left[f_1^{(\mathrm{O})}(j-1/2,n-1/2) - (1+\nu)u_j^n\right]. \quad (2.42)$$

According to Eq. (2.41), in the case where the spatial domain is finite and (j,n) is a mesh point at the left boundary, $f_1^{(\mathrm{O})}(j,n)$ can be evaluated in terms of $f_2^{(\mathrm{O})}(j+1/2,n-1/2)$ if the value of u_j^n is given. On the other hand, according to Eq. (2.42), $f_2^{(\mathrm{O})}(j,n)$ can be evaluated in terms of $f_1^{(\mathrm{O})}(j-1/2,n-1/2)$ for a right-boundary mesh point (j,n) if the value of u_j^n is given. Thus, one concludes that the marching can proceed through all time levels if the values of u_j^n are specified at all boundary mesh points.

This section concludes with the following remarks:

(a) The a–μ scheme has the simplest stencil, i.e., a triangle with a vertex at the upper time level and the other two vertices at the lower time level. Furthermore, the number of the independent marching variables associated with a mesh point $(j,n) \in \Psi$ is identical to the number of the mesh points at the $(n-1/2)$th time level that are part of the stencil. Note that the same relation also holds for many 2D and 3D CE/SE schemes [2–4, 18, 20].

(b) Let (j,n) be an interior mesh point. According to Eqs. (2.21), (2.22), and (2.29), u_j^n and $(u_x^+)_j^n$ can be determined in terms of the outgoing fluxes $f_1^{(\mathrm{O})}(j-1/2,n-1/2)$ and $f_2^{(\mathrm{O})}(j+1/2,n-1/2)$.

(c) Let (j,n) be a mesh point on the right boundary. According to Eqs. (2.16), (2.20), and (2.21), one has

$$\left[(1+\nu)u_j^n - (1-\nu^2-\xi)(u_x^+)_j^n\right] = f_1^{(\mathrm{O})}(j-1/2,n-1/2). \quad (2.43)$$

Thus, $(u_x^+)_j^n$ can be determined in terms of the outgoing flux $f_1^{(\mathrm{O})}(j-1/2,n-1/2)$ if the boundary value u_j^n is given. Similarly, for a mesh point (j,n) on the left boundary, $(u_x^+)_j^n$ can be determined in terms of $f_2^{(\mathrm{O})}(j+1/2,n-1/2)$ if u_j^n is given.

(d) As a preliminary for a discussion of the consistency of the a–μ scheme in Section 3, note that, by using Eqs. (2.8) and (2.9) repeatedly, one has [1, p. 20]

$$\begin{aligned} u_j^{n+1} = {} & \frac{1}{2}\left[\nu + \frac{\xi(1-\nu)}{1-\nu^2+\xi}\right]\left[(1+\nu)u_{j-1}^n + (1-\nu^2-\xi)(u_x^+)_{j-1}^n\right] \\ & + \frac{1-\nu^2}{1-\nu^2+\xi}\left[(1-\nu^2)u_j^n - \nu(1-\nu^2-\xi)(u_x^+)_j^n\right] \\ & - \frac{1}{2}\left[\nu - \frac{\xi(1+\nu)}{1-\nu^2+\xi}\right]\left[(1-\nu)u_{j+1}^n - (1-\nu^2-\xi)(u_x^+)_{j+1}^n\right] \end{aligned} \quad (2.44)$$

and

$$(u_x^+)_j^{n+1} = -\frac{1}{2}\left[\nu + \frac{\xi(1-\nu)}{1-\nu^2+\xi}\right]\left[\frac{1-\nu^2}{1-\nu^2+\xi}u_{j-1}^n + (1-\nu)\frac{1-\nu^2-\xi}{1-\nu^2+\xi}(u_x^+)_{j-1}^n\right]$$
$$+ \left[\frac{1-\nu^2}{1-\nu^2+\xi}\right]^2\left[\nu u_j^n + (1-\nu^2-\xi)(u_x^+)_j^n\right] - \frac{1}{2}\left[\nu - \frac{\xi(1+\nu)}{1-\nu^2+\xi}\right]$$
$$\times \left[\frac{1-\nu^2}{1-\nu^2+\xi}u_{j+1}^n - (1+\nu)\frac{1-\nu^2-\xi}{1-\nu^2+\xi}(u_x^+)_{j+1}^n\right]. \tag{2.45}$$

(e) Note that the results presented here are only the special case $b = 0$ of those presented in [1], where b (a constant) is the speed of the moving mesh considered there. Also, according to a Fourier error analysis given in Section 5 of [1], for a given Δx, the accuracy of the a–μ scheme ($\mu > 0$) will reach a peak if Δt is chosen such that

$$1 - \nu^2 = \sqrt{3}\xi. \tag{2.46}$$

The potency of this analytical result will be numerically demonstrated in Section 4.2.

2.2. *The Inviscid Burgers a–ϵ–α–β Scheme*

Consider the inviscid Burgers equation; i.e.,

$$\frac{\partial u}{\partial t} + \frac{\partial g}{\partial x} = 0 \quad (g \stackrel{\text{def}}{=} u^2/2). \tag{2.47}$$

The integral conservation form of Eq. (2.47) is Eq. (2.2) with

$$\mathbf{h} = (g, u). \tag{2.48}$$

In the current scheme, for any $(x, t) \in \text{SE}(j, n)$, $u(x, t)$, $g(x, t)$, and $\mathbf{h}(x, t)$ are approximated by $u^*(x, t; j, n)$, $g^*(x, t; j, n)$ and $\mathbf{h}^*(x, t; j, n)$, respectively. Here (i) $u^*(x, t; j, n)$ is defined by Eq. (2.3); (ii)

$$g^*(x, t; j, n) = g_j^n + (g_x)_j^n(x - x_j) + (g_t)_j^n(t - t^n) \tag{2.49}$$

and (iii)

$$\mathbf{h}^*(x, t; j, n) = (g^*(x, t; j, n),\ u^*(x, t; j, n)). \tag{2.50}$$

Note that, in Eq. (2.49),

$$g_j^n \stackrel{\text{def}}{=} \left(u_j^n\right)^2/2, \quad (g_x)_j^n \stackrel{\text{def}}{=} u_j^n(u_x)_j^n, \quad \text{and} \quad (g_t)_j^n \stackrel{\text{def}}{=} u_j^n(u_t)_j^n. \tag{2.51}$$

Obviously, the above expressions are the numerical analogues of the analytical expressions $g = u^2/2$, $\partial g/\partial x = u\partial u/\partial x$, and $\partial g/\partial t = u\partial u/\partial t$, respectively.

Furthermore, let $u = u^*(x, t; j, n)$ and $g = g^*(x, t; j, n)$ satisfy Eq. (2.47) (i.e., $\nabla \cdot \mathbf{h}^* = 0$) within $\text{SE}(j, n)$. As a result,

$$(u_t)_j^n = -(g_x)_j^n = -u_j^n(u_x)_j^n. \tag{2.52}$$

Thus again, u_j^n and $(u_x)_j^n$ are the only independent marching variables at any $(j, n) \in \Psi$.

Let (j, n) denote an interior mesh point. We assume that Eq. (2.7) is valid for both $V = \mathrm{CE}_+(j, n)$ and $V = \mathrm{CE}_-(j, n)$. Then, with the aid of Eqs. (2.3), (2.11), (2.49), and (2.50), one concludes that

$$u_j^n - u_{j\pm1/2}^{n-1/2} \mp \frac{\Delta t}{\Delta x}\left(g_j^n - g_{j\pm1/2}^{n-1/2}\right) \pm \left\{(u_x^+)_j^n + (u_x^+)_{j\pm1/2}^{n-1/2} + \frac{(\Delta t)^2}{4\Delta x}\left[(g_t)_j^n + (g_t)_{j\pm1/2}^{n-1/2}\right]\right\} = 0. \tag{2.53}$$

By summing over the above two expressions and using Eqs. (2.51) and (2.52), one has

$$u_j^n = \frac{1}{2}\left[u_{j-1/2}^{n-1/2} + u_{j+1/2}^{n-1/2} + s_{j-1/2}^{n-1/2} - s_{j+1/2}^{n-1/2}\right]. \tag{2.54}$$

Here, for any $(j, n) \in \Psi$,

$$s_j^n \stackrel{\text{def}}{=} \left[1 - \left(\nu_j^n\right)^2\right](u_x^+)_j^n + \frac{1}{2}\,\nu_j^n\,u_j^n \tag{2.55}$$

with $\nu_j^n \stackrel{\text{def}}{=} (u_j^n \Delta t)/\Delta x$. Moreover, by substituting Eq. (2.54) into any one of the two expressions in Eq. (2.53) and assuming $1 - (\nu_j^n)^2 \neq 0$, it can be shown that

$$(u_x^+)_j^n = \left(u_x^{a+}\right)_j^n \stackrel{\text{def}}{=} \frac{u_{j+1/2}^{n-1/2} - u_{j-1/2}^{n-1/2} - s_{j+1/2}^{n-1/2} - s_{j-1/2}^{n-1/2} + \nu_j^n\,u_j^n}{2\left[1 - \left(\nu_j^n\right)^2\right]}. \tag{2.56}$$

At this juncture, note that Eq. (2.54) can be obtained directly from the assumption that Eq. (2.7) is valid for $V = \mathrm{CE}(j, n)$ where $\mathrm{CE}(j, n)$ is the union of $\mathrm{CE}_+(j, n)$ and $\mathrm{CE}_-(j, n)$ (see Fig. 1e). As explained in [2], the last assumption follows directly from the assumptions that Eq. (2.7) is valid for $V = \mathrm{CE}_\pm(j, n)$.

The scheme formed by Eqs. (2.54) and (2.56) is referred to as the inviscid Burgers *a* scheme. It is a nonlinear extension of the nondissipative *a* scheme (i.e., the inviscid version of the a–μ scheme). Such an extension generally is unstable; it must be modified to become a stable scheme. Note that the superscript symbol "a" in $(u_x^{a+})_j^n$ is introduced to remind the reader that Eq. (2.56) is valid for the inviscid Burgers *a* scheme only.

For the modified scheme, we impose a less stringent conservation condition, i.e., for each $(j, n) \in \Psi$, the modified scheme satisfies Eq. (2.7) with $V = \mathrm{CE}(j, n)$. Because this condition is equivalent to Eq. (2.54), the latter equation is also part of the modified scheme.

To proceed further, consider any $(j, n) \in \Psi$. Then, $(j \pm 1/2, n - 1/2) \in \Psi$. Let

$$u'^n_{j\pm1/2} \stackrel{\text{def}}{=} [u + (\Delta t/2)u_t]_{j\pm1/2}^{n-1/2}. \tag{2.57}$$

To simplify notation, in the above and hereafter we adopt a convention that can be explained using the expression on the right side of Eq. (2.57) as an example; i.e.,

$$[u + (\Delta t/2)u_t]_{j\pm1/2}^{n-1/2} = u_{j\pm1/2}^{n-1/2} + (\Delta t/2)(u_t)_{j\pm1/2}^{n-1/2}.$$

With the aid of Eq. (2.52) and $\nu_j^n \stackrel{\text{def}}{=} (u_j^n \Delta t)/\Delta x$, Eq. (2.57) implies that

$$u'^n_{j\pm1/2} = [u - 2\nu u_x^+]_{j\pm1/2}^{n-1/2}. \tag{2.58}$$

Note that, by definition, $(j \pm 1/2, n) \notin \Psi$ if $(j, n) \in \Psi$. Thus, $u'^n_{j\pm1/2}$ is not associated with a mesh point $\in \Psi$.

According to Eq. (2.57), $u'^n_{j\pm1/2}$ can be interpreted as a first-order Taylor's approximation of u at $(j \pm 1/2, n)$. Thus,

$$\left(u_x^{c+}\right)_j^n \stackrel{\text{def}}{=} \frac{u'^n_{j+1/2} - u'^n_{j-1/2}}{4} = \frac{\Delta x}{4}\left(\frac{u'^n_{j+1/2} - u'^n_{j-1/2}}{\Delta x}\right) \tag{2.59}$$

is a central-difference approximation of $\partial u/\partial x$ at (j, n), normalized by the same factor $\Delta x/4$ that appears in Eq. (2.11). Note that the superscript "c" is used to remind the reader of the central-difference nature of the term $(u_x^{c+})_j^n$.

Furthermore, let

$$\left(u_{x\pm}^{c+}\right)_j^n \stackrel{\text{def}}{=} \pm\frac{1}{2}\left(u'^n_{j\pm1/2} - u_j^n\right) = \pm\frac{\Delta x}{4}\left(\frac{u'^n_{j\pm1/2} - u_j^n}{\Delta x/2}\right). \tag{2.60}$$

By their definitions, $(u_{x+}^{c+})_j^n$ and $(u_{x-}^{c+})_j^n$ are two normalized numerical analogues of $\partial u/\partial x$ at (j, n), with one being evaluated from the right and another from the left. It can be shown that

$$\left(u_x^{c+}\right)_j^n = \frac{1}{2}\left[\left(u_{x+}^{c+}\right)_j^n + \left(u_{x-}^{c+}\right)_j^n\right]; \tag{2.61}$$

i.e., $(u_x^{c+})_j^n$ is the simple average of $(u_{x+}^{c+})_j^n$ and $(u_{x-}^{c+})_j^n$. Next, let the function W_o be defined by (i) $W_o(0, 0; \alpha) = 0$ and (ii)

$$W_o(x_-, x_+; \alpha) = \frac{|x_+|^\alpha x_- + |x_-|^\alpha x_+}{|x_+|^\alpha + |x_-|^\alpha} \quad (|x_+| + |x_-| > 0), \tag{2.62}$$

where x_+, x_- and $\alpha \geq 0$ are real variables. Note that (i) to avoid dividing by zero, in practice a small positive number such as 10^{-60} is added to the denominator in Eq. (2.62); and (ii) $W_o(x_-, x_+; \alpha)$, a nonlinear weighted average of x_- and x_+, becomes their simple average if $\alpha = 0$ or $|x_-| = |x_+|$. Furthermore, let

$$\left(u_x^{w+}\right)_j^n \stackrel{\text{def}}{=} W_o\left(\left(u_{x+}^{c+}\right)_j^n, \left(u_{x-}^{c+}\right)_j^n; \alpha\right). \tag{2.63}$$

Note that the superscript "w" is used to remind the reader of the weighted-average nature of the term $(u_x^{w+})_j^n$. With the aid of the above definitions, the modified scheme, referred to as the inviscid Burgers a–ϵ–α–β scheme, is formed by Eq. (2.54) and

$$(u_x^+)_j^n = \left(u_x^{a+}\right)_j^n + 2\epsilon\left(u_x^{c+} - u_x^{a+}\right)_j^n + +3\left(u_x^{w+} - u_x^{c+}\right)_j^n. \tag{2.64}$$

Here (i) $\epsilon \geq 0$ and $\beta \geq 0$ are adjustable parameters; and (ii) $(u_x^{w+})_j^n$ is implicitly dependent on the adjustable parameter α.

The expression on the right side of Eq. (2.64) contains three parts. The first part is a nondissipative term $(u_x^{a+})_j^n$. The second part is the product of 2ϵ and the difference between the central-difference term $(u_x^{c+})_j^n$ and the nondissipative term $(u_x^{a+})_j^n$. The third part is the product of β and the difference between a weighted average of $(u_{x+}^{c+})_j^n$ and $(u_{x-}^{c+})_j^n$ and their simple average (see Eq. (2.61)). Numerical dissipation of ϵ-type, i.e., that results from

adding the second part to the right side of Eq. (2.64), generally is effective in damping out numerical instabilities that arise from the smooth region of a solution. However, it is less effective in suppressing numerical wiggles that often occur near a discontinuity. On the other hand, numerical dissipation of α–β-type, i.e., that results from adding the third part, is very effective in suppressing numerical wiggles. Moreover, because the condition $|(u_{x+}^{c+})_j^n| = |(u_{x-}^{c+})_j^n|$ more or less prevails and thus the weighted average is nearly equal to the simple average (see comment (ii) given immediately following Eq. (2.62)) in the smooth region of the the solution, numerical dissipation introduced by the third part has very slight effect in the smooth region.

This section is concluded with the following comments:

(a) According to numerical evidence, stability of the current solver generally requires that (i) $0 \le \epsilon \le 1$, (ii) $\beta \ge 0$, (iii) $\alpha \ge 0$, and (iv) $|\nu_j^n| \le 1$ for all $(j, n) \in \Psi$.

(b) Let $\epsilon = 1/2$ and $\beta = 1$. Then the current scheme is formed by Eq. (2.54) and

$$(u_x^+)_j^n = \left(u_x^{w+}\right)_j^n. \tag{2.65}$$

For this special case, one does not need to evaluate $(u_x^{a+})_j^n$ and thus the condition that $1 - (\nu_j^n)^2 \ne 0$ can be ignored (see Eq. (2.56)). Moreover, the value of α is the only adjustable parameter allowed in the reduced scheme. Generally, with a choice of $\alpha = 1$ or $\alpha = 2$, the numerical dissipation introduced is sufficient to suppress numerical wiggles. Because it is totally explicit and has the simplest stencil, the reduced scheme is also highly compatible with parallel computing. Furthermore, it will be shown in Section 3 that the scheme can accurately capture shocks and contact discontinuities with high resolution and no numerical oscillations.

(c) For other remarks on the parameters ϵ, α, and β, the reader is referred to Section 5.5 in Ref. [3].

3. CONSISTENCY AND TRUNCATION ERROR

In this section, the consistency and the truncation error of the scheme formed by Eqs. (2.44) and (2.45), i.e., the circumstances under which an analytical solution may "satisfy" the above two discrete equations, will be investigated. As a preliminary to this investigation and to provide a basis for analyzing the numerical results given in Section 4, this section will begin with a discussion of several critical concepts.

First note that, in a typical numerical scheme, a physical variable is associated with a single numerical variable. Thus, a system of two coupled physical equations involving two independent physical variables generally is modeled by a system of two coupled discrete equations involving two independent numerical variables. Also, one would expect that the two coupled discrete equations are consistent with the two coupled physical equations. Thus, in general, one would not expect that two coupled discrete equations be consistent with only a single PDE.

The scheme formed by Eqs. (2.44) and (2.45) is nontraditional in one key respect. Even though it is introduced to model a single PDE (i.e., Eq. (2.1)) with a single dependent variable u, it is formed by two coupled discrete equations involving two *independent* numerical variables u_j^n and $(u_x)_j^n$.

The numerical variables u_j^n and $(u_x)_j^n$ could be "interpreted" as the numerical analogues of u and $\partial u/\partial x$, respectively. However, it should be understood that this interpretation is

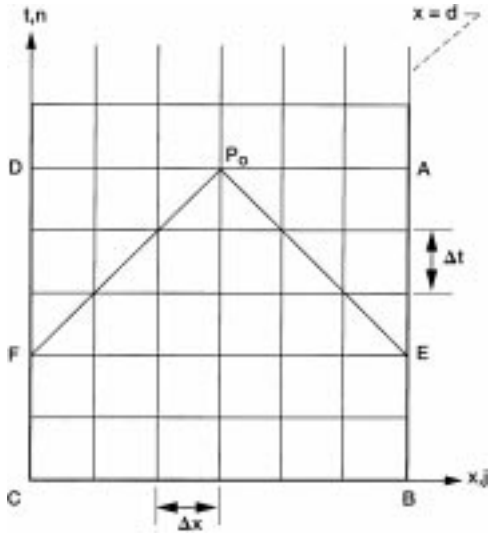

FIG. 3. A computational domain with $d \geq x \geq 0$ and $t \geq 0$ ($x = j\Delta x, t = n\Delta t$).

not exact in nature and that certainly it does not invalidate the fact that u_j^n and $(u_x)_j^n$ are independent numerical variables. As a result, one would expect that Eqs. (2.44) and (2.45) be consistent with a system of two PDEs, with one of them being Eq. (2.1).

Next we will discuss a general limitation on the ability of an explicit scheme to solve an initial-value/boundary-value problem accurately. As an example, consider Eq. (2.1) ($\mu > 0$) over a domain with $d \geq x \geq 0$ and and $t \geq 0$ (see Fig. 3). Let the initial data $u(x, 0)$ ($d \geq x \geq 0$) and the boundary data $u(0, t)$ and $u(d, t)$ ($t > 0$) be given. Let x_0 and t_0 be the coordinates of a fixed point P_0. Let $u(P_0)$ and $\underline{u}(P_0)$, respectively, denote the values of analytical and discrete solutions at P_0. Since a characteristic of Eq. (2.1) is represented by $t =$ constant, the domain of dependence of $u(P_0)$ is the union of $\overline{AB}$, $\overline{BC}$, and $\overline{CD}$. In other words, $u(P_0)$ is dependent on all the initial data, and the boundary data with $t \leq t_0$. Assuming that the discrete solution is generated by an explicit solver, then the domain of dependence of $\underline{u}(P_0)$, contrarily, will include only a subset of the mesh points located on $\overline{AB}$, $\overline{BC}$, and $\overline{CD}$. As an example, consider an explicit scheme with the marching variables at the mesh point $(j, n + 1)$ being determined by those at the mesh points at (j, n) and $(j \pm 1, n)$. As a result, the domain of dependence of $\underline{u}(P_0)$ includes only the mesh points on $\overline{EB}$, $\overline{BC}$, and $\overline{CF}$. (Note: Here, a line segment includes its end points.) Because (i) the mesh points that lie on $\overline{AB}$ but not $\overline{EB}$ and those that lie on $\overline{CD}$ but not $\overline{CF}$ do not belong to the domain of dependence of $\underline{u}(P_0)$, and (ii) the lengths of $\overline{AE}$ and $\overline{FD}$ are proportional to the ratio $\Delta t/\Delta x$ if the values of x_0 and t_0 remain fixed as $\Delta t, \Delta x \to 0$, one may conclude that, as $\Delta t, \Delta x \to 0$, the *discrete* solution (considered a function of Δt and Δx) cannot converge to its analytical counterpart unless $\Delta t/\Delta x \to 0$. It follows from Lax's equivalence theorem [22, p. 45] that, for an explicit solver of Eq. (2.1) with $\mu > 0$, the condition that $\Delta t/\Delta x \to 0$ as $\Delta t, \Delta x \to 0$ must be required by consistency or stability or both. As an example, for the MacCormack scheme (see Section 4), the last condition is required by the necessary stability condition $\mu\Delta t/(\Delta x)^2 \leq 0.75$ (see Fig. 4). On the other hand, for the a–μ scheme (as will be shown shortly), it is required by consistency. Note that the stability of the a–μ scheme [1, 2] is limited only by the condition $|\nu| \leq 1$, which does not require that $\Delta t \to 0$ as $\Delta t, \Delta x \to 0$.

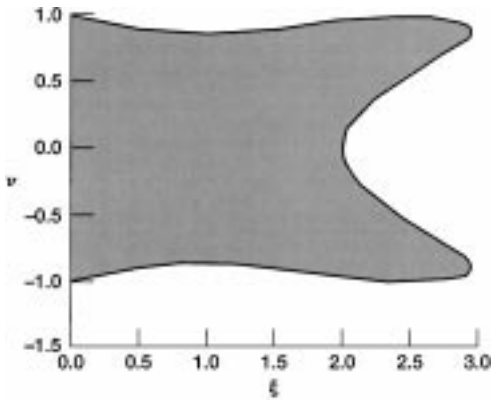

FIG. 4. Stability region (shaded area) of the MacCormack scheme on the ξ–ν plane.

Also note that, for a problem with an unbounded domain and a periodic initial condition, a discussion similar to that given above is given in [1, p. 55].

Furthermore, as a result of above considerations, and the fact that the analytical domain of dependence can be matched by the domain of dependence of an implicit scheme, one concludes that, for an initial-value/boundary-value problem, an explicit solver is generally not as accurate as an implicit solver. Generally, an explicit solver should not be used to solve such a problem except for the special circumstance in which errors caused by neglecting certain initial/boundary data are relatively small. The factors that help achieve the above special circumstance include: (i) a small value of $\Delta t/\Delta x$, (ii) a small time rate of change of boundary data, and (iii) a small contribution of the diffusion terms relative to that of the convection terms.

On the other hand, for a pure initial-value problem, such as a problem involving Eq. (2.1) with $\mu = 0$, an implicit solver generally is not as accurate as an explicit solver. This is because the domain of dependence of the former solver may be far greater than the analytical domain of dependence and, as a result, an implicit solution tends to be contaminated by extraneous information.

For simplicity, the consistency of the finite discrete equations, Eqs. (2.44) and (2.45), will be investigated fully here only for the special case with $a = 0$ and $\mu > 0$. For the general case, the reader is referred to Section 6 in [1].

Because (i) $\nu = 0$ if $a = 0$; and (ii) $(u_x^+)_j^n = (\Delta x/4)(u_x)_j^n$, for the special case under consideration, Eqs. (2.44) and (2.45) reduce to

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} - \mu \left\{ \frac{2}{1+\xi} \cdot \frac{u_{j+1}^n + u_{j-1}^n - 2u_j^n}{(\Delta x)^2} + \frac{\xi - 1}{\xi + 1} \cdot \frac{(u_x)_{j+1}^n - (u_x)_{j-1}^n}{2\Delta x} \right\} = 0 \quad (3.1)$$

and

$$\frac{(u_x)_j^{n+1} - (u_x)_j^n}{\Delta t} - \mu \left\{ \frac{2(\xi - 1)}{(\xi + 1)^2} \cdot \frac{(u_x)_{j+1}^n + (u_x)_{j-1}^n - 2(u_x)_j^n}{(\Delta x)^2} + \frac{8}{(\xi + 1)^2} \cdot \frac{u_{j+1}^n - u_{j-1}^n - 2\Delta x (u_x)_j^n}{(\Delta x)^3} \right\} = 0, \quad (3.2)$$

respectively. Let $q_1(x, t)$ and $q_2(x, t)$ be two smooth functions of x and t. If $u_j^n = q_1(x_j, t^n)$

and $(u_x)_j^n = q_2(x_j, t^n)$ form a solution to Eqs. (3.1) and (3.2), then with the aid of Taylor's formula with reminder, one has

$$\left[\frac{\partial q_1}{\partial t} - \mu\frac{\partial^2 q_1}{\partial x^2}\right] - \mu\frac{\xi - 1}{\xi + 1}\frac{\partial}{\partial x}\left(q_2 - \frac{\partial q_1}{\partial x}\right) + O(\Delta t, (\Delta x)^2) = 0 \tag{3.3}$$

and

$$\left[q_2 - \frac{\partial q_1}{\partial x}\right] + \frac{(\Delta x)^2 - 4\mu\Delta t}{8}\frac{\partial^2 q_2}{\partial x^2} - \frac{(\Delta x)^2}{6}\frac{\partial^3 q_1}{\partial x^3} + \frac{1}{16\mu}\left(\frac{4\mu\Delta t}{\Delta x} + \Delta x\right)^2\left(\frac{\partial q_2}{\partial t} + O(\Delta t, (\Delta x)^2)\right) = 0. \tag{3.4}$$

Note that, to emphasize the fact that u_j^n and $(u_x)_j^n$ are two independent marching variables, here new symbols q_1 and q_2 are introduced to denote the functions that are the analytical counterparts of u_j^n and $(u_x)_j^n$, respectively. Also, it should be understood that in Eqs. (3.3) and (3.4), q_1 and q_2 and their derivatives represent the values at the mesh point (j, n).

Equations (3.1) and (3.2) can be considered the numerical approximations to the PDEs

$$\frac{\partial q_1}{\partial t} - \mu\frac{\partial^2 q_1}{\partial x^2} = 0 \quad \text{and} \quad q_2 - \frac{\partial q_1}{\partial x} = 0, \tag{3.5}$$

respectively. According to Eqs. (3.3) and (3.4), the truncation errors of these approximations are the terms after the brackets in Eqs. (3.3) and (3.4). Let q_1 and q_2 uniformly satisfy the PDEs given in Eq. (3.5). Then obviously the first term after the brackets in Eq. (3.3) vanishes. As a result, it is easy to see that the truncation errors that appear in Eqs. (3.3) and (3.4) $\to$ 0 in the limit of $\Delta t, \Delta x \to 0$ if the mesh is refined in such a manner that $\Delta t/\Delta x \to 0$ as $\Delta t, \Delta x \to 0$. In other words, assuming the above rule of mesh refinement, Eqs. (3.1) and (3.2) are consistent with the system of the PDEs given in Eq. (3.5). Moreover, one can conclude that the above truncation errors are second order in Δx if the rule of mesh refinement is such that ξ remains bounded as $\Delta x \to 0$ and $\Delta t \to 0$.

For the more general case in which $a \neq 0$ and $\mu > 0$, it is shown in Section 6 of [1] that Eqs. (2.44) and (2.45) are consistent with the system of the PDEs

$$\frac{\partial q_1}{\partial t} + a\frac{\partial q_1}{\partial x} - \mu\frac{\partial^2 q_1}{\partial x^2} = 0 \quad \text{and} \quad q_2 - \frac{\partial q_1}{\partial x} = 0. \tag{3.6}$$

Also the truncation errors $\to 0$ in the limit of $\Delta t, \Delta x \to 0$, assuming the mesh is refined in such a manner that $\Delta t/\Delta x \to 0$ as $\Delta t, \Delta x \to 0$

Note that Eq. (2.1) reduces to

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = 0 \tag{3.7}$$

when $\mu = 0$. For this special case, it is shown in Section 6 of [1] that Eqs. (2.44) and (2.45) are consistent with the system of the PDEs

$$\frac{\partial q_1}{\partial t} + a\frac{\partial q_1}{\partial x} = 0 \tag{3.8}$$

and

$$\frac{\partial}{\partial t}\left(q_2 - \frac{\partial q_1}{\partial x}\right) - a\frac{\partial}{\partial x}\left(q_2 - \frac{\partial q_1}{\partial x}\right) + \frac{\partial}{\partial x}\left(\frac{\partial q_1}{\partial t} + a\frac{\partial q_1}{\partial x}\right) = 0. \tag{3.9}$$

Because, in this case, the truncation errors are all second order in Δx and Δt, consistency does not require that the mesh be refined in such a manner that $\Delta t/\Delta x \to 0$ as $\Delta t, \Delta x \to 0$. Obviously, this conclusion is consistent with the fact that Eq. (3.7) is associated with a pure initial-value problem.

Finally, note that the consistency and truncation error of the a–ϵ scheme are discussed in Section 7 of [8].

4. NUMERICAL EVALUATION

The accuracy of the numerical schemes described in Section 2 is evaluated here by comparing their numerical results with the exact solutions and the results generated by other traditional schemes.

4.1. *The a Scheme*

The a scheme is the special case of the a–μ scheme with $\mu = 0$. It is the only two-level, explicit, and nondissipative solver of Eq. (3.7) known to the authors. As a matter of fact, it is shown in [2] that the two amplification factors of the a scheme are identical to those of the "decoupled" Leapfrog scheme. Note that the ordinary three-level Leapfrog scheme [23, p. 100] is formed by two completely decoupled schemes. Because these two decoupled schemes are identical in structure, any one of them is referred to as the decoupled Leapfrog scheme. Using the mesh depicted in Fig. 1a, the decoupled scheme can be expressed as (see Eq. (A.9) in [1])

$$u_j^n = u_j^{n-1} + \nu\left(u_{j-1/2}^{n-1/2} - u_{j+1/2}^{n-1/2}\right). \tag{4.1}$$

Obviously, like the a scheme, the mesh points associated with the decoupled Leapfrog scheme also are staggered in space–time. However, unlike the two-level a scheme, the three-level scheme Eq. (4.1) needs to be supplemented by a two-level starting scheme. In this paper, the starting scheme used is

$$\frac{u_j^{1/2} - u_j^0}{\Delta t/2} + a\frac{u_{j+1/2}^0 - u_{j-1/2}^0}{\Delta x} = 0 \tag{4.2}$$

In the following, the accuracy and the operation count of the a scheme will be compared against those of the decoupled Leapfrog scheme and the Lax–Wendroff scheme. Note that the three schemes under comparison (excluding the starting scheme Eq. (4.2)) are all second-order accurate in space and time. Using the mesh depicted in Fig. 5, the Lax–Wendroff scheme can be expressed as

$$u_j^{n+1} = \frac{\nu'(\nu'+1)}{2}u_{j-1}^n + (1-\nu'^2)u_j^n + \frac{\nu'(\nu'-1)}{2}u_{j-1}^n, \quad \nu' \stackrel{\text{def}}{=} \frac{a\Delta t'}{\Delta x'}. \tag{4.3}$$

In Fig. 5, for a reason which will become clear shortly, the spatial mesh interval and the time-step size are denoted by the new symbols $\Delta x'$ and $\Delta t'$, respectively.

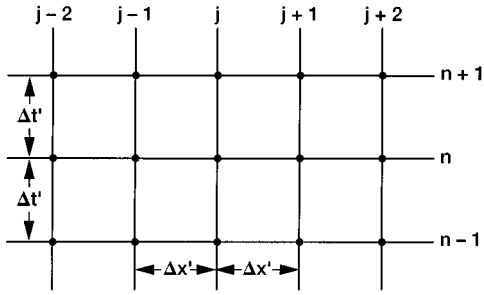

FIG. 5. A regular space–time mesh.

Consider a model problem involving Eq. (3.7). Let (i) $a = 0.5$ and (ii)

$$u(x, 0) = \sin(\pi x), \quad -\infty < x < \infty \tag{4.4}$$

Then the exact solution is

$$u(x, t) = \sin(\pi(x - 0.5t)), \quad -\infty < x < \infty; \quad t \geq 0. \tag{4.5}$$

Obviously, the exact solution represents a wave motion with the wavelength $\lambda = 2$ and the period $T = 4$. Thus, one may limit the computational domain to $-1 \leq x \leq 1$ and $t \geq 0$ and use the periodic boundary condition. Furthermore, it will be assumed that $u_j^0 = u(x_j, 0)$ and $(u_x)_j^0 = \frac{\partial u}{\partial x}(x_j, 0)$.

Let (i) $\Delta x = 0.04$ and $\Delta t = 10/131$ (i.e., $\nu = 125/131 = 0.954$) for the a scheme and the decoupled Leapfrog scheme; and (ii) $\Delta x' = 0.02$ and $\Delta t' = 5/131$ (i.e., $\nu' = 125/131 = 0.954$) for the Lax–Wendroff scheme. Let

$$\epsilon_j^n \stackrel{\text{def}}{=} u_j^n - u(x_j, t^n). \tag{4.6}$$

Given the above definitions, the error distributions for the above three schemes at $t = 10 = 2.5T$ are depicted in Fig. 6a. Note that totally there are (i) 51 data points for the a scheme and also for the Leapfrog scheme and (ii) 101 data points for the Lax–Wendroff scheme. From the results shown, one concludes that the errors of the a scheme and the Lax–Wendroff scheme vary smoothly in the x-direction and they almost fall on each other at all spatial locations where the mesh points of these two schemes coincide. On the other hand, the errors of the decoupled Leapfrog scheme fluctuate rather erratically from one mesh point to another neighboring mesh point. Even though the mean value of $|\epsilon_j^n|$ at $t = 10$ for the decoupled Leapfrog scheme is only about 8% higher than those for the a scheme and the Lax–Wendroff scheme, the maximum of $|\epsilon_j^n|$ for the first scheme is more than twice those for the last two schemes.

Assuming the same values of Δx, Δt, $\Delta x'$, and $\Delta t'$, the error distributions for the above three schemes at $t = 100$ are depicted in Fig. 6b. At this time, the erratic behavior of the errors of the decoupled Leapfrog scheme is much less pronounced and, as a result, the errors of all three schemes are more or less identical at all spatial locations where the mesh points of these schemes coincide. It appears that the erratic behavior of the errors of the decoupled Leapfrog scheme at the earlier time is due to the errors introduced by the starting scheme

(which has only first-order accuracy in time). Because, as time increases, these initial errors will eventually become negligible compared with other numerical errors accumulated over time, the total errors will behave in a less erratic manner as time increases.

It has been shown that, assuming $\Delta x' = \Delta x/2$ and $\Delta t' = \Delta t/2$, the *a* scheme and the Lax–Wendroff scheme have almost identical accuracy. Furthermore, assuming that the same values of Δx and Δt are used, the *a* scheme and the decoupled Leapfrog scheme, on the average, have more or less identical accuracy, albeit the latter scheme may have much higher maximal local error when t is small. In the following, the above three schemes will be further judged on the basis of their operation counts.

According to Eq. (4.1) and the comments made following Eq. (2.38), for the decoupled Leapfrog scheme and the *a* scheme, at each time level these schemes require, respectively, *one* and *two* multiplications per mesh point to advance the numerical solution by the time interval $\Delta t/2$ (i.e., to advance by a single marching step). On the other hand, according to Eq. (4.3) and Fig. 5, for the Lax–Wendroff scheme, at each time level it takes *three* multiplications per mesh point to advance by the time interval $\Delta t'$ (i.e., to advance by a single marching step in the Lax–Wendroff scheme). Here, it is again assumed that (i) additions and subtractions, which can be performed much faster than multiplications, are ignored in operation counts; and (ii) the three combination coefficients on the right side of Eq. (4.3) are to be evaluated once and stored for repeated later calculations. Because the numerical results shown in Fig. 6 are generated assuming (i) $\Delta t' = \Delta t/2$ and (ii) $\Delta x' = \Delta x/2$ (i.e., the number of mesh points per time level used in the Lax–Wendroff scheme is twice that used in the *a* scheme and the decoupled Leapfrog scheme), one concludes that the Lax–Wendroff scheme can achieve the same accuracy as that of the *a* scheme only at the expense of an operation count that is three times that of the latter scheme.

The section is concluded with the following comments:

(a) A solution of the ordinary Leapfrog scheme is formed by two completely decoupled solutions. As time increases, these solutions will gradually deviate from the correct solution and, because of their decoupled nature, from each other. Thus, one of these two solutions is completely redundant. For this reason, using the ordinary Leapfrog scheme instead of the

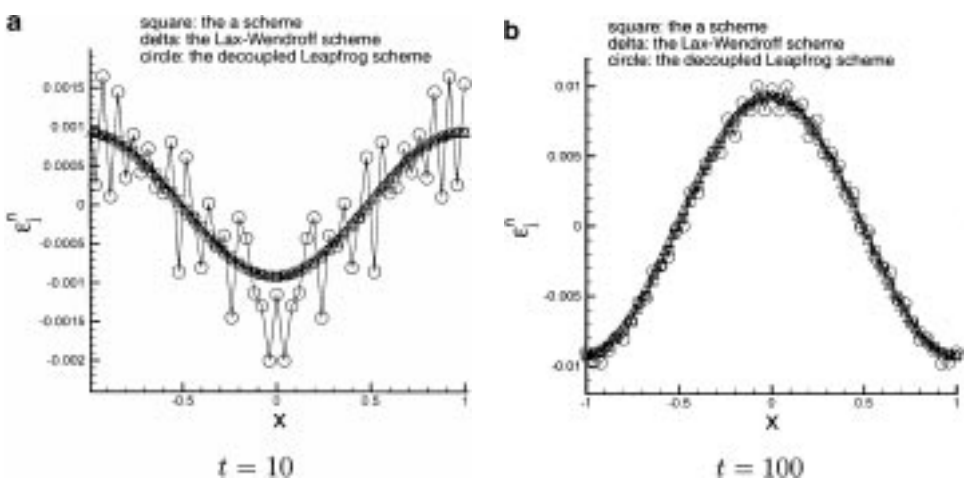

FIG. 6. The error distributions of the *a* scheme, the decoupled Leapfrog scheme, and the Lax–Wendroff scheme ($\Delta x = 0.04$, $\Delta t = 10/131$, $\Delta x' = 0.02$, and $\Delta t' = 5/131$).

decoupled Leapfrog scheme is pointless, i.e., it simply doubles computer cost without any gain in accuracy. The same observation is also applicable to other schemes that have the same decoupled nature (e.g., the DuFort–Frankel scheme and the Lax scheme).

(b) Even though the *a* scheme and the decoupled Leapfrog scheme have the same amplification factors, the former scheme has more compact stencil than the latter scheme. As it turns out, this compactness and the fact that the mesh values of both the dependent variable and its spatial derivative are carried at each mesh point make it much easier to construct robust and accurate generalizations of the *a* scheme.

4.2. *The μ Scheme*

The μ scheme is the special case of the a–μ scheme with $a = 0$. It is shown in [1] that the two-amplification factors of the μ scheme are identical to those of the "decoupled" DuFort–Frankel scheme. Using the mesh depicted in Fig. 1a, the decoupled scheme can be expressed as (see Eq. (A.9) in [2])

$$u_j^n = \frac{1-\xi}{1+\xi} u_j^{n-1} + \frac{\xi}{1+\xi}\left(u_{j-1/2}^{n-1/2} + u_{j+1/2}^{n-1/2}\right). \tag{4.7}$$

Obviously, like the μ scheme, the mesh points associated with the decoupled DuFort–Frankel scheme also are staggered in space–time. However, unlike the two-level μ scheme, the three-level scheme Eq. (4.7) needs to be supplemented by a two-level starting scheme. In this paper, for the *interior* mesh points $(j, 1/2)$, the starting scheme used is

$$\frac{u_j^{1/2} - u_j^0}{\Delta t/2} = \mu \frac{u_{j+1/2}^0 + u_{j-1/2}^0 - 2u_j^0}{(\Delta x/2)^2} = 0. \tag{4.8}$$

In the following, the accuracy and the operation count of the μ scheme will be compared against those of the decoupled DuFort–Frankel scheme and the forward-time central-space (FTCS) scheme. Using the mesh depicted in Fig. 5, the FTCS scheme can be expressed as

$$u_j^n = u_j^{n-1} + \alpha'\left(u_{j+1}^{n-1} + u_{j-1}^{n-1} - 2u_j^{n-1}\right) \quad (\alpha' \stackrel{\text{def}}{=} \mu \Delta t'/(\Delta x')^2). \tag{4.9}$$

Consider a model problem [24] that is defined by (i) the PDE

$$\frac{\partial u}{\partial t} - \mu \frac{\partial^2 u}{\partial x^2} = 0, \tag{4.10}$$

where $\mu = 2.17 \times 10^{-4}$; (ii) the initial condition $(t = 0)$

$$u = \begin{cases} 40 & \text{if } x = 0; \\ 0 & \text{if } 0 < x \le 0.04; \end{cases} \tag{4.11}$$

and the boundary condition $(t > 0)$

$$u = \begin{cases} 40 & \text{if } x = 0; \\ 0 & \text{if } x = 0.04. \end{cases} \tag{4.12}$$

Note that the exact solution of this problem is given in [24]. Also note that, as a result of the above definitions, (i) the computational domain is limited to $0 \le x \le 0.04$ and $t \ge 0$;

and (ii) it is assumed that $u_j^n = 40$ if (j, n) is on the left boundary ($x = 0$) and $u_j^n = 0$ if (j, n) is on the right boundary ($x = 0.04$). Moreover, the reader is reminded that, for the μ scheme, $(u_x)_j^n$ at any boundary mesh point (j, n) is a marching variable to be evaluated.

In the numerical simulation involving the μ scheme and the decoupled DuFort–Frankel scheme, again the staggered mesh depicted in Fig. 1a is used. Let (i) $x_j = j\Delta x$ and $t^n = n\Delta t$; and (ii) Δx and an integer J be chosen such that $x_J = 0.04$. Furthermore, it is assumed that: (i) for $n = 0, 1, 2, \ldots$, $(j, n) \in \Psi$ if and only if $j = 1/2, 3/2, \ldots, (J - 1/2)$; and (ii) for $n = 1/2, 3/2, \ldots$, $(j, n) \in \Psi$ if and only if $j = 0, 1, 2, \ldots, J$. As a result, the origin $(0, 0) \notin \Psi$. Thus, one can safely assume $u_j^0 = (u_x)_j^0 = 0$ for $j = 1/2, 1, 3/2, 2, \ldots, (J - 1/2)$. Note that: (i) to apply the starting scheme Eq. (4.8), u_j^0 must be specified at $j = 1/2, 1, 3/2, 2, \ldots, (J - 1/2)$, not only at those values of j with $(j, 0) \in \Psi$; and (ii) because the initial value has a spurious jump at the origin, the fact that one does not need to specify initial value there generally results in more accurate numerical results.

On the other hand, the regular mesh depicted in Fig. 5 is used in the numerical simulation involving the FTCS scheme. In this case, it is assumed that $x_j = j\Delta x'$, $t^n = n\Delta t'$ and $x_{J'} = 0.04$ with J' being an integer. According to Fig. 5 and Eq. (4.9), in the case of the FTCS scheme, the origin must be a mesh point. Thus, the numerical initial conditions are defined by $u_0^0 = 40$ and $u_j^0 = 0$, $j = 1, 2, \ldots, J'$.

Let $\Delta x = \Delta t = 0.001$ and $\Delta x' = \Delta t' = 0.0005$. Then the error distributions of the above three schemes at $t = 0.18$ are those depicted in Fig. 7a. On the other hand, the distributions at $t = 1.08$ are those depicted in Fig. 7b. It is seen that, for both $t = 0.18$ and $t = 1.08$, the errors of the μ scheme and the decoupled DuFort–Frankel scheme are almost identical at all mesh points, and they are smaller than those of the FTCS scheme at most mesh points. Note that for the current dissipative case, ϵ_j^n decays with u_j^n as time increases.

Note that for the FTCS scheme (see Eq. (4.9)), it requires one multiplication per mesh point to advance by the time interval $\Delta t'$ (i.e., to advance by one marching step). On the other hand, for the decoupled DuFort–Frankel scheme and the μ scheme (see Eq. (4.7) and the comments made following Eq. (2.38)), it requires two multiplications per mesh point to advance by the time interval $\Delta t/2$ (i.e., to advance by one marching step). Because (i)

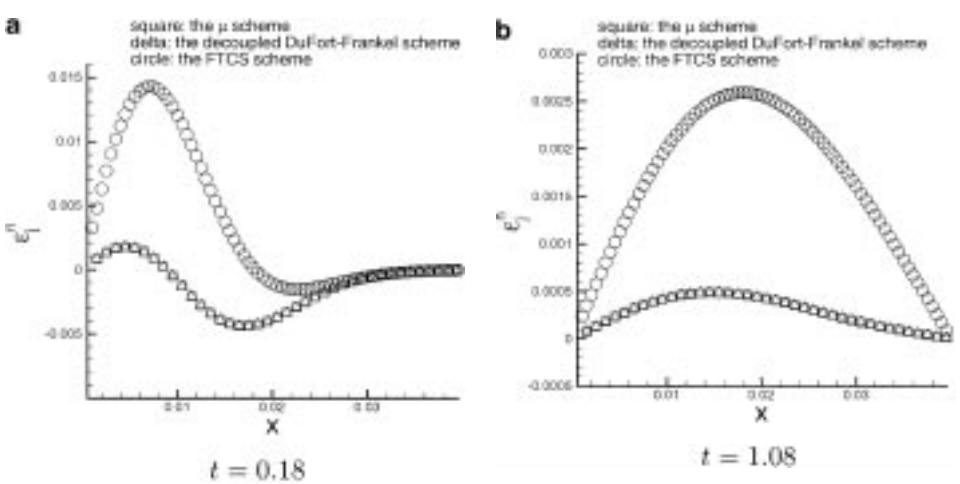

FIG. 7. The error distributions of the μ scheme, the decoupled DuFort–Frankel scheme, and the FTCS scheme ($\Delta x = \Delta t = 0.001$, and $\Delta x' = \Delta t' = 0.0005$).

$\Delta t' = \Delta t/2$ and (ii) $\Delta x' = \Delta x/2$ (i.e., the number of mesh points per time level used in the FTCS scheme is twice that used in the μ scheme and the decoupled DuFort–Frankel scheme), one concludes that the results shown in each of Figs. 7a and 7b are obtained with approximately the same operation count for each of the three schemes considered.

This section is ended with a numerical evaluation of an analytical prediction given earlier, i.e., for a given Δx, the accuracy of the a–μ scheme will reach a peak if Δt are chosen such that Eq. (2.46) is satisfied. Obviously, for the current case $a = 0$, Eq. (2.46) reduces to $\xi = 1/\sqrt{3}$. With the aid of Eq. (2.11), the last expression implies that

$$\Delta t = \Delta t_0 \stackrel{\text{def}}{=} \frac{(\Delta x)^2}{4\sqrt{3\mu}} \quad (a = 0). \tag{4.13}$$

Note that, in [1], the same Fourier error analysis from which Eq. (2.46) was derived also was used to study the Leapfrog/DuFort–Frankel scheme—a scheme which reduces to the ordinary Leapfrog scheme when $\mu = 0$ and to the ordinary DuFort–Frankel scheme when $a = 0$. By accounting for the fact that a regular mesh was used in that study and for the differences in notations (such as the fact that Δx and Δt used in the study of the Leapfrog/DuFort–Frankel scheme correspond to $\Delta x/2$ and $\Delta t/2$ in this paper), one may infer from Eq. (5.68) of [1] that the accuracy of the decoupled DuFort–Frankel scheme also will reach a peak if $\Delta t = \Delta t_0$.

Let (i) $\Delta x = 0.004$, (ii) $\mu = 2.17 \times 10^{-4}$ and (iii) $\Delta t = \Delta t_0 \doteq 0.0106424$. Then the error distributions of the μ scheme and the decoupled DuFort–Frankel scheme at $t = 17\Delta t_0 \doteq 0.1809$ are those depicted in Fig. 8a. On the other hand, the distributions at $t = 101\Delta t_0 \doteq 1.0749$ are those depicted in Fig. 8b. It is seen that the error distributions of the above two schemes again are almost identical at all mesh points. Furthermore, a comparison of the results shown in Figs. 7a–8b implies that, by choosing $\Delta t = \Delta t_0$, it is possible to improve the accuracy of the μ scheme and the decoupled DuFort–Frankel scheme while simultaneously using much bigger Δx and Δt. This "strange" phenomenon can be explained by a fact established by numerical experiments (i.e., with Δt and Δx being related by Eq. (4.13)), the above two schemes effectively achieve an accuracy which is fourth order in Δx.

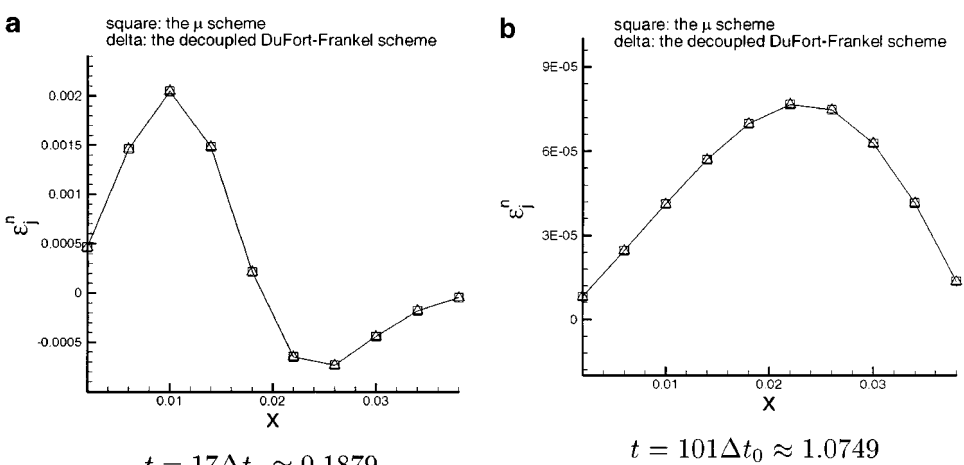

FIG. 8. The error distributions of the μ scheme and the decoupled DuFort–Frankel scheme ($\Delta x = 0.004$, $\Delta t = \Delta t_0 \approx 0.0106424$).

At this junction, note that, according to a discussion given in Section 3, for an initial-value/boundary-value problem, an explicit solver is generally not as accurate as an implicit solver. As a result, two implicit schemes for solving Eq. (2.1) have been constructed using the CE/SE method [9, 10]. In the inviscid case, both schemes reduce to the current explicit a scheme. On the other hand, in the pure diffusion case, these two schemes also become the same scheme and their principal amplification factor is identical to the amplification factor of the implicit Crank–Nicolson scheme. As a result, in the pure diffusion case, the Crank–Nicolson scheme and the two CE/SE implicit schemes are of similar accuracy [10].

4.3. *The Inviscid Burgers a–ϵ–α–β Scheme*

Consider the hyperbolic problem defined by Eq. (2.47) and the initial condition

$$u(x,0) = \begin{cases} 1 & \text{if } x < 0; \\ 0 & \text{if } x > 0. \end{cases} \tag{4.14}$$

The weak solution to this problem is [23, p. 142]

$$u(x,t) = \begin{cases} 1 & \text{if } x - t/2 < 0; \\ 0 & \text{if } x - t/2 > 0; \end{cases} \tag{4.15}$$

i.e., the discontinuity propagates in the x-direction with a speed of $1/2$.

The above problem is solved by the simplest inviscid Burgers a–ϵ–α–β scheme (i.e., that formed by Eqs. (2.54) and (2.65) with $\alpha = 1$). The computational domain ($-2 \le x \le 2$ and $0 \le t$) is covered by a space–time staggered mesh with $\Delta x = 0.1$. The locations of mesh points (dots in Fig. 1(a)) are determined by the assumptions: (i) $x_j = j\Delta x$ and $t^n = n\Delta t$, and (ii) $(j,n) \in \Psi$ if and only if $j + n$ is a half-integer. With $J \stackrel{\text{def}}{=} 2/\Delta x = 20$, the initial conditions used are:

$$u_j^0 = u(x_j, 0) \quad \text{and} \quad (u_x)_j^0 = 0, \tag{4.16}$$

where $j = -J + 1/2, -J + 3/2, \ldots, -1/2, 1/2, \ldots, J - 3/2, J - 1/2$. The boundary conditions used are the simple extrapolation conditions

$$u_{\pm J}^n = u_{\pm J \mp 1/2}^{n-1/2} \quad \text{and} \quad (u_x)_{\pm J}^n = (u_x)_{\pm J \mp 1/2}^{n-1/2}, \tag{4.17}$$

where $n = 1/2, 3/2, 5/2, \ldots$.

The numerical solutions obtained at $t = 1.8$ with $\Delta t = 0.1$ (i.e., the maximal Courant number $\nu_m = 1.0$) and $\Delta t = 0.06$ (i.e., $\nu_m = 0.6$) are shown in Fig. 9a. It is seen that the current very simple shock-capturing scheme can generate nearly perfect solutions. For each solution shown in Fig. 9a, the shock is resolved almost by a single mesh interval and no numerical wiggles are detected in its vicinity. Obviously, the best shock resolution occurs when $\nu_m = 1.0$.

The numerical solutions obtained at $t = 4.2$ with $\Delta t = 0.1$ and $\Delta t = 0.06$ are shown in Fig. 9b. At this time, the shock is located at $x = 2.1$; i.e., it has just exited the computational domain. Thus, the exact solution is $u = 1$ within this domain. The maximum magnitude of the errors in the numerically computed values of u is less than 10^{-4} (10^{-3}) in the case $\Delta t = 0.1$ ($\Delta t = 0.06$). Thus, one concludes that the simple extrapolation conditions Eq. (4.17) are excellent nonreflecting boundary conditions if they are applied in conjunction

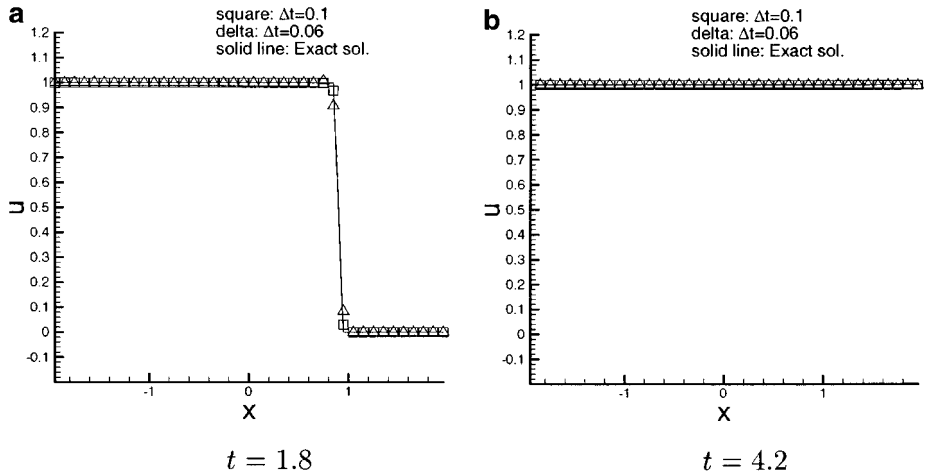

FIG. 9. Comparisons of the CE/SE solutions of the inviscid Burgers equation with the exact solution.

with the CE/SE method. This conclusion is also consistent with the theoretical results presented in [11].

At this juncture, note that, for each $(j, 0) \in \Psi$, the line segment joining the two space–time points $(x_j \pm \Delta x/2; 0)$ is part of $\mathrm{SE}(j, 0)$. As a result, the space–time flux passing through the above line segment can be evaluated using either the exact or the numerical initial condition. The resulting two values are identical if and only if

$$\int_{x_j-\Delta x/2}^{x_j+\Delta x/2} u^*(x, 0; j, 0)\, dx = \int_{x_j-\Delta x/2}^{x_j+\Delta x/2} u(x, 0)\, dx. \tag{4.18}$$

Because of its flux-based nature, accuracy of the CE/SE method generally will suffer (particularly if the exact initial condition is not continuous) if the numerical initial condition specified does not satisfy Eq. (4.18). Obviously, the numerical initial condition used here does satisfy Eq. (4.18). Note that, in case $(j, 0) \in \Psi$ for $j = 0, \pm 1, \pm 2, \ldots, \pm J$, Eq. (4.18) is satisfied by the initial condition: (i) $u_j^0 = u(x_j, 0)$ and $(u_x)_j^0 = 0$ if $j \neq 0$; and (ii) $u_0^0 = 1/2$ and $(u_x)_0^0 = c$ where c is an arbitrary constant.

5. CONCLUSIONS

Many important topics left untreated in [2], such as the consistency, accuracy, and operation count of the a–μ scheme, were discussed in this paper. As part of these discussions, an equivalent yet numerically more efficient and physically more appealing form of the a–μ scheme was introduced. The key conclusions of the discussions include the following:

(a) It is shown that the a–μ scheme is consistent with a system of two PDEs involving two dependent variables, with one of the PDEs being Eq. (2.1). This result corresponds closely to the fact that, for each $(j, n) \in \Psi$, the a–μ scheme is formed by two equations involving two *independent* unknowns u_j^n and $(u_x)_j^n$.

(b) It is shown that as $\Delta t, \Delta x \to 0$, a discrete solution of an *explicit* solver of the convection–diffusion equation Eq. (2.1) with $\mu > 0$ cannot converge to its analytical

counterpart unless the mesh is refined in such a manner that $\Delta t/\Delta x \to 0$ as $\Delta t, \Delta x \to 0$. As a result, Lax's equivalence theorem implies that, for such a solver, convergence requires that the above rule of mesh refinement be imposed through consistency or stability or both. For many explicit schemes, such as the MacCormack scheme, the above rule of mesh refinement is imposed through a stability condition. On the other hand, for the a–μ scheme, it is imposed as a requirement of consistency.

(c) It is shown that, in spite of the fact that both schemes are second order in accuracy, the a scheme (i.e., the a–μ scheme with $\mu = 0$) may achieve the same accuracy as that of the Lax–Wendroff scheme with an operation count being only one-third of that of the latter scheme.

(d) Excluding its two-level starting scheme, the ordinary three-level Leapfrog scheme is formed by two completely identical and decoupled subschemes. Any one of these subschemes is referred to as the decoupled Leapfrog scheme. The amplification factors of the decoupled Leapfrog scheme are identical to those of the a scheme. However, the actual accuracy of the former scheme is degraded by the first-order errors introduced by its starting scheme. As a result, the decoupled Leapfrog scheme is less accurate than the a scheme, a fact that is most prominent in the earlier time during which the errors of the former scheme fluctuate erratically from one mesh point to another.

(e) As in the case of the Leapfrog scheme, any one of the two completely identical and decoupled subschemes that form the ordinary three-level DuFort–Frankel scheme is referred to as the decoupled DuFort–Frankel scheme. The amplification factors of the decoupled DuFort–Frankel scheme are identical to those of the μ scheme (i.e., the a–μ scheme with $a = 0$). Assuming that the same values of Δx and Δt are used, it is shown that the decoupled DuFort–Frankel scheme and the μ scheme have (i) the same operation count, and (ii) almost the same accuracy. Note that because of the effect of the viscosity, the first-order errors introduced by the starting scheme associated with the DuFort–Frankel scheme rapidly become negligible compared with other errors accumulated over time.

(f) Assuming that the spatial mesh intervals and time-step sizes are chosen such that the total operation counts are equal among them, it is shown that the FTCS scheme is less accurate than the μ scheme and the decoupled DuFort–Frankel scheme.

(g) With Δt and Δx being related by Eq. (4.13), both the decoupled DuFort–Frankel scheme and the μ scheme effectively achieve an accuracy that is fourth-order in Δx.

(h) The a–ϵ scheme described in [2] was extended to become a family of solvers for the inviscid Burgers equations. It was shown that (i) the simplest among these solvers is capable of generating nearly perfect shock solutions for the inviscid Burgers equation; and (ii) the simple extrapolation conditions Eq. (4.17) are excellent nonreflecting boundary conditions if they are applied in conjunction with the CE/SE method.

REFERENCES

1. S. C. Chang and W. M. To, *A New Numerical Framework for Solving Conservation Laws—The Method of Space–Time Conservation Element and Solution Element*, NASA TM 104495, August 1991.
2. S. C. Chang, The method of space–time conservation element and solution element—A new approach for solving the Navier–Stokes and Euler equations, *J. Comput. Phys.* **119**, 295–324 (1995).
3. S. C. Chang, X. Y. Wang, and C. Y. Chow, The space–time conservation element and solution element method: A new high-resolution and genuinely multidimensional paradigm for solving conservation laws, *J. Comput. Phys.* **156**, 89–136 (1999).

4. X. Y. Wang and S. C. Chang, A 2D non-splitting unstructured triangular mesh Euler solver based on the space–time conservation element and solution element method, *Comput. Fluid Dyn. J.* **8**(2), 309–325 (1999).
5. S. C. Chang, S. T. Yu, A. Himansu, X. Y. Wang, C. Y. Chow, and C. Y. Loh, The method of space–time conservation element and solution element—A new paradigm for numerical solution of conservation Laws, *in* "Computational Fluid Dynamics Review 1998," edited by M. M. Hafez and K. Oshima (World Scientific, Singapore, 1998), Vol. 1.
6. T. Molls and F. Molls, Space–time conservation method applied to Saint Venant equations, *J. Hydraul. Eng.* **124**(5), 501 (1998).
7. C. Zoppou and S. Roberts, Space–time conservation method applied to Saint Venant equations: A discussion, *J. Hydraul. Eng.* **125**(8), 891 (1999).
8. S. C. Chang, *New Developments in the Method of Space–Time Conservation Element and Solution Element—Applications to the Euler and Navier–Stokes Equations*, NASA TM 106226, August 1993.
9. S. C. Chang, X. Y. Wang, C. Y. Chow, and A. Himansu, The Method of space–time conservation element and solution element—Development of a new implicit solver, in *Proceedings of the Ninth International Conference on Numerical Methods in Laminar and Turbulent Flow, Atlanta, July 10–14, 1995*. [Also published as NASA TM 106897]
10. S. C. Chang, A. Himansu, and X. Y. Wang, Implicit space–time conservation element and solution element schemes, submitted for publication.
11. S. C. Chang, A. Himamsu, C. Y. Loh, X. Y. Wang, S. T. Yu, and P. Jorgenson, Robust and simple non-reflecting boundary conditions for the space–time conservation element and solution element method, *in* "A Collection of Technical Papers, 13th AIAA CFD Conference, Snowmass, Colorado, 1997," AIAA Paper 97-2077.
12. S. T. Yu and S. C. Chang, Applications of the space–time conservation element/solution element method to unsteady chemically reactive flows, *in* "A Collection of Technical Papers, 13th AIAA CFD Conference, Snowmass, Colorado, 1997," AIAA Paper 97-2099.
13. C. Y. Loh, L. S. Hultgren, and S. C. Chang, *Computing Waves in Compressible Flow Using the Space–Time Conservation Element and Solution Element Method*, AIAA Paper 98-0369.
14. S. T. Yu, S. C. Chang, P. Jorgenson, S. J. Park, and M. C. Lai, Eds., Treating stiff source terms in conservation laws by the space–time conservation element and solution element method, *in* "Proceedings of the 16th International Conference on Numerical Method in Fluid Dynamics, Arcachon, France," 1998.
15. C. Y. Loh, L. S. Hultgren, and S. C. Chang, *Vortex Dynamics Simulation in Aeroacoustics by the Space–Time Conservation Element and Solution Element Method*, AIAA Paper 99-0359.
16. N. S. Liu and K. H. Chen, *Flux: An Alternative Flow Solver for the National Combustion Code*, AIAA Paper 99-1079.
17. A. Onorati, G. Ferrari, and G. D'Errico, Fluid dynamics modeling of the gas flow with chemical specie transport through the exhaust manifold of a four cylinder SI engine, presented at the SAE Int. Congress & Exp., Detroit, Michigan, March 1999, Paper 1999-01-0557.
18. X. Y. Wang and S. C. Chang, A 3D structured/unstructured Euler solver based on the space–time conservation element and solution element method, *in* A Collection of Technical Papers, 14th AIAA CFD Conference, *Norfolk, VA, 1999*, AIAA Paper 99-3278.
19. Z. C. Zhang, S. T. Yu, S. C. Chang, A. Himansu, and P. Jorgenson, A modified space–time CE/SE method for solving Euler and Navier–Stokes equations, *in* "A Collection of Technical Papers, 14th AIAA CFD Conference, *Norfolk, VA, 1999*," AIAA Paper 99-3277.
20. K. H. Chen and N. S. Liu, *Navier–Stokes Solution of the Flux Code: A Module for the NCC Solver Using the Concept of Conservation Element and Solution Element Method*, AIAA Paper 2000-0455.
21. Other CE/SE references available at http://www.grc.nasa.gov/www/ microbus.
22. R. D. Richtmyer and K. W. Morton, *Difference Methods for Initial-Value Problems*, 2nd ed., Interscience, New York, 1967.
23. D. A. Anderson, J. C. Tannehill, and R. H. Pletcher, *Computational Fluid Mechanics and Heat Transfer* (Hemisphere, New York, 1984).
24. K. A. Hoffmann and S. T. Chiang, *Computational Fluid Dynamics for Engineers* (Engineering Education System, Wichita, KS, 1993), Vol. 1.