Ann. nucl. Energy, Vol. 11, No. 12, pp. 629–646, 1984
Printed in Great Britain. All rights reserved

0306-4549/84 $3.00 + 0.00

COMPARISON OF FINITE-DIFFERENCE AND VARIATIONAL SOLUTIONS TO ADVECTION–DIFFUSION PROBLEMS

C. E. LEE and K. E. WASHINGTON
Department of Nuclear Engineering, Texas A&M University, College Station, TX 77843, U.S.A.

(*Received* 4 *June* 1984)

Abstract—Two numerical solution methods are developed for 1-D time-dependent advection–diffusion problems on infinite and finite domains. Numerical solutions are compared with analytical results for constant coefficients and various boundary conditions. A finite-difference spectrum method is solved exactly in time for periodic boundary conditions by a matrix operator method and exhibits excellent accuracy compared with other methods, especially at late times, where it is also computationally more efficient. Finite-system solutions are determined from a conservational variational principle with cubic spatial trial functions and solved in time by a matrix operator method. Comparisons of problems with few nodes show excellent agreement with analytical solutions and exhibit the necessity of implementing Lagrangian conservational constraints for physically-correct solutions.

INTRODUCTION

Considerable investigation of efficient numerical calculational methods for advection and diffusion processes has occurred over the past several decades. These processes have been modeled by the initial-value problem

$$\partial\phi/\partial t + \partial[U(x)\phi]/\partial x - \partial[D(x)\partial\phi/\partial x]/\mathrm{d}x + \Sigma(x)\phi = S(x), \quad (1)$$

where ϕ is a field variable (e.g. isotope concentration), U is an advection velocity, D is a diffusion coefficient, Σ is a decay constant or removal cross section and S is a source. Various limiting forms are applicable to the modeling of neutron diffusion, fission-product transport, conduction of thermally-induced waves, porous media mass transport, traffic flow on congested highways, viscous fluid flow, weather front propagation and charged-particle diffusion in an electric field. Although analytic solutions are readily developed for constant coefficients, boundary conditions and simple initial conditions, problems with spatial-coefficient dependency and non-linear extensions are more difficult (or practically impossible) to obtain. These limitations provide a major stimulus for development of more accurate and efficient solution methods so that non-linear extensions can be considered. Significant numerical modeling difficulties still arise with many classical numerical schemes in determining the late-time solutions of the hyperbolic pure advection from

$$\partial\phi/\mathrm{d}t + \partial[U(x)\phi]/\partial x = 0. \quad (2)$$

Finite-difference (FD) and variational methods have been applied extensively to the advection–diffusion problem. General reviews of low-order FD methods have been presented by Roach (1976) and Turkell (1980). A Taylor series approximation of the derivatives represents a straightforward approach. Numerous methods for the pure advection problem were compared by Stuhmiller and Ferguson (1979). Although implementation of low-order FD methods for equation (2) is relatively easy and seemingly requires minimal central processing unit (CPU) investment per time step, computational results frequently exhibit severe amplitude dispersion and phase errors at late times ($> 100\ \Delta t$). This can be understood from a von Neumann stability analysis of the amplitude and phase error (Richtmeyer, 1957). Although lower amplitude errors apparently result in applications of low-order FD methods to equation (1) rather than to equation (2), this may be somewhat deceptive since a fictitious numerical diffusion in the advection term is not readily discernible. Reid (1980) reported studies of various explicit, implicit, single-level and multilevel FD methods applied to equation (2) with periodic boundary conditions and triangular initial conditions. The implicit and multilevel schemes generally were superior to the explicit and single-level schemes. Acceptable errors are obtained only with large numbers of nodes and small time steps. Considerable success was achieved by Reid (1980) in solving equation (2) with the 'filter' method, where the interpolating function includes many nodes rather than just two. Applications of the method of characteristics to advection are described by Gardner *et al.* (1964). The classical Lax–Wendroff (1960) approach was altered by Fromm (1968), who noted that major difficulties were caused by a severe phase lag for non-unity Courant numbers.

Although Fromm's modifications showed marked improvement in phase lag, the method still suffered from amplitude error and consequently gave poor late-time solutions. Shapiro (1975) extended the FD equations to fourth order in space and time. The second-order scheme exhibited significant improvement over the first-order Lax–Wendroff method, but the fourth-order approximation was required for further improvement. Explicit schemes in that analysis failed at late times and implicit schemes tended to consume excessive computer time for acceptable accuracy. Cubic and quintic pseudo-upstream difference equations were developed by Davies (1980) for convective–diffusion modeling. Predictable improvement in amplitude dispersion and phase error over FD methods was reported. Reduced storage requirements were achieved compared to more complicated FD methods. Chan (1978) developed some very promising implicit FD schemes for modeling pure advection based on the balance expansion technique (BETA). Although a relatively large CPU investment for accurate late-time solutions seems to be requisite, the BETA methods have been extended to solve the steady-state version of equation (1).

Numerous investigations based on variational methods have been performed. Price *et al.* (1968) developed a Galerkin variational numerical solution to the advection–diffusion problem without sources or absorption. Linear, cubic and quintic trial functions were utilized. Linear trial functions yielded solutions accuracies comparable to the BETA FD methods. Further accuracy improvement occurs with cubic or quintic trial functions when used in conjunction with a moving node scheme. The method of Price *et al.* involves time discretization which made late-time solutions computationally costly to obtain. Guymon (1970) solved the same problem by transforming the advection–diffusion problem to a diffusion expression with subsequent application of the Ritz method. He reported results for constant coefficients and linear trial functions. Apparently the transformations used produce erroneous results for large values of UL/D (Guymon, 1972). Kermadis (1980) developed a unified finite-element method (FEM) for solving equation (1) with linear and cubic trial functions. In agreement with Price *et al.*, the Kermadis cubic model gave more accurate results than the linear mode for the same CPU investment. Hennart (1979) considered semidiscrete Galerkin techniques in conjunction with finite spatial elements to discretize the space–time reactor kinetics equations. Lee and Wilson (1981, 1984) developed a conservational variational approach for the time-dependent multicomponent diffusion equation, but neglected the advection term. Significantly improved computational speed was demonstrated for a specified accuracy over previous FD formulations by Apperson *et al.* (1979) in applications by Horton (1980) to fission-product release from pebble bed fuel elements.

In this paper we develop two numerical solution methods. First, however, we outline the time-dependent analytic solutions which serve to benchmark the schemes on infinite and finite spatial domains and allow direct pointwise relative error comparisons to be made. Then, we develop a high-order FD method for the advection and advection–diffusion problem on the infinite domain, and a conservational variation principle with advection and diffusion on finite domains. In both numerical methods only the spatial function is approximated either by FD or cubic trial functions. The time variable is maintained continuous, and the resulting matrix equations have the form

$$\mathrm{d}\mathbf{X}/\mathrm{d}t = \mathbb{A}\mathbf{X}+\mathbf{S}, \tag{3}$$

where $\mathbf{X}$ and $\mathbf{S}$ are vectors (time-dependent nodal values) and $\mathbb{A}$ is a matrix. This equation is readily solved in exponential matrix form with the ASH program (Lee, 1980), when $\mathbb{A}$ is constant over the time interval considered. Consequently, numerical errors are predominantly due to spatial discretization and coupling since the matrix solutions are essentially 'exact' (to within computer significance) in time.

ANALYTIC SOLUTIONS

The non-dispersive pure advection process described by the hyperbolic partial differential equation, equation (2), is frequently investigated on an infinite domain with constant advection velocity. The infinite domain can be simulated with periodic boundary conditions on a finite mesh $[0, L]$ by

$$\phi(0,t) = \phi(L,t). \tag{4}$$

Comparisons for an arbitrary initial condition

$$\phi(x,0) = F(x) \tag{5}$$

will be made. Since the Fourier component $\exp[-i\xi(x-Ut)]$, $i=\sqrt{-1}$, satisfies equation (2) for any wavenumber ξ, application of the Fourier transform and the initial condition gives

$$\phi(x,t) = [1/(2\pi)]\int_{-\infty}^{\infty}\mathrm{d}x'F(x')\cdot\int_{-\infty}^{\infty}\exp\{i[x'-(x-Ut)]\xi\}\,\mathrm{d}\xi, \tag{6}$$

which reduces to

$$\phi(x,t) = \int_{-\infty}^{\infty}\mathrm{d}x'\,F(x')\delta(x'-(x\text{–}Ut)) \tag{7}$$

or

$$\phi(x, t) = F(x-Ut), \qquad (8)$$

using the integral definition of the Dirac delta function $\delta(x)$. This simple analytic pure advection solution, a translation of the initial condition along the lattice at speed U, is useful for numerical comparisons with a variety of initial conditions.

The advection–diffusion solution of equation (1) with constant coefficients is conveniently expressed in terms of dimensionless groups. With the definitions $\lambda = UL/D$, $\tau = Dt/L^2$, $Q = SL^2/D$, $\eta = \Sigma L^2/D$ and $x \to x/L$, equation (1) takes the form

$$\partial\phi/\partial T + \lambda\,\partial\phi/\partial x - \partial^2\phi/\partial x^2 + \eta\phi = Q. \qquad (9)$$

Two solutions are considered. First, we construct the zero source ($Q = 0$) solution on the infinite domain, and then the finite-domain constant source solution with general boundary conditions.

For a zero source the Fourier component $\exp[-i(x-hT)-(\xi^2+\eta)\tau]$ satisfies equation (9) for any wavenumber ξ. The analytic solution is formed from

$$\phi(x, T) = (1/2\pi)\,\mathrm{e}^{-\eta\tau}\int_{-\infty}^{\infty} \mathrm{d}x'\,F(x')$$
$$\cdot\int_{-\infty}^{\infty} \mathrm{d}\xi\,\exp[-i\xi(x-x'-\lambda\tau)-\xi^2\tau]. \qquad (10)$$

Completing the square on the exponential argument and evaluating the resulting error function yields

$$\phi(x, T) = [1/(2(\pi\tau)^{1/2})]\,\mathrm{e}^{-\eta\tau}\int_{-\infty}^{\infty} \mathrm{d}x'$$
$$\cdot F(x')\exp[-(x-x'-\lambda\tau)^2/4\tau], \qquad (11)$$

where the initial condition, $F(x)$, must still be specified. For numerical comparisons we assume a Gaussian initial condition given by

$$F(x) = \alpha/(2\pi)^{1/2}\exp[-(x-x_0)^2/2\sigma^2], \qquad (12)$$

where $\alpha = 2[\ln(4)]^{1/2}$ and σ = full width at half maximum (FWHM)/α. Performing the integration with equation (12) gives

$$\phi(x, T) = (\alpha/2)[2\pi\tau A(\tau)]^{1/2}\exp[B^2(x,\tau)/4A(\tau)$$
$$-C(x,\tau)-\eta\tau], \qquad (13)$$

where

$$A(\tau) = (2\tau+\sigma^2)/(4\tau\sigma^2),$$

$$B(x,\tau) = 2[2\tau x_0+\sigma^2(x-\lambda\tau)]/(4\tau\sigma^2)$$

and

$$C(x,\tau) = [2\tau x_0^2+\sigma^2(x-\lambda\tau)^2]/(4\tau\sigma^2). \qquad (14)$$

The analytic solution is not equivalent to the finite-domain problem with periodic boundary conditions because the mass diffusing beyond the subdomain boundaries $[0, 1]$ is lost in the infinite case, but reappears at the other boundary in the periodic case. Thus, for numerical comparisons, this result is valid only at times before diffusion occurs to the mesh edges $[0, 1]$.

The analytic solution of equation (9) for finite domain $[0, 1]$ with constant coefficients can be determined for the general boundary conditions

$$\mathrm{a}_1\phi(0, T)+\mathrm{a}_2\,\partial\phi(0, T)/\partial x = \mathrm{a}_3$$

and (15)

$$\mathrm{b}_1\phi(1, T)+\mathrm{b}_2\,\partial\phi(1, T)/\partial x = \mathrm{b}_3,$$

where the a_i and b_i, $1 \leqslant i \leqslant 3$, are arbitrary constants. Considerable solution simplification results from assuming zero initial conditions.

Performing the Laplace transformation of equations (9) and (15) for a constant source Q_0 and zero initial conditions results in

$$\partial^2\theta/\partial x^2 - \lambda\,\partial\theta/\partial x - (\eta+s)\theta = -Q_0/s,$$

$$\mathrm{a}_1\theta(0, s)+\mathrm{a}_2\,\partial\theta(0, s)/\partial x = \mathrm{a}_3/s$$

and (16)

$$\mathrm{b}_1\theta(1, s)+\mathrm{b}_2\,\partial\theta(1, s)/\partial x = \mathrm{b}_3/s,$$

where $\theta(x, s) = \mathscr{L}[\phi(x, t)]$, the Laplace transform of $\phi(x, t)$. The solution is

$$\theta(x, s) = \{\mathrm{e}^{-\lambda/2}/[2s\gamma(s)]\}$$
$$\cdot\{[\mathrm{b}_3-Q_0/(\eta+s)](\mathrm{a}_1+\mathrm{a}_2r_2)$$
$$-[\mathrm{a}_3-Q_0/(\eta+s)](\mathrm{b}_1+\mathrm{b}_2r_2)\exp(r_2)\}$$
$$\cdot\exp(r_1x)\{[\mathrm{a}_3-Q_0/(\eta+s)](\mathrm{b}_1+\mathrm{b}_2r_1)\exp(r_1)$$
$$-[\mathrm{b}_3-Q_0/(\eta+s)](\mathrm{a}_1+\mathrm{a}_2r_1)\}$$
$$\cdot\exp(r_2x)+Q_0/[s(\eta+s)], \qquad (17)$$

where

$$r_1(s) = \lambda/2+\mu(s),$$

$$r_2(s) = \lambda/2-\mu(s),$$

$$\mu(s) = [(\lambda/2)^2+\eta+s]^{1/2},$$

$$Y(s) = (K_1-\mathrm{a}_2\mathrm{b}_2\mu^2)\sinh(\mu)-K_2\cosh(\mu),$$

$$K_1 = \mathrm{a}_1\mathrm{b}_1+(\lambda/2)(\mathrm{a}_2\mathrm{b}_1+\mathrm{a}_1\mathrm{b}_2)+(\lambda/2)^2\mathrm{a}_2\mathrm{b}_2$$

and

$$K_2 = \mathrm{a}_2\mathrm{b}_\mathrm{d}-\mathrm{a}_1\mathrm{b}_2. \qquad (18)$$

The inverse Laplace transform is obtained by the Cauchy residue theorem upon summing over the poles of $\theta(x, s)$ in equation (17). The apparent singularity at

$s+\eta = 0$ gives zero residue contribution. Steady-state solutions result from $s = 0$ poles and the transient contribution arises from the infinite set of roots of the transcendental equation $\gamma(s) = 0$. The residue terms are evaluated by a (complex arithmetic) computer program to achieve convergence for specified accuracy of the real function $\phi(x, t)$. These analytical evaluations were used in the comparisons described below.

FD ADVECTION–DIFFUSION METHODS

Numerous FD schemes have been developed for the pure advection problem (Roach, 1976). We briefly summarize the general first order Lax–Wendroff FD method for later comparisons. The forward time-centered scheme with constant advection velocity is given by (Nakamura, 1977)

$$(\phi_i^{n+1}-\phi_i^n)/\Delta t = \beta U(\phi_{i+1}^{n+1}-\phi_{i-1}^{n+1})/(2\Delta x)$$
$$+(1-\beta)U(\phi_{i-1}^n-\phi_{i-1}^n)/(2\Delta x), \quad (19)$$

where $\phi_i^n = \phi\ (i\Delta x, n\Delta t)$, $1 \leqslant i \leqslant I$, the number of spatial nodes. This expression can be derived from a Taylor series expansion of ϕ_{i+1} and ϕ_{i-1} by retaining only the first two terms. The parameter β denotes the implicitness with the most widely used values being 0.0, 0.5 and 1.0, corresponding to explicit, time averaged (centered) and fully implicit, respectively.

A potential computational advantage exists if the spatial derivatives are discretized while the time variable is retained as continuous. With this assumption the Lax–Wendroff method applied to the pure advection problem on a periodic spatial finite mesh takes the form

$$\mathbb{T}\ \mathrm{d}\boldsymbol{\phi}/\mathrm{d}t = \mathbb{A}\boldsymbol{\phi}, \quad (20)$$

where

$$\boldsymbol{\phi} = (\phi_1, \phi_2, \ldots, \phi_I)^{\mathrm{T}} \quad (21)$$

and the matrix elements of $\mathbb{T}$ and $\mathbb{A}$ are given by

$$\begin{aligned} T_{i,i} &= 1.0,\ 1 \leqslant i \leqslant I, \\ A_{i,i} &= 0.0,\ 1 \leqslant i \leqslant I, \\ A_{i,i+1} &= -U/2\Delta,\ 1 \leqslant i \leqslant I-1, \\ A_{i,i-1} &= U/2\Delta,\ 2 \leqslant i \leqslant I, \quad (22) \\ A_{1,I} &= U/2\Delta \end{aligned}$$

and

$$A_{I,1} = -U/2\Delta.$$

The two off-tridiagonal terms $A_{1,I}$ and $A_{I,1}$ result from the applied periodicity boundary conditions. If the matrix $\mathbb{B} = \mathbb{T}^{-1}\mathbb{A}$ is constant in time, as assumed, the solution to equation (20) is

$$\boldsymbol{\phi}(t) = \exp(\mathbb{B}t)\boldsymbol{\phi}(0), \quad (23)$$

where $\boldsymbol{\phi}(0)$ is the initial condition vector. The exponential matrix is evaluated using the ASH program technique, as summarized in the Appendix. Although the solution method is essentially 'exact' in time, this spatial approximation is still too low an order and insufficiently coupled to obtain accurate late-time solutions. Mitigation of this difficulty is obtained by inclusion of higher-order spatial terms. One could derive a general Nth-order FD scheme based on the Taylor series approach, but the algebra tends to rapidly become intractable, and, besides, this direct approach does not necessarily yield an optimal coupling in space and time. An alternative approach is therefore followed, whereby the desired Nth-order accuracy is obtained for both the pure advection and the advection–diffusion problems.

The advection–diffusion FD derivation is simplified to constant coefficients and equispaced nodes. Defining E as the wavenumber, the spatial mesh separation $\Delta = x_{i+1}-x_i$, $\theta = \xi\Delta$, the phase angle, $\phi^{(n)} = \partial n\phi/\partial x^n$, and $\psi = \partial\phi/\partial t$, equations (2) and (9) may be written as

$$\psi + U\phi' = 0 \quad (24)$$

and

$$\psi - \phi'' + \lambda\phi' + \eta\phi = 0, \quad (25)$$

respectively. Expanding ϕ_{i+1} and ϕ_{i-1} about x_i, the pure advection problem, equation (24), takes the form

$$\psi_i + (U/2\Delta)(\phi_{i+1}-\phi_{i-1}) - S_i = 0, \quad (26)$$

where

$$S_i = U\sum_{k=1}^{\infty} \Delta^{2k}\phi_i^{(2k+1)}/(2k+1)!. \quad (27)$$

The approximation $S_i = 0$ in equation (26) corresponds to the first-order Lax–Wendroff scheme, and S_i represents the higher-order terms usually neglected. For equation (25), if we approximate ϕ' and ϕ'' with the Taylor series expansion, the advection–diffusion problem can be written as

$$\psi_i - (\phi_{i-1} - 2\phi_i + \phi_{i-1})/$$
$$\Delta^2 + (\lambda/2\Delta)(\phi_{i+1}-\phi_{i-1}) + \eta\phi_i + S_i = 0, \quad (28)$$

where

$$S_i = \lambda\sum_{k=1}^{\infty} \Delta^{2k}\phi_i^{(2k+1)}/(2k+1)! - 2$$
$$\cdot \sum_{k=1}^{\infty} \Delta^{2k}\phi_i^{(2k+2)}/(2k+2)! \quad (29)$$

The first term in equation (29) corresponds to the advection contribution and the second term arises from the diffusion contribution.

In order to approximate S_i uniformly, we assume a finite expansion form in terms of ϕ and $\psi = \partial\phi/\partial t$ given by

$$S_i = C_0\psi_i + \sum_{k=1}^{M} C_k(\psi_{i+k} + \psi_{i-k})$$

$$+ A_0\phi_i + \sum_{k=1}^{M} A_k(\phi_{i+k} + \phi_{i-k}), \quad (30)$$

where $M \leqslant (I-1)/2$. Substituting the discrete Fourier component in the assumed form for S_i and equation (28), using the MacLaurin series expansion for the sine and cosine of the phase angle, θ, and equating real and imaginary parts (after algebraic manipulation), we obtain the spectral expressions

$$C_0 + 2\sum_{k=1}^{M} C_k \cos(k\theta) = 1 - (1/\theta)\sin(\theta)$$

and

$$A_0 + 2\sum_{k=1}^{M} A_k \cos(k\theta) = 2\xi^2\{-1/2 + [1-\cos(\theta)]/\theta^2\}$$

$$+ (\xi^2 + \eta)[1 - (1/\theta)\sin(\theta)]. \quad (31)$$

The coefficients C_k arise from advection and the A_k are due to diffusion. In order to determine these coefficients, we select $M \leqslant (I-1)/2$ phase angles θ which best approximate the spectral expressions. Given a function $f(\theta)$ defined on $[0, n]$, the approximation

$$f(\theta) = B_0 + 2\sum_{k=1}^{N-1} B_k \cos(k\theta) \quad (32)$$

has minimal error with the coefficients

$$B_k = (1/N)\sum_{i=1}^{N} f(\theta_i)\cos(k\theta_i), \quad 0 \leqslant k \leqslant N-1, \quad (33)$$

for the phase angle choice

$$\theta_i = \cos[(2i-1)\pi/2N], \quad 1 \leqslant i \leqslant N, \quad (34)$$

by the Chebyshev summation coefficient theorem (Dodes, 1978). Thus, the coefficients C_k and A_k in equation (31) can be evaluated explicitly and optimally.

Writing the resultant matrix approximation in the form of equation (20), we identify the interior elements of $\mathbb{T}$ and $\mathbb{A}$ as

$$\begin{aligned} T_{i,i} &= 1 - C_0, \\ T_{i,i+k} &= T_{i,i-k} = -C_k, \\ A_{i,i} &= A_0 - n - 2/\Delta^2, \\ A_{i,i+1} &= A_1 + 1/\Delta^2 - \lambda/2\Delta, \\ A_{i,i-1} &= A_1 + 1/\Delta^2 + \lambda/2\Delta \end{aligned} \quad (35)$$

and

$$A_{i,i+k} = A_{i,i-k} = A_k.$$

Periodic boundary conditions are readily implemented by imposing a periodicity mapping of these matrix elements.

Equations (20), (22a,b) and (35) constitute the basis of the SPECTRUM method upon which comparisons are reported below. Several features of this advection–diffusion approximation are noteworthy. The pure advection case is obtained if the equations are written in dimensional form and the diffusion coefficient D, and removal cross section, Σ, are set to zero. Secondly, the centered first-order Lax–Wendroff spatial FD method is recovered with the coefficients A_k and C_k set to zero. An ASH FD method produces results for advection–diffusion similar to the pure advection problem discussed previously. Finally, since the matrix coefficients of $\mathbb{T}$ and $\mathbb{A}$ were derived for periodic boundary condition problems, direct extension to finite-medium problems would result in a significant loss of nodal coupling near the boundaries. Such extensions require separate considerations.

CONSERVATIONAL VARIATIONAL METHOD

Although the SPECTRUM method developed above results in significantly improved solutions for a limited class of infinite-domain problems, as simulated by periodic boundary conditions, it is not particularly well-suited for finite-domain problems. This is partially due to the implementation of the infinite-medium Fourier component in the matrix coefficient derivation and partially because a significant reduction in nodal coupling would occur near the finite-domain boundaries without additional assumptions. The method seems to be reasonably difficult to extend with high-order accuracy to non-constant coefficient problems or mixed boundary conditions.

Instead of directly implementing standard FEM methods (Chung, 1978) with upwind differencing, we considered the problem from the basic viewpoint of variational calculus and the Euler–Lagrange formulation. The conservational variational principle (CVP) method results. A functional

$$G(\psi,(*) = \int_0^T \mathrm{d}t \int_0^L \mathrm{d}x\, L(\psi,\psi^*,\psi',\psi'^*,\nabla\psi,\nabla\psi^*) \tag{36}$$

in terms of the function of ψ, its adjoint ψ^*, time derivatives ψ' and ψ'^*, and spatial gradients $\nabla\psi$ and $\nabla\psi$, is stationary provided the Euler–Lagrange equations (Morse and Feshbach, 1953; Lanczos, 1949)

$$\partial/\partial x\,[\partial L/\partial(\nabla\psi^*)] + \partial/\partial t\,[\partial L/\partial\psi'^*] - \partial L/\partial\psi^* = 0$$

and (37)

$$\partial/\partial x\,[\partial L/\partial(\nabla\psi)] + \partial/\partial t\,[\partial L/\partial\psi] - \partial L/\partial\psi = 0$$

are satisfied, and reproduce the equations to be solved, equation (1). The choice

$$L(\psi,\psi^*,\psi',\psi'^*,\nabla\psi,\nabla\psi^*) = \psi\psi^{*\prime} + U(x)\psi\nabla\psi^*$$
$$-D(x)\nabla\psi\cdot\nabla\psi^* - \Sigma(x)\psi\psi^* + S(x)\psi^* + S^*(x)\psi \tag{38}$$

yields

$$K\psi = \psi' + \nabla\cdot(U\psi) - \nabla\cdot(D\nabla\psi) + \Sigma\psi - S = 0 \tag{39}$$

and

$$K^*\psi^* = -\psi^{*\prime} - U\nabla\psi^* - \nabla\cdot(D\nabla\psi^*) + \Sigma\psi^* - S^* = 0. \tag{40}$$

Dividing the spatial range into I cells of size $\Delta x_i = x_{i+1} - x_i$, the functions are approximated within each spatial cell. For diffusion solutions the continuity of flux (ψ) and current (J) at interior material interfaces should be satisfied, namely, we impose the solution constraints that

$$\psi_{i-1}(x_{i-},t) = \psi_i(x_{i+},t) \tag{41}$$

and

$$J_{i-1}(x_{i-},t) = J_i(x_{i+},t), \tag{42}$$

where

$$J_i(x,t) = -D_i(x)\nabla\psi_i(x,t). \tag{43}$$

The subscript on ψ and J is a cell index and the subscript on x is a cell boundary index with the ± appendage indicating evaluation from the right (+) or left (−) of the interior boundary. Since there are four continuity conditions on flux and current satisfied in each cell, a unique determination of the four coefficients of a cubic spatial trial function can be made. If the spatial cubic trial function in the ith cell is $\psi_i(x,t)$, application of the continuity conditions yields

$$\psi_i(p,t) = \theta_i(t)s_1(\rho) + \theta_{i+1}(t)s_2(\rho)$$
$$-(\Delta_i/D_{i+})J_i(t)s_3(\rho) - (\Delta_i/D_{i+1-})J^{(t)}_{i+1}s_4(\rho), \tag{44}$$

where

$$\rho = (x - x_i)/\Delta_i, \quad 0 \leqslant \rho \leqslant 1, \tag{45}$$

$$\Delta_i = x_{i+1} - x_i, \quad 1 \leqslant i \leqslant I, \tag{46}$$

and the shape function polynomials $s_k(\rho)$, $1 \leqslant k \leqslant 4$, are determined as

$$s_1(\rho) = (1+2\rho)(1-\rho)^2,$$

$$s_2(\rho) = (3-2\rho)\rho^2,$$

$$s_3(\rho) = \rho(1-\rho)^2,$$

and

$$s_4(\rho) = \rho^2(1-\rho) \tag{47}$$

from the continuity conditions (Hennart, 1973). The adjoint, $\psi_i^*(x,t)$, expressions are obtained similarly, involving the adjoint nodal values, $\theta_i^*(t)$, $\theta_{i+1}^*(t)$, $J_i^*(t)$ and $J_{i+1}^*(t)$, respectively, and the same shape functions, equation (47).

The functional is partitioned into I cellwise components in terms of the time-dependent nodal functions θ_i, θ_i^*, J_i and J_i^*, $1 \leqslant i \leqslant I+1$. We have assumed a spatial trial function form which imposes and satisfies the continuity conditions, equations (41) and (42). In order to guarantee particle and adjoint conservation for each spatial cell, we must add a Lagrangian conservation constraint to the functional involving the spatial integrals of equations (39) and (40), as shown previously for steady-state (Lee *et al*., 1984) and time-dependent (Lee and Wilson, 1984) diffusion solutions. Thus, we consider the spatially discretized functional

$$G(\psi,\psi^*) = G(\boldsymbol{\theta},\boldsymbol{\theta}^*,\mathbf{J},\mathbf{J}^*,\mathbf{E},\mathbf{E}^*)$$
$$= \int_0^T \mathrm{d}t \sum_{i=1}^{I} \left[\Delta_i \int_0^1 L_i(\theta_i,\theta_i^*,J_i,J_i^*)\,\mathrm{d}\rho \right.$$
$$+ E_i^*\Delta_i \int_0^1 K\psi_i(\rho,t)\,\mathrm{d}\rho$$
$$\left. + E_i\Delta_i \int_0^1 K^*\psi_i^*(\rho,t)\,\mathrm{d}\rho \right], \tag{48}$$

where

$$L_i = \psi_i\,\partial\psi_i^*/\partial t + [U_i(\rho)/\Delta_i]\psi_i\,\partial\psi_i^*/\partial\rho$$
$$-[D_i(\rho)/\Delta_i^2]\,\partial\psi_i/\partial\rho\,\partial\psi_i^*/\partial\rho$$
$$-\Sigma_i(\rho)\psi_i\psi_i^* + S_i(\rho)\psi_i^* + S_i^*(\rho)\psi_i \tag{49}$$

and the operators K and K^* are defined by equations (39) and (40).

Substituting the explicit forms for $\psi_i(\rho,t)$ and $\psi_i(\rho,t)$ in terms of θ_i(t), $\theta_i^*(t)$, $J_i(t)$ and $J_i^*(t)$ into equation (49) for L_i and applying Gauss' divergence theorem to $\nabla\cdot(U\psi)$

and $\nabla \cdot (D\nabla\psi)$ in the conservation constraint terms of equation (48), we obtain the functional $G(\boldsymbol{\theta}, \boldsymbol{\theta}^*, \mathbf{J}, \mathbf{J}^*, \mathbf{E}, \mathbf{E}^*)$ which is extremal for the Euler–Lagrange equations. The 'equations of motion' are the time-dependent equations for the nodal flux ($\boldsymbol{\theta}$), current ($\mathbf{J}$) and Lagrange multiplier ($\mathbf{E}$) and adjoints ($\boldsymbol{\theta}^*$, $\mathbf{J}^*$ and $\mathbf{E}^*$). The boundary conditions are imposed by using the time derivative of equation (15) evaluated with the trial function nodal values, namely

$$\partial/\partial t[\mathrm{a}_1\theta_1(t)-(\mathrm{a}_2/D_1)J_1(t)] = 0$$

and (50)

$$\partial/\partial t[\mathrm{b}_2\theta_I(t)-(\mathrm{b}_2/D_I)J_I(t)] = 0,$$

where the initial condition vector for $\boldsymbol{\theta}$ is set to a_3 and b_3, respectively. We find the nodal matrix equations

$$\mathbb{T}\, \mathrm{d}\mathbf{X}/\mathrm{d}t = \mathbb{M}\mathbf{X}+\mathbf{Q}, \tag{51}$$

which are equivalent to equation (3) with the transformation $\mathbb{A} = \mathbb{T}^{-1}\mathbb{M}$ and $\mathbf{S} = \mathbb{T}^{-1}\mathbf{Q}$, where $\mathbb{T}$ is a non-singular symmetric block tridiagonal matrix and the inverse is evaluated numerically (Hornbeck, 1975). We have ordered the elements of $\mathbf{X}$ as $X_i = [\phi_i, J_i, E_i]^{\mathrm{T}}$. As the advection velocity, U, approaches zero, the matrix $\mathbb{M}$ limits to the form derived previously for pure diffusion. The accuracy of the continuous spatial resolution depends upon the number and placement of nodal points.

RESULTS AND COMPARISONS

A number of solutions have been compared analytically and numerically for pure advection and advection–diffusion problems on infinite and finite domains. The infinite-domain problems reported here all had a Gaussian initial condition, $\phi(x, 0) = F(x)$, equation (12), with a 0.1 pulse width, shown in Fig. 1. This initial condition eliminates the Gibb's phenomena standardly encountered from step function initial conditions in numerical and analytical (finite term) solutions. Periodic boundary conditions were imposed on all infinite-medium problems.

The pure advection problem, equation (24), was solved on the subdomain [0, 1]. The ASH FD, BETA and SPECTRUM methods for 31 and 101 nodes are compared with the analytical solutions in Figs 2–7. The solid lines represent the numerical solution, and the dashed lines (if visible) indicate the analytical solution, equation (8). The absolute value of the difference between the numerical and analytical solutions is shown beneath each solution graph. For infinite-medium problems the cycle variable indicates the number of times that the pulse passed through the subdomain [0, 1].

The SPECTRUM method is more accurate than either the ASH FD or BETA methods. This significant improvement is both interesting and important since the BETA method has been reported by Chan (1978) to be better than most other FD methods for modeling advection equations. As the number of nodes is increased from 31 to 101, each method exhibits considerable improvement. However, the SPECTRUM method still has the smallest errors, as is particularly evident at later times (>64 cycles) when most FD methods have failed severely. At 1024 cycles, the SPECTRUM method is still reasonably accurate, as shown in Fig. 7. Since the BETA method was

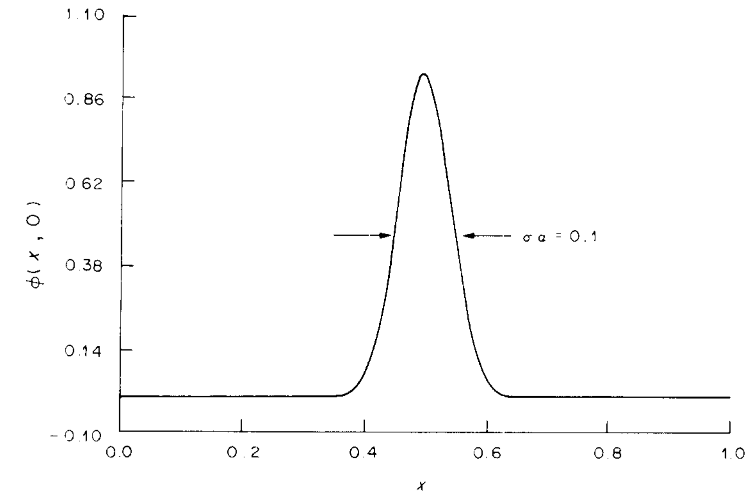

Fig. 1. Gaussian normal distribution initial condition.

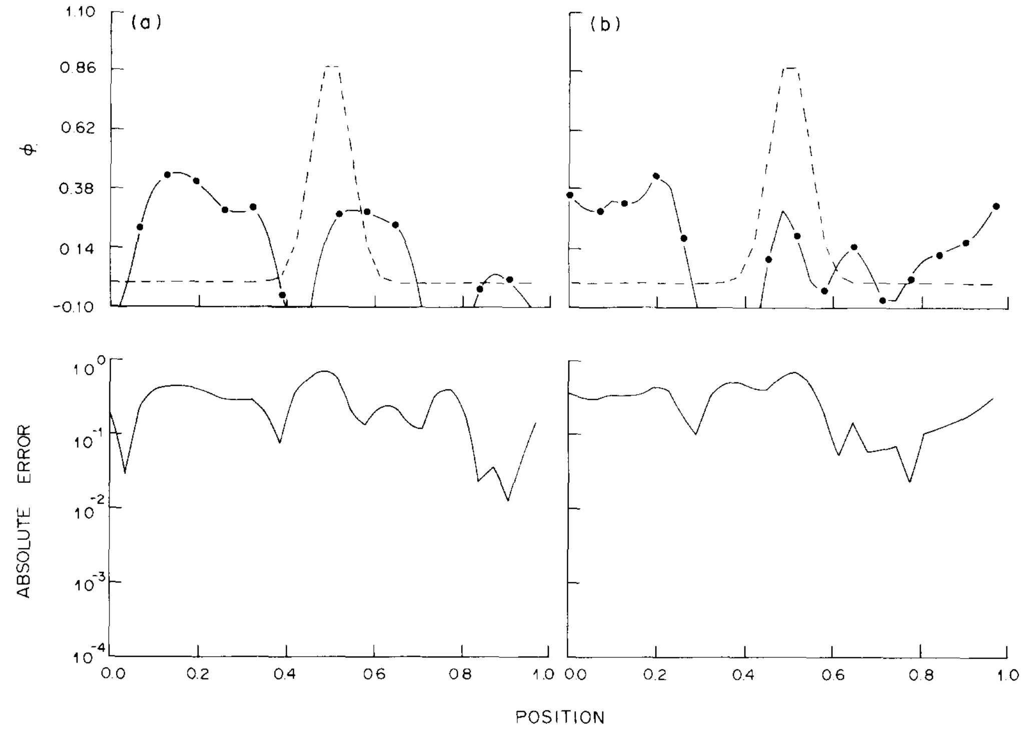

Fig. 2. *Upper curves*: analytic (---) and numerical comparison (●—●) of the ASH FD method vs position for 31 nodes at (a) cycle 32 and (b) cycle 512. *Lower curves*: absolute errors vs position.

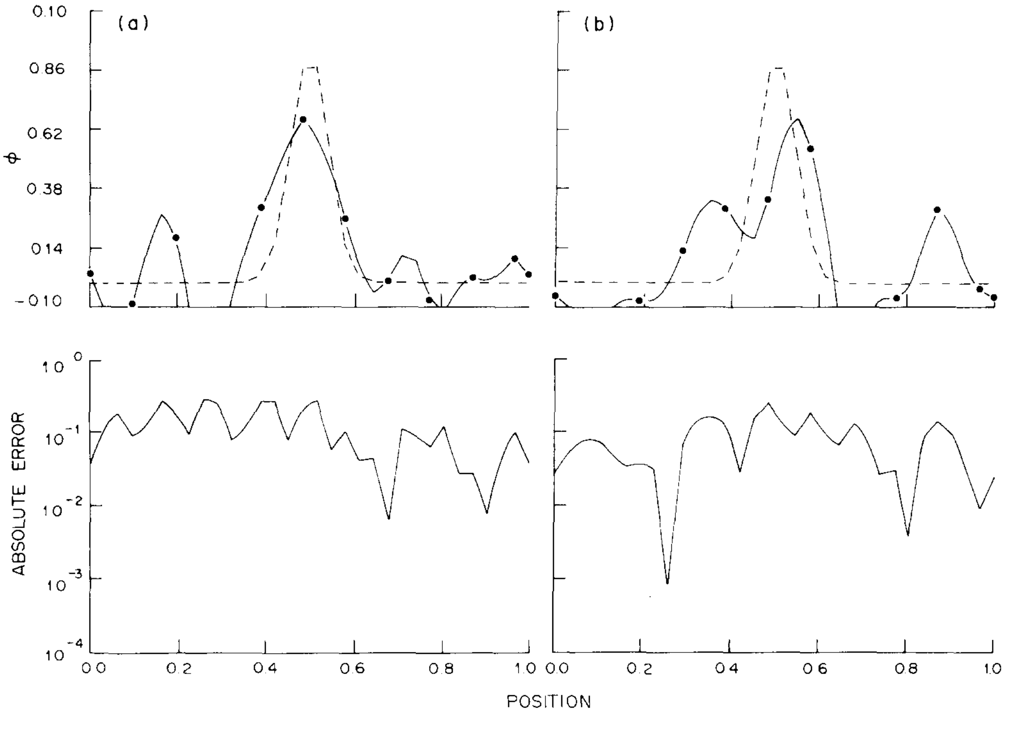

Fig. 3. *Upper curves*: analytic (---) and numerical comparison (●—●) of the BETA FD method vs position for 31 nodes at (a) cycle 32 and (b) cycle 512. *Lower curves*: absolute errors vs position.

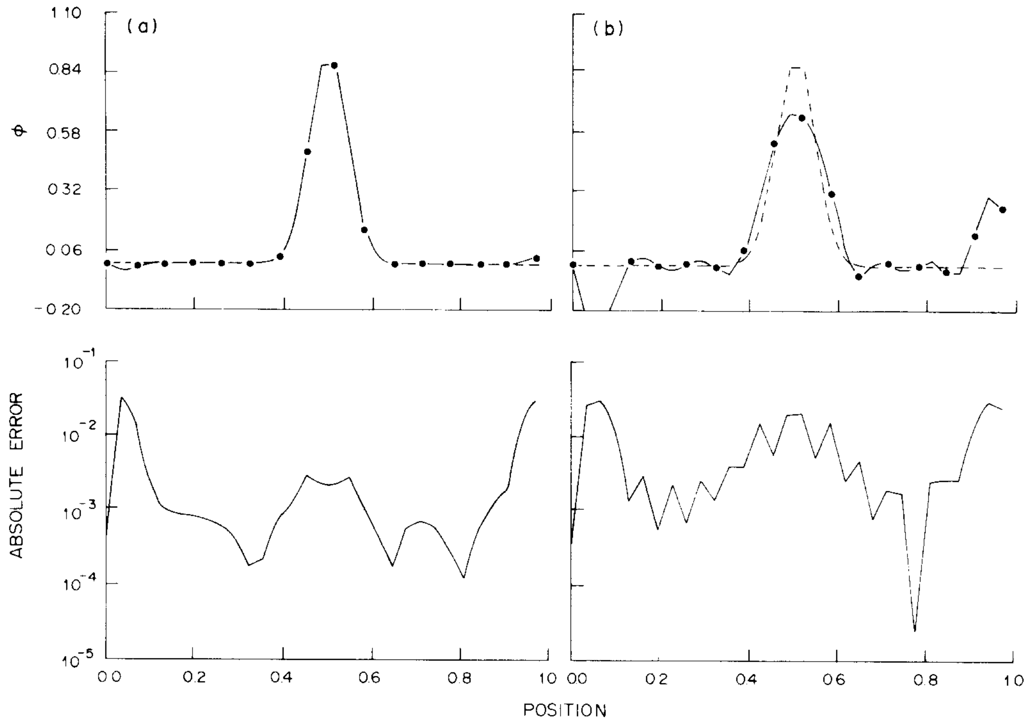

Fig. 4. *Upper curves*: analytic (---) and numerical comparison (●—●) of the SPECTRUM method vs position for 31 nodes at (a) cycle 32 and (b) cycle 512. *Lower curves*: absolute errors vs position.

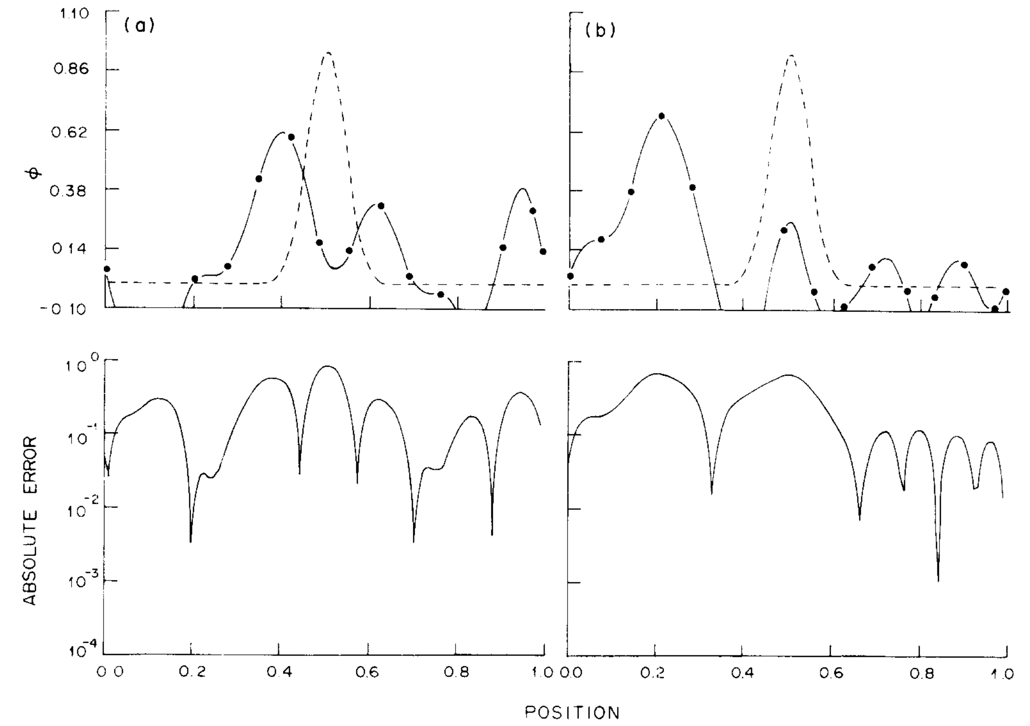

Fig. 5. *Upper curves*: analytic (---) and numerical comparison (●—●) of the ASH FD method vs position for 101 nodes at (a) cycle 32 and (b) cycle 512. *Lower curves*: absolute errors vs position.

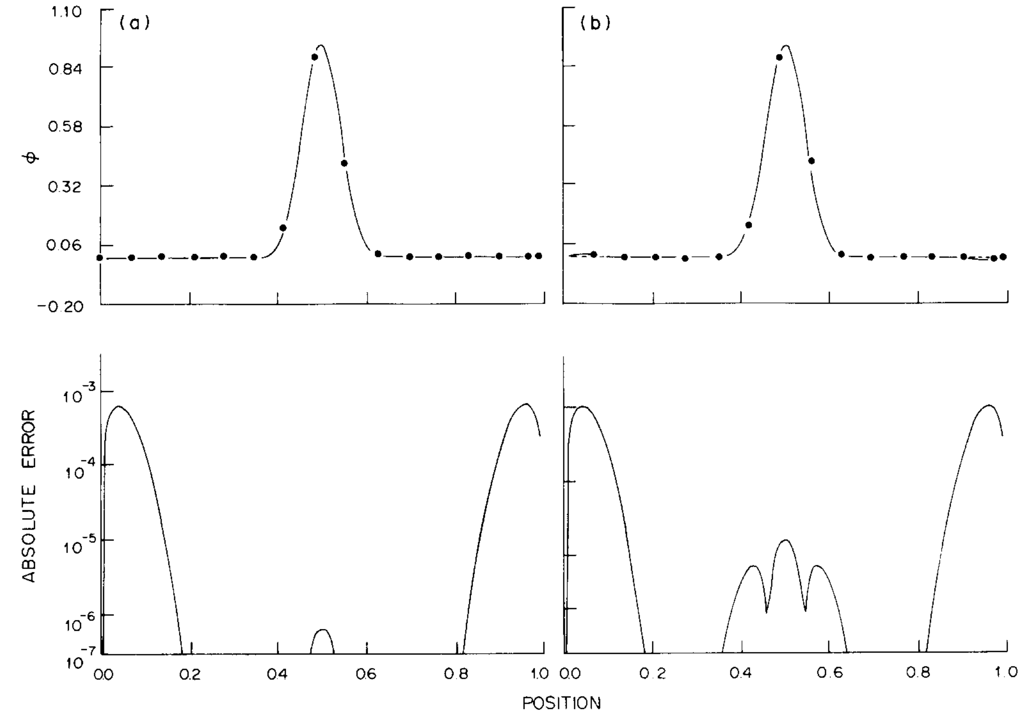

Fig. 6. *Upper curves*: analytic (---) and numerical comparison (●—●) of the SPECTRUM method vs position for 101 nodes at (a) cycle 32 and (b) cycle 512. *Lower curves*: absolute errors vs position.

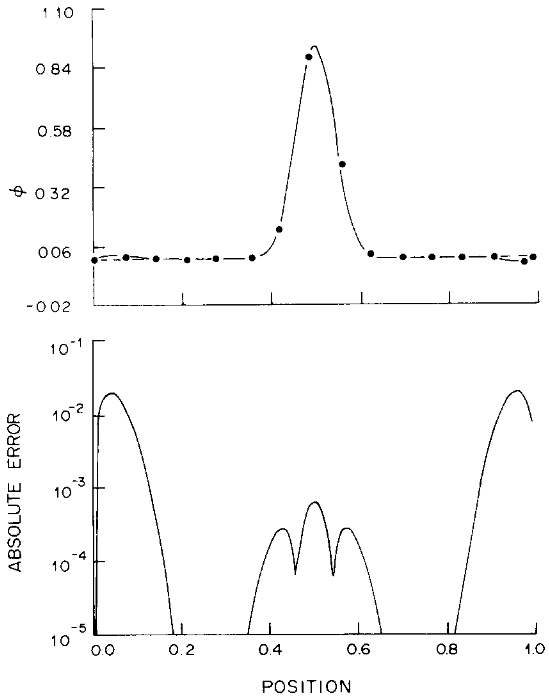

Fig. 7. *Upper curves*: analytic (---) and numerical comparison (●—●) of the SPECTRUM method vs position for 101 nodes at cycle 1024. *Lower curves*: absolute errors vs position.

significantly slower in execution, it was not run to 1024 cycles. The SPECTRUM method accuracy results from the retention of the higher-order terms in the Taylor series expansion and the optimal Chebyshev approximation to the corresponding wavenumber spectrum. As the number of nodes increases, more terms are retained automatically in the $\mathbb{A}$ matrix with the SPECTRUM method, which is probably always more accurate than other FD methods for this problem. The ASH FD and SPECTRUM methods have a distinct advantage over standard discretized FD methods in that late-time solutions are computationally less costly to obtain. This results directly from the ASH time-solution technique, outlined in the Appendix. Obtaining the solution at double the final time requires only an additional matrix multiplication by ASH, but twice the computational effort with standard FD methods to find the solution to double the final time.

The ASH FD, SPECTRUM and CVP methods were compared for the advection–diffusion on the infinite domain,

$$\partial\phi/\partial t+\lambda\,\partial\phi/\partial x-\partial^2\phi/\partial x^2=0,$$

with a Gaussian initial condition and periodic boundary conditions. Numerical solutions can only be compared with the infinite-medium analytic solution, discussed previously, if the initial pulse has not diffused to the boundaries. For this problem, the comparison is valid until approximately cycle 64. The solutions are compared to the analytic solutions at cycles 16 and 64 in Figs 8–10. The FD solution, even when evaluated using ASH, exhibits reasonably large absolute errors, of the order of0.1, as shown in Fig. 8. However, this solution is still an improvement over the usual time-discretized schemes since the spatial equations are solved exactly in time by ASH. The CVP method solves three times as many equations as the FD methods. Consequently significantly longer running times and greater array storage are expected for the same number of nodes. In order to reduce the CPU requirements of the CVP method and adjust its computational times more closely to those of the FD methods, the number of nodes was reduced to 61. Although this might appear to be a significant disadvantage for the CVP method, in fact, the method gives quite reasonable absolute errors, 10^{-3}, at early times, and, 10^{-4}, at late problem times for a comparable CPU investment as shown in Fig. 10. This relative success of the CVP method is a direct result of the piecewise continuous solution and, in particular, the constraint of particle conservation in each computational cell, since both methods used ASH for the time solution. The SPECTRUM method produces an accurate solution with absolute error typically less than about 10^{-4}. This is not particularly surprising since the method is quite high order with 101 nodes and was specifically designed to solve this periodic boundary condition problem. The small absolute errors obtained with the SPECTRUM method on this diffusion problem are also even less than the errors produced in solving the corresponding pure advection problem. This is partially due to the natural error damping introduced by diffusion. Finally, it is noted that there is an increased error near the boundary at cycle 64 which is due to the previously discussed breakdown of the approximate infinite-medium analytic solution.

Even though the SPECTRUM method gives a good solution of the periodic boundary condition problem for a Gaussian initial condition, the CVP method also yields acceptable results, and offers considerable versatility in solving other problems. Application of the SPECTRUM method seems somewhat more difficult for finite-domain problems with non-constant coefficients. However, such problems are readily treated with CVP, where quite acceptable errors are obtained.

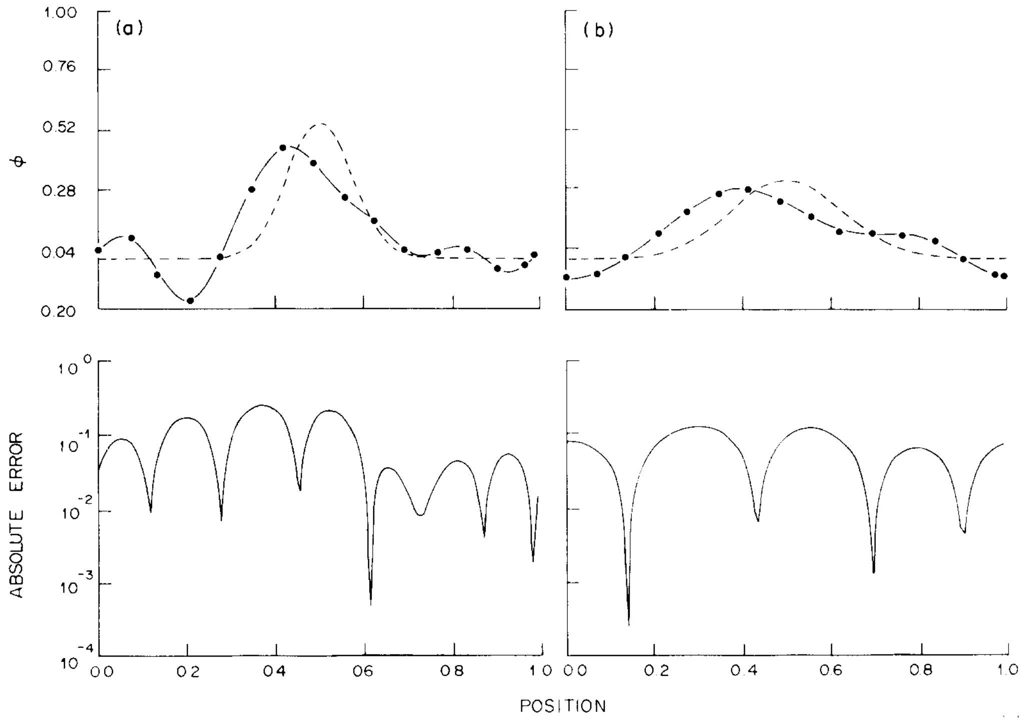

Fig. 8. *Upper curves*: analytic (---) and numerical comparison (●—●) of the ASH FD method vs position with $\lambda = 9 \times 10^3$ for 101 nodes at (a) cycle 16 and (b) cycle 64. *Lower curves*: absolute errors vs position.

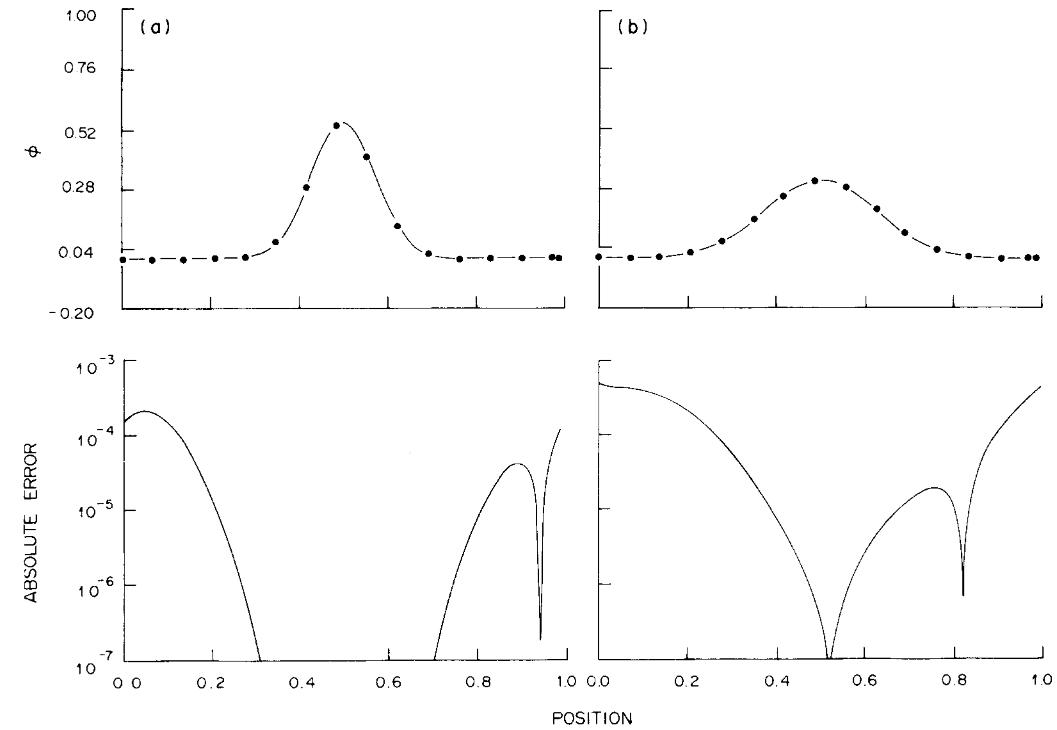

Fig. 9. *Upper curves*: analytic (---) and numerical comparison (●—●) of the SPECTRUM method vs position with $\lambda = 9 \times 10^3$ for 101 nodes at (a) cycle 16 and (b) cycle 64. *Lower curves*: absolute errors vs position.

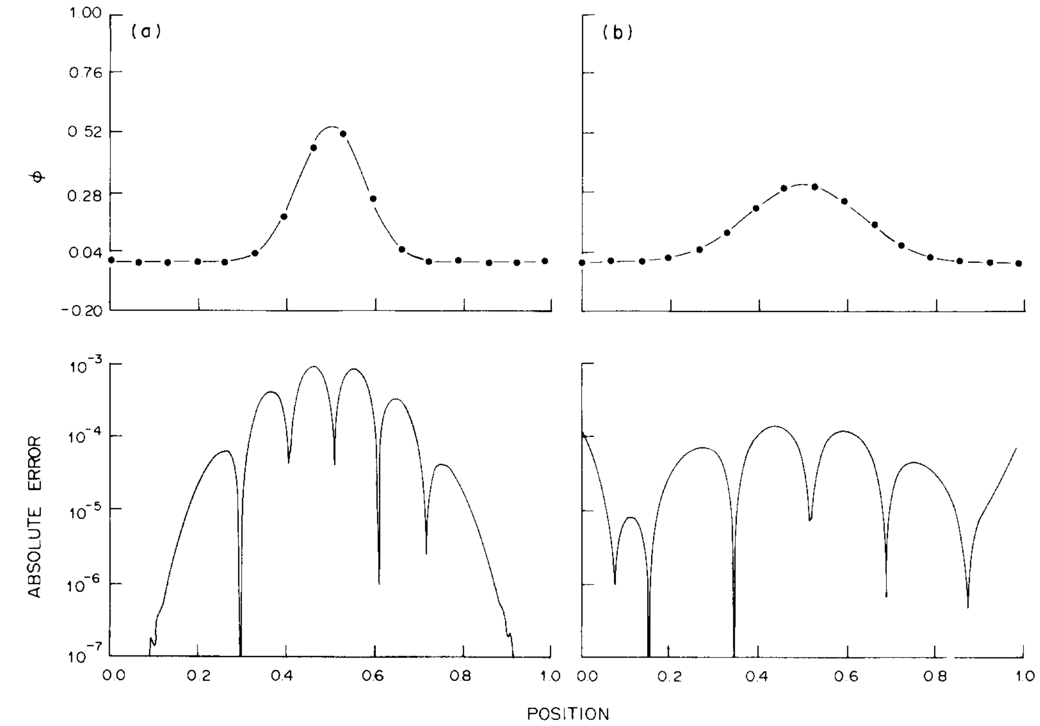

Fig. 10. *Upper curves*: analytic (---) and numerical comparison (●—●) of the CVP method vs position with $\lambda = 9 \times 10^3$ for 61 nodes at (a) cycle 16 and (b) cycle 64. *Lower curves*: absolute errors vs position.

Table 1. Finite domain advection–diffusion problems

Case	No. of nodes	λ	η	Q	T	Left boundary	Right boundary	Fig. no.
I	6	0.0	4.0	4.0	1.0	$\phi' = 0.0$	$\phi = 0.0$	11
II	6	0.1, 1.0	4.0	4.0	1.0	$\phi' = 0.0$	$\phi = 0.0$	12
	11	1.0, 10.0	4.0	4.0	0.5	$\phi' = 0.0$	$\phi = 0.0$	13
III	11	1.0, 10.0	0.0	0.0	1.0, 0.5	$\phi = 0.0$	$\phi = 1.0$	14
	31	10.0, 40.0	0.0	0.0	0.125	$\phi = 0.0$	$\phi = 1.0$	15
IV	31	1.0	4.0	4.0	1.5625E−2 0.5	$\phi = 1.0$	$\phi = 0.0$	16
V	31	10.0	0.0	0.0	3.125E−2 6.25E−2	$\phi = 1.0$	$\phi' = 0.0$	17

Several finite-medium problems were solved and compared with analytic solutions for a range of parameters and for a variety of boundary conditions, as summarized in Table 1. The CVP results are compared to the analytic solution in Figs 11–17.

Case I involves pure diffusion in a finite symmetric slab with zero outside boundary condition. The importance of retaining particle conservation in the solution method is emphasized in the comparison exhibited in Fig. 11, where approximately one order of magnitude in accuracy has been lost because solution conservation was not enforced. Increased CPU time is required in order to obtain particle conservation solutions using a Lagrangian constraint. However, the required CPU time is still less with the conservation constraint than required with the non-conserving solution in obtaining equivalent accuracy by running more nodes. This result is in agreement with previous steady-state and time-dependent results (Lee *et al.*, 1984; Lee and Wilson, 1981, 1984). Standard Galerkin methods without specific built-in (Lagrangian) conservation constraints will not necessarily conserve and thereby require considerably increased numbers of nodes for comparable accuracy.

Case II is similar to Case I with the addition of the non-zero advection term. For a fixed number of nodes, the CVP numerical accuracy compared to the analytical solution is clearly reduced as the advection

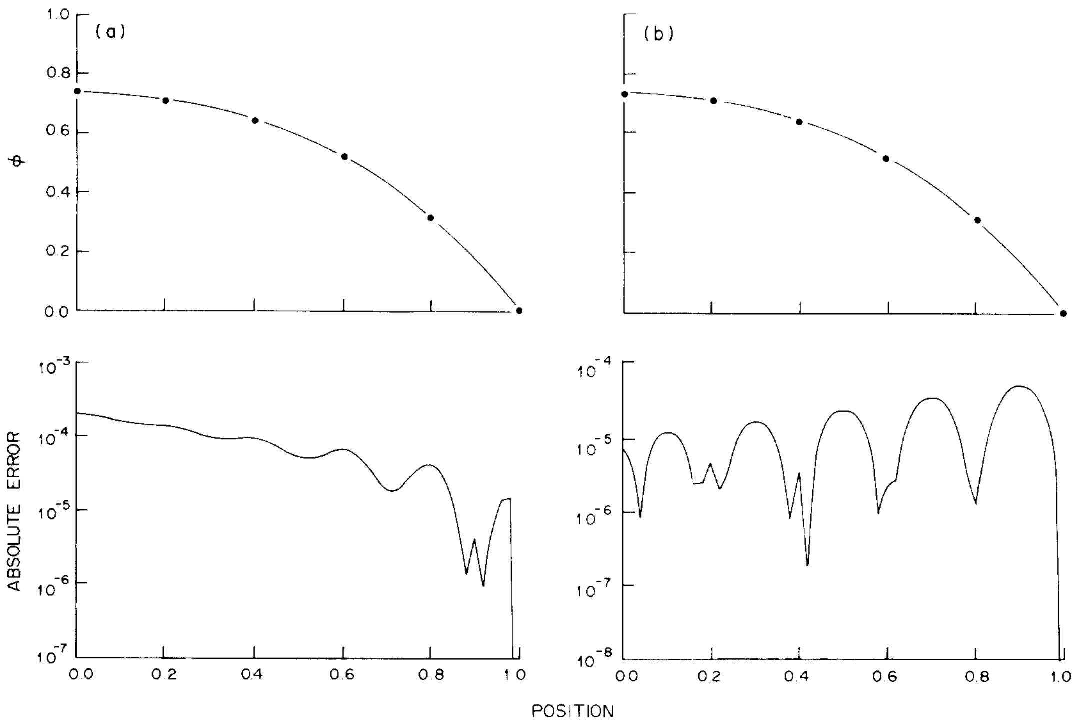

Fig. 11. *Upper curves*: comparison of (a) non-conserving and (b) conserving variational solutions for Case I with 6 nodes. *Lower curves*: absolute errors vs position.

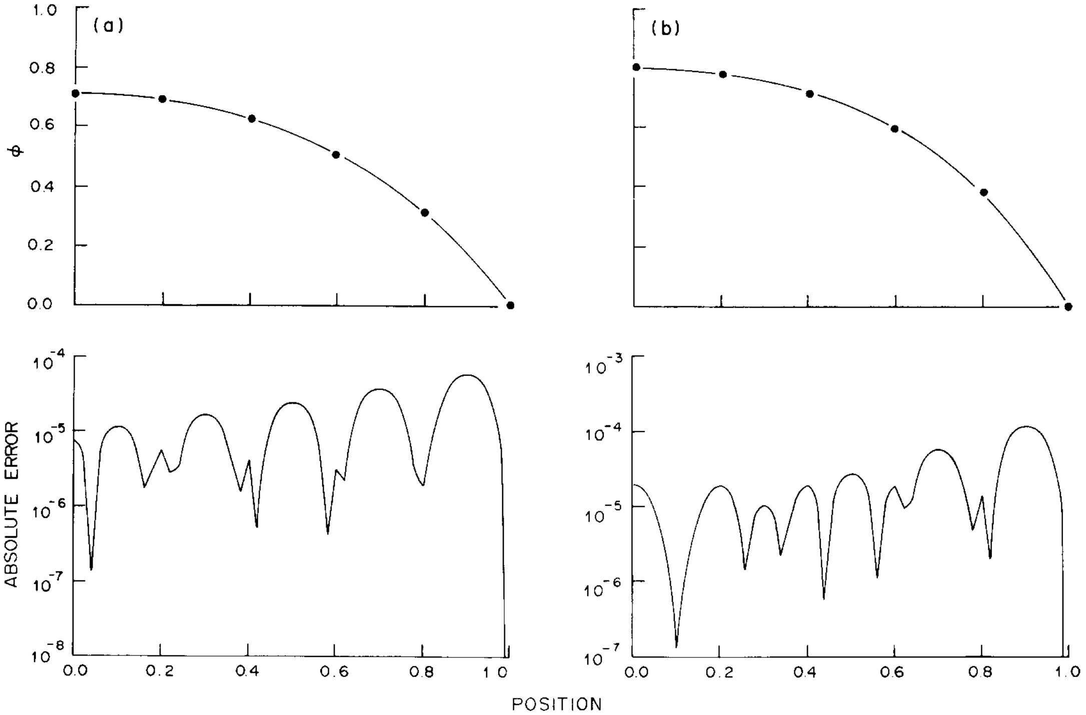

Fig. 12. *Upper curves*: comparison of the CVP method for Case II with 6 nodes for (a) $\lambda = 0.1$ at $T = 0.50$ and (b) $\lambda = 1.0$ at $T = 1.0$. *Lower curves*: absolute errors vs position.

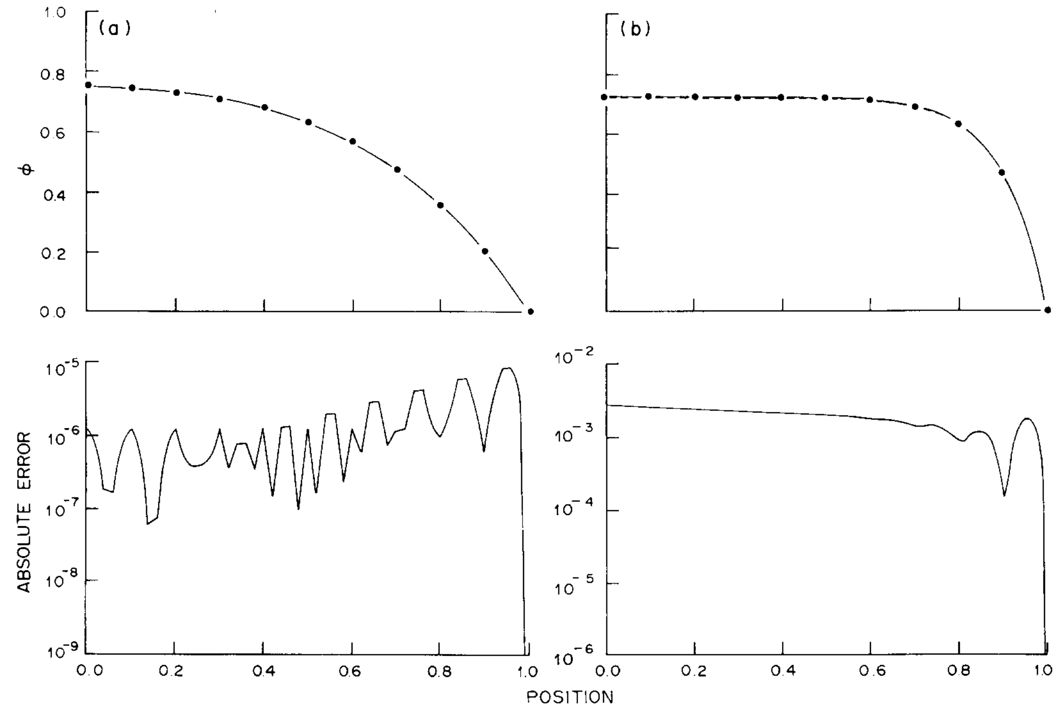

Fig. 13. *Upper curves*: comparison of the CVP method for Case II with 11 nodes for (a) $\lambda = 1.0$ at $T = 0.5$ and (b) $\lambda = 10$ at $T = 0.5$. *Lower curves*: absolute errors vs position.

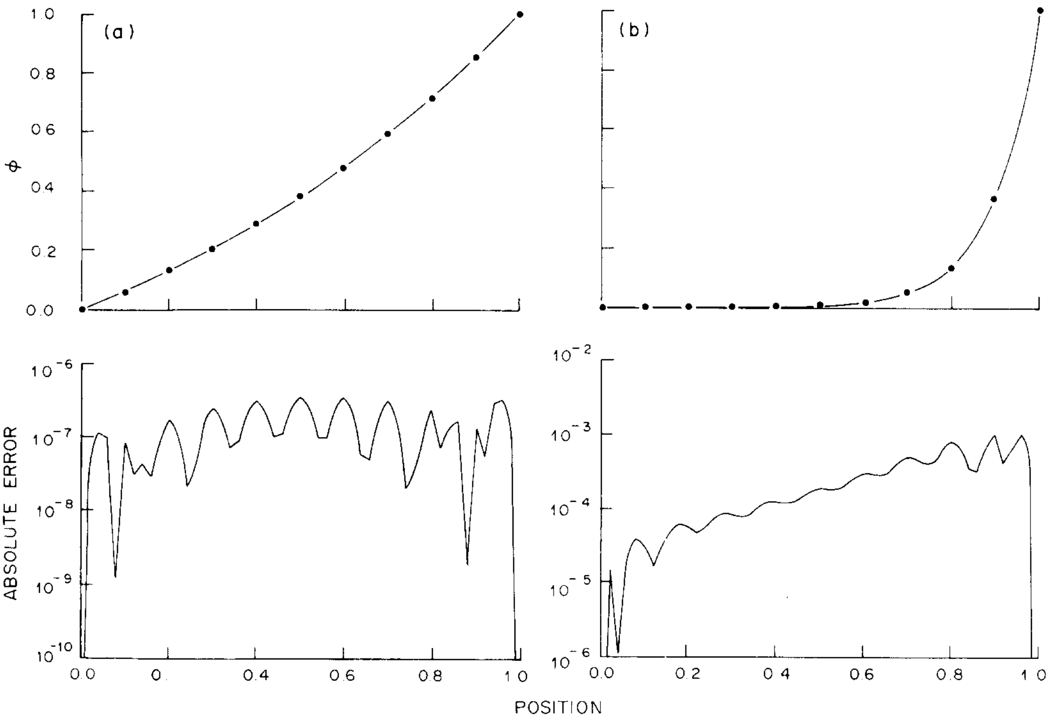

Fig. 14. *Upper curves*: comparison of the CVP method for Case III with 11 nodes for (a) $\lambda = 1.0$ at $T = 1.0$ and (b) $\lambda = 10$ at $T = 0.5$. *Lower curves*: absolute errors vs position.

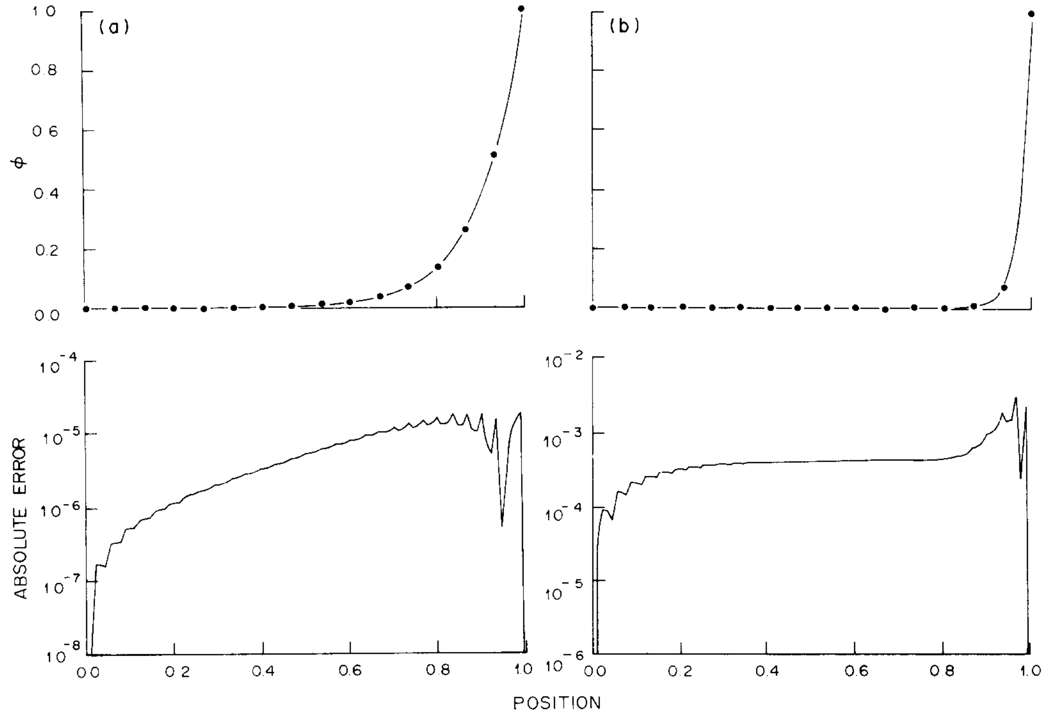

Fig. 15. *Upper curves*: comparison of the CVP method for Case III with 31 nodes for (a) $\lambda = 1.0$ at $T = 0.125$ and (b) $\lambda = 40$ at $T = 0.125$. *Lower curves*: absolute errors vs position.

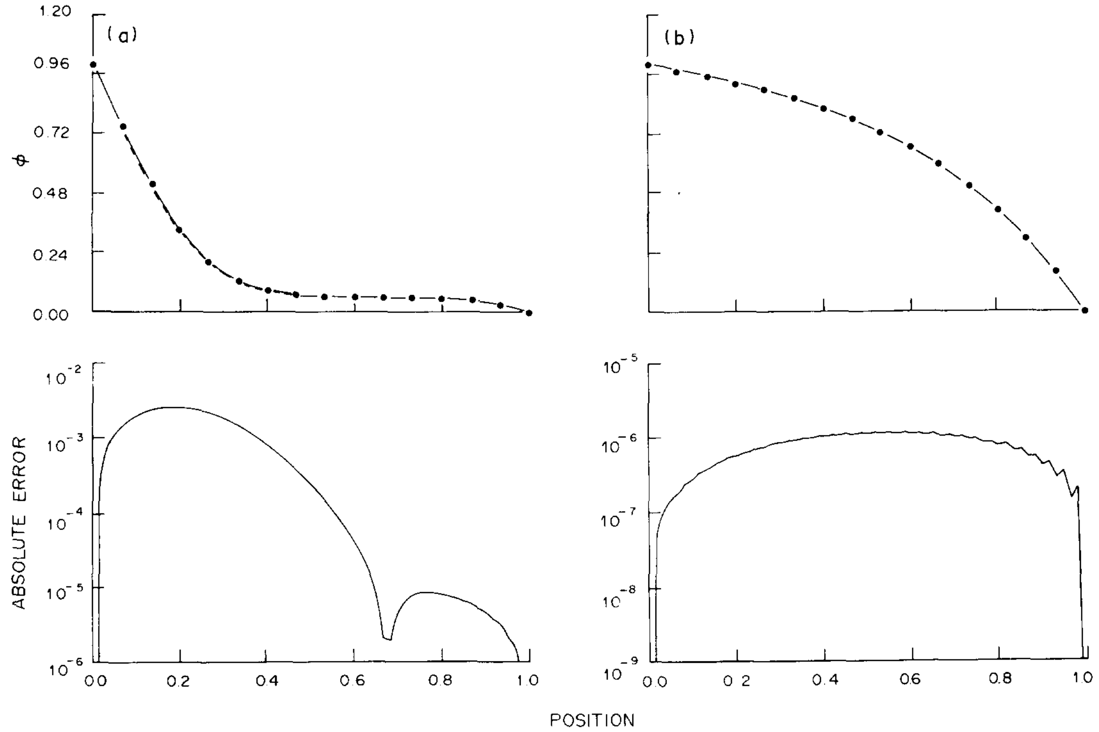

Fig. 16. *Upper curves*: comparison of the CVP method for Case IV with 31 nodes for $\lambda = 1.0$ at (a) $T = 1.5625 \times 10^{-2}$ and (b) $T = 0.5$. *Lower curves*: absolute errors vs position.

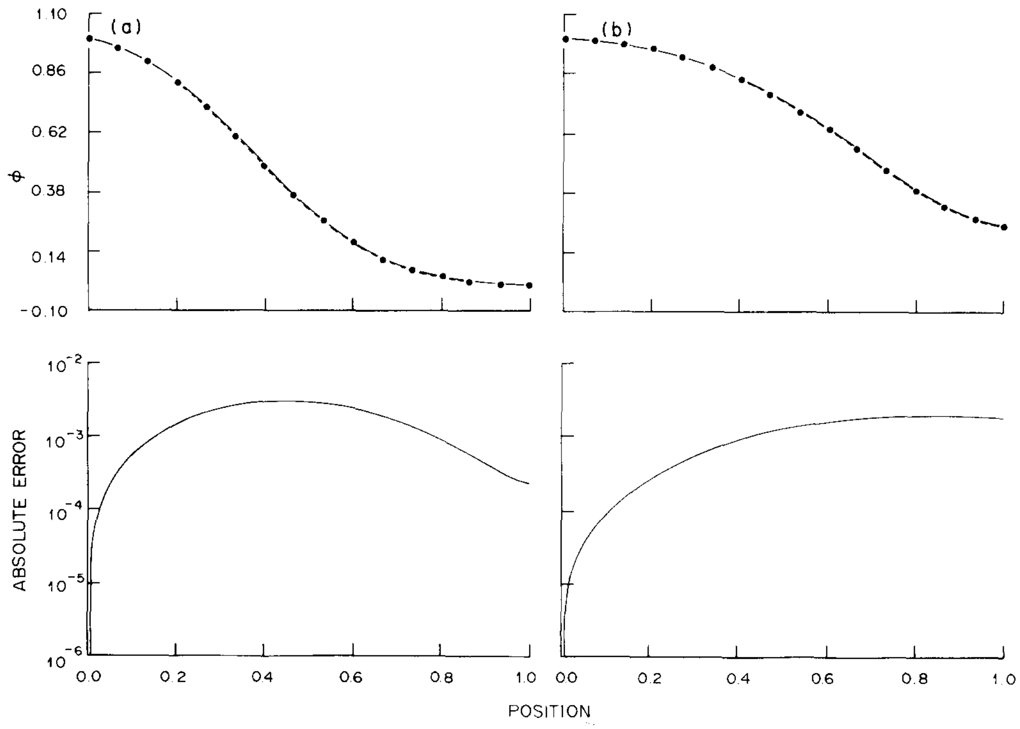

Fig. 17. *Upper curves*: comparison of the CVP method for Case V with 31 nodes for $\lambda = 10$ at (a) $T = 3.125 \times 10^{-2}$ and (b) $T = 6.25 \times 10^{-2}$. *Lower curves*: absolute errors vs position.

velocity is increased, as illustrated in Fig. 12 with 6 nodes and Fig. 13 with 11 nodes at early and late times. However, for a fixed value of λ, the desired numerical accuracy is readily obtained by increasing the number of nodes. This is demonstrated for $\lambda = 1.0$ in Fig. 12b and Fig. 13a.

When the boundary conditions are changed to those of Case III, the same general behavior pattern is exhibited in Fig. 14 for 11 nodes and Fig. 15 for 31 nodes. The errors increase with increasing advection for a fixed number of nodes, but increasing the number of nodes for a fixed advection value decreases the error. Additionally, it should be noted in Fig. 15 that increased spatial resolution is required in order to resolve sharp fronts. This feature is also evident in the solutions of Cases IV and V below.

Numerical solutions to Case IV are shown in Fig. 16 at two different times illustrating the facility of the CVP method to accurately track transient motion. The second time frame is essentially at steady state where the accuracy has actually improved over early times. This error reduction at late times is characteristic of the CVP method when steady-state solutions are approached. It is due to two basic features incorporated in the method: cubic trial functions and conservation. As steady state is approached $\partial\phi/\partial t$ approaches zero as it must physically. When the conservation constraint condition is removed, $\partial\phi/\partial t$ does not approach zero and large numerical errors result in disagreement with the analytical solution.

Finally, the solution to Case V, a problem previously addressed by Price *et al.* (1968) and Guymon (1970) is shown in Fig. 17 for 31 nodes. The agreement with the analytical solution is of particular interest here since both the boundary conditions are exactly satisfied by CVP. This is in contrast to approximate analytical solutions often quoted in the literature which fail for moderate h values near the right-hand boundary.

In summary, the five finite-domain test cases discussed above indicate that the CVP method generally has solutions with errors bounded by about 10^{-3} for a moderate number of nodes on the domain [0, 1]. The numerical errors of the method in comparison with analytical solutions can be further reduced by increasing the number of nodes. Increasing the number of nodes will, of course, increase the computational investment. For these problems quite acceptable accuracy is obtained by the CVP method using only 31 nodes.

CONCLUSIONS

Low-order FD methods are generally inadequate for modeling late-time advection processes. Even using the ASH technique to solve the spatial Lax–Wendroff FD equations exactly in time, the method is inadequate for obtaining late-time solutions. The necessary space–time coupling and high-order accuracy imposed by the SPECTRUM method results in accurate solutions, even at late times, for both the pure advection and advection–diffusion problems with periodic boundary conditions. Comparisons with the BETA FD method show that the SPECTRUM method yields considerably smaller numerical errors for a fraction of the CPU investment. Late-time solutions were less costly computationally to obtain than standard discretized solutions as a result of the ASH time-solution technique. However, the SPECTRUM method seems limited to solving constant coefficient problems with periodic or homogeneous Dirchlet boundary conditions. This limitation, unless removed, excludes consideration of many physical problems which are often of primary interest.

The CVP method was demonstrated to be applicable to a wide range of physical problems on the finite domain [0, 1]. This includes implementation of a general mixed boundary condition into numerical method the matrix representation. Accurate numerical solutions, compared to exact analytical solutions, were obtained for representative problems. The importance of obtaining particle conservation within each numerical cell and in the overall system was illustrated. When particle conservation was turned off for any solution, generally one order of magnitude in accuracy was lost.

The CVP solution method applied to infinite-medium problems, modeled by periodic boundary conditions, yielded excellent agreement with analytical solutions. However, the SPECTRUM method gave better results for the same infinite-domain problem with a Gaussian initial condition. The SPECTRUM method is therefore preferred for solving that particular problem, assuming that the model equation can satisfy the limitations of constant coefficients. The CVP method, on the other hand, is not limited by the restriction of constant model coefficients. The CVP method represents an accurate viable solution technique for a rather wide range of 1-D time-dependent advection–diffusion problems. Extensions appear possible to areas involving alternate geometries, multicomponents and non-linearities.

REFERENCES

Apperson C. E. Jr, Lee C. E. and Carruthers L. M. (1979) Report LA-7793-MS.

Chan R. K.-C. (1978) *Int. J. numer. Meth. Engng* **12,** 1131.

Chung T. J. (1978) *Finite Element Analysis in Fluid Dynamics.* McGraw-Hill, New York.

Davies H. C. (1980) *J. comput. Phys.* **37,** 280.

Dodes I. A. (1978) *Numerical Analysis for Computer Science*, p. 386. North-Holland, New York.

Fromm J. E. (1968) *J. comput. Phys.* **3,** 176.

Gantmacher F. R. (1960) *The Theory of Matrices*. Chelsa, New York.

Gardner A. O., Peaceman D. W. and Pozzi A. L. Jr (1964) *Soc. Petrol. Engrs J.* **4,** 26.

Guymon G. L. (1970) *Wat. Resour. Res.* **6,** 204.

Guymon G. L. (1972) *Wat. Resour. Res.* **8,** 1357.

Hennart J. P. (1973) *Nucl. Sci. Engng* **50,** 185.

Hennart J. P. (1979) *ANS Top. Meet. on Computational Methods in Nuclear Engineering*, Vol. 1, Sect. 3, pp. 119–133.

Hornbeck R. W. (1975) *Numerical Methods*. Quantum, New York.

Horton C. E. (1980) Master of Engineering Report, Texas A&M Univ., College Station, Tex.

Kermadis G. A. (1980) Naval Research Lab. Memorandum Report 4225.

Lanczos C. (1949) *The Variational Principles of Mechanics*. Unit. of Toronto Press, Toronto, Canada.

Lax P. D. and Wendroff B. (1960) *Communs pure appl. Math.* **15,** 363.

Lee C. E. (1980) *Int. Conf. on Nuclear Waste Transmutation, Austin, Tex.*, pp. 651–672.

Lee C. E. and Wilson B. C. (1981) *Trans. Am. nucl. Soc.* **39,** 961.

Lee C. E. and Wilson B. C. (1984) *J. comput. Phys.* Submitted for publication.

Lee C. E., Fan W. C. P. and Bratton R. L. (1984) *Ann. nucl. Energy* **11,** 493.

Morse P. M. and Feshbach H. (1953) *Methods in Theoretical Physics*. McGraw-Hill, New York.

Nakamura S. (1977) *Computational Methods in Engineering and Science with Applications to Fluid Dynamics and Nuclear Systems*. Wiley, New York.

Price H. S., Cavendish and Varga (1968) *Soc. Petrol. Engrs J.* **243,** 293.

Reid R. O. (1980) *Notes for Short Course on Computational Hydraulics*, 26–30 May 1980. Texas A&M Univ., College Station, Tex.

Richtmeyer R. D. (1957) *Difference Methods for Initial Value Problems*. Interscience, New York.

Roach P. J. (1976) *Computational Fluid Dynamics*, pp. 53–105. Hermosa, Albuquerque, N.Mex.

Shapiro R. (1975) Report AFCRL-TR-0212, pp. 15–24.

Shuhmiller J. H. and Ferguson R. E. (1979) EPRI Report NP-1236.

Turkell E. (1980) *Computational Fluid Dynamics*, Vol. 2 (Edited by Kollman W.), pp. 127–262.

APPENDIX

The matrix equation, equation (3), can be solved by an exponential matrix method. Assuming that $\mathbb{A}$ is constant over the time interval of interest, $[0, t]$, the Volterra method of the multiplicative integral gives the solution as (Lee, 1980; Gantmacher, 1960)

$$\mathbf{X}(t) = e^{\mathbb{A}t}\mathbf{X}(0) + \mathbb{A}^{-1}(e^{\mathbb{A}t} - I)\mathbf{S}_0,$$

where the source $\mathbf{S}_0$ is assumed constant, and $\mathbf{X}(0)$ is the initial condition vector. Defining the matrix $\mathbb{C} = \mathbb{A}t$ and the matrix operator $D(\mathbb{C})$ by

$$D(\mathbb{C}) = \mathbb{C}^{-1}(e^{\mathbb{C}} - I),$$

the solution can be written in the form

$$\mathbf{X}(t) = [I - \mathbb{C}D(\mathbb{C})]\mathbf{X}(0) + tD(\mathbb{C})\mathbf{S}_0.$$

The matrix operators $e^{\mathbb{C}}$ and $D(\mathbb{C})$ have the series representations

$$e^{\mathbb{C}} = \sum_{n=0}^{\infty} \mathbb{C}^n/n! = I + \mathbb{C}\sum_{n=0}^{\infty} \mathbb{C}^n/(n+1)!$$

and

$$D(\mathbb{C}) = \mathbb{C}^{-1}(e^{\mathbb{C}} - I) = \sum_{n=0}^{\infty} \mathbb{C}^n/(n+1)!.$$

The operator $D(\mathbb{C})$ exists even if the matrix $\mathbb{C}$ is singular. Evaluation of the series representations of both $e^{\mathbb{C}}$ and $D(\mathbb{C})$ is difficult if the eigenvalues of $\mathbb{C}$ exceed unity. A scaled matrix, $\mathbb{H} = 2^{-p}\,\mathbb{C}$, with norm less than 0.5, evaluates in the series expressions without difficulty. The appropriate value of p is determined from

$$p = 1 + [\ln(t) + (1/2)\ln(\Sigma_{ij}C_{ij}^2)]/\ln 2.$$

Evaluation of $e^{\mathbb{C}}$ and $D(\mathbb{H})$ is performed with a finite number, M, of series terms until

$$|\mathbb{H}|^{M+1}/(M+2)! = 1/[2^{M+1}(M+2)!] < \varepsilon,$$

for some small prescribed value ε. The evaluation of $D(\mathbb{C})$ is obtained from the recursion relationship

$$D(2^n\mathbb{H}) = D(2^{n-1}\mathbb{H})[I - (1/2)(2^{n-1}\mathbb{H})D(2^{n-1}\mathbb{H})],$$

for $0 \leqslant n \leqslant p$, which can be proved by induction (Lee, 1980).