JOURNAL OF THERMOPHYSICS AND HEAT TRANSFER
Vol. 7, No. 3, July–Sept. 1993

Explicit vs Implicit Schemes for the Spectral Method for the Heat Equation

G. E. Schneider* and K. Wittich†
University of Waterloo, Waterloo, Ontario, Canada

Three different boundary condition enforcement/time-stepping formulations are detailed for the Chebyshev collocation spectral method. Application is made to the one-dimensional heat equation with a saw-tooth initial condition and several boundary condition combinations. It is observed that for a small time step, the explicit time stepping/strong boundary condition implementation method and the implicit time-stepping method give essentially the same solutions for various numbers of collocation points. The explicit time-stepping/weak boundary condition enforcement scheme has significantly larger error at low numbers of collocation points, and requires considerably more collocation points to achieve convergence. While implicit schemes can employ significantly larger time steps than can the explicit schemes, implicit schemes require some form of matrix inversion, and this can be costly. Conversely, the explicit schemes can exploit the use of fast transform techniques to economize each time-step solution, but stability considerations limit the size of the time step that can be employed. A series of parametric studies are conducted to assess the relative economies of explicit vs implicit time-stepping schemes. The one-dimensional heat equation is used as the basis problem and Robin boundary conditions are considered. The maximum explicit time-step limit is determined and a correlation provided that includes the dependence on boundary condition parameter values.

Nomenclature

A = matrix or correlation coefficient
a = boundary condition constant
a_k = Chebyshev expansion function coefficients
$a_k^{(n)}$ = Chebyshev coefficients for the nth derivative
B = matrix
Bi = Biot modulus, hL/k
b = boundary condition constant
C = matrix or correlation coefficient
c = coefficient
D = matrix
D_{ij} = Chebyshev collocation differentiation matrix
k = thermal conductivity or sequence index
L = half-width of the domain
N = number of collocation points minus one
N_t = number of time steps
T_k = Chebyshev polynomials
t = time
$u_N(x)$ = approximate series expansion of a function $u(x)$
x = spatial coordinate
x_j = collocation points
α = thermal diffusivity
β = boundary condition coefficient
γ = boundary condition coefficient
Δ = difference
δ = boundary condition coefficient
ζ = boundary condition coefficient
θ = temperature

Superscripts
new = new time step
old = old time step

Presented as Paper 91-0158 at the AIAA 29th Aerospace Sciences Meeting and Exhibit, Reno, NV, Jan. 7–10, 1991; received Aug. 12, 1991; revision received Aug. 3, 1992; accepted for publication Aug. 3, 1992.

*Professor, Department of Mechanical Engineering. Associate Fellow AIAA.
†Graduate Student, Department of Mechanical Engineering.

Introduction

THE Chebyshev spectral collocation method has been widely applied in fluid dynamics due to its ease of use, accuracy, and rapid convergence characteristics. The authors are interested in using the above method as an efficient numerical solution method to a binary solid/liquid phase change problem with convection. In this numerical solution, the effective implementation of Neumann and Robin boundary conditions will be extremely important. These phase-change problems have boundary conditions of the general Robin type, as well as interfacial compatibility constraints.[1] The nonlinearity of enthalpy and concentration at the a priori unknown solid/liquid interface can require many solution iterations at each time step. When complete temporal evolution of the thermal field is required, it is important that the solution procedure be economical in its implementation. A fundamental decision that must be made is whether to use implicit time-stepping techniques or to use explicit techniques requiring more, but less costly, time steps due to stability considerations. This article addresses this basic question and evaluates the relative computational advantages of the two different time-stepping schemes. For the explicit methods, fast transform techniques will be employed, whereas, for the implicit schemes, direct matrix inversion is employed.

The model problem chosen for this study is the one-dimensional transient heat conduction problem, which is described by the equation

$$\frac{\partial \theta}{\partial t} = \alpha \frac{\partial^2 \theta}{\partial x^2} \tag{1}$$

Saw-tooth, top hat, and cosine initial distributions are considered.

Chebyshev Collocation Spectral Method

Chebyshev expansion functions in spectral approximations have the advantage of being able to handle nonperiodic boundary conditions and to use the computationally efficient fast Fourier transforms (FFTs).[2] Chebyshev polynomials are defined on the interval $[-1, 1]$ by

$$T_k(x) = \cos(k \cos^{-1} x) \quad k = 0, 1, \ldots, N \tag{2}$$

The collocation points are chosen for convenience to be

$$x_j = \cos(\pi j/N) \quad j = 0, 1, \ldots, N \tag{3}$$

where N is the number of collocation points in the solution domain.[2] With this choice of collocation points, the Chebyshev expansion functions reduce to

$$T_k(x_j) = \cos(\pi jk/N) \quad k = 0, 1, \ldots, N \tag{4}$$

The approximate series representation $u_N(x)$ of a function $u(x)$ takes the form

$$u_N(x) = \sum_{k=0}^{N} a_k T_k(x) \tag{5}$$

where a_k are the expansion function coefficients. These coefficients[2] are defined by

$$a_k = \frac{2}{c_k N} \sum_{j=0}^{N} c_j^{-1} u_j \cos \frac{\pi jk}{N} \quad k = 0, 1, \ldots, N \tag{6}$$

where

$$u_j = u(x_j)$$

$$c_j = \begin{cases} 2; & j = 0 \text{ or } N \\ 1; & j = 1, \ldots, N-1 \end{cases}$$

The nth derivative of the series approximation given by Eq. (5) is

$$\frac{\partial^{(n)} u_N}{\partial x^{(n)}} = \sum_{k=0}^{N} a_k^{(n)} T_k(x) \tag{7}$$

The coefficients of the first derivative are found by the recursive formula

$$\begin{aligned} a_{N+1}^{(1)}(t) &= 0 \\ a_N^{(1)}(t) &= 0 \\ c_k a_k^{(1)}(t) &= a_{k+2}^{(1)}(t) + 2(k+1)a_{k+1}(t) \\ k &= N-1, \quad N-2, \ldots, 0 \end{aligned} \tag{8}$$

where the superscript (1) denotes first derivative. Similarly, the coefficients for the second derivative are generated by the formula

$$\begin{aligned} a_{N+1}^{(2)}(t) &= 0 \\ a_N^{(2)}(t) &= 0 \\ c_k a_k^{(2)}(t) &= a_{k+2}^{(2)}(t) + 2(k+1)a_{k+1}^{(1)}(t) \\ k &= N-1, \quad N-2, \ldots, 0 \end{aligned} \tag{9}$$

The first and second derivatives of the expansion [Eq. (5)] are found by substituting Eqs. (8) and (9) into Eq. (7), respectively.[2]

Boundary Condition Implementation

Using the Chebyshev spectral collocation method to approximate the derivative on the right side of Eq. (1), either explicit or implicit time-stepping schemes may be used in combination with strong or weak boundary condition implementation. Strong implementations require that the approximating polynomial satisfy the boundary conditions exactly. Weak implementations satisfy the boundary conditions in a less direct way.

In a time explicit scheme, the derivative on the right side of Eq. (1) is

$$\left.\frac{\theta_N^{\text{new}} - \theta_N^{\text{old}}}{\Delta t}\right|_{x=x_j} = \alpha \sum_{k=0}^{N} a_k^{(2)}(t) \cos\left(\frac{\pi jk}{N}\right) \tag{10}$$

where Δt is the time step. Rearranging this, the values of θ_N at the new time step are given by

$$\theta_N^{\text{new}} = \alpha \Delta t \sum_{k=0}^{N} a_k^{(2)}(t) \cos\left(\frac{\pi jk}{N}\right) + \theta_N^{\text{old}} \tag{11}$$

Here, the second derivative expansion coefficients $a_k^{(2)}$ are found using the solution at the old time step.

Time explicit solution with strong imposition of Dirichlet boundary conditions is the simplest procedure relative to other combinations of boundary condition, implementation, and time-stepping scheme. The first step here proceeds by generating the coefficients for the second derivative. This is effected by substituting the initial conditions into Eq. (6) and generating the $a_k^{(2)}$ coefficients by Eqs. (8) and (9). With the $a_k^{(2)}$ coefficients known, the right side of Eq. (10) is evaluated. The values of θ_N at the new time step, θ_N^{new}, are then given by Eq. (11). The Dirichlet boundary conditions are now enforced at the endpoints, x_0 and x_N, by directly overwriting the calculated values to satisfy the boundary constraints in favor of the governing equation at the boundary. The values of θ_N after the second time step are now found by calculating the coefficients for the second derivative $a_k^{(2)}$ by way of Eq. (9), and by repeating the procedure outlined above.

General Robin boundary conditions of the form

$$\begin{aligned} \beta \frac{\partial \theta}{\partial x} + \zeta\theta &= a \quad \text{at} \quad x = 1 \\ \delta \frac{\partial \theta}{\partial x} + \gamma\theta &= b \quad \text{at} \quad x = -1 \end{aligned} \tag{12}$$

also may be applied strongly. As with the Dirichlet case, the solution at the interior nodes is advanced by satisfying the explicit spectral equation. The solution values at the end points x_0 and x_N are found by satisfying the following 2×2 system of equations:

$$\begin{aligned} (\zeta + \beta D_{00})\theta_N(x_0) + \beta D_{0N}\theta_N(x_N) &= -\beta \sum_{j=1}^{N-1} D_{0j}\theta_N(x_j) + a \\ (\gamma + \delta D_{NN})\theta_N(x_N) + \delta D_{N0}\theta_N(x_0) &= -\delta \sum_{j=1}^{N-1} D_{nj}\theta_N(x_j) + b \end{aligned} \tag{13}$$

The spectral differentiation matrix D_{ij} is easily found.[3] Here, the boundary conditions are exactly satisfied at the new time step.

A time explicit scheme with weak enforcement of general Robin boundary conditions may also be implemented. This has been examined by Schneider and Wittich,[4] who concluded that the weak enforcement method suffers from significantly degraded accuracy while offering no computational advantage of consequence.

Implicit time stepping may be desirable due to the explicit time-step restriction. For the time implicit solution of Eq. (1), the backward difference formula is chosen for simplicity to approximate the temporal derivative

$$\left.\frac{\theta_N^{\text{new}} - \theta_N^{\text{old}}}{\Delta t}\right|_{x=x_j} = \alpha \sum_{k=0}^{N} a^{(2)}(t) \cos\left(\frac{\pi jk}{N}\right) \tag{14}$$

This equation, it is recalled, is the same as that used for explicit time stepping. However, when solved implicitly, the right side of Eq. (14) is based on the solution at the new time step. By expanding θ_N^{new} using Eq. (5), we may write the time implicit form of Eq. (14) in matrix notation as

$$[T_k(x_j)]\{a_k^{(2)}\} - (1/\alpha\Delta t)[T_k(x_j)]\{a_k\} = (-1/\alpha\Delta t)\{\theta_N^{old}\} \quad (15)$$

where

$$[T_k(x_j)] = \begin{bmatrix} T_0(x_0) & T_1(x_0) & \cdots & T_N(x_0) \\ T_0(x_1) & T_1(x_1) & & \\ \vdots & & \ddots & \\ T_0(x_N) & T_1(x_N) & & T_N(x_N) \end{bmatrix} \quad (16)$$

The general Robin boundary conditions, Eq. (12), can be incorporated by collocating the interior nodes as indicated by Eq. (15), but requiring the endpoints to satisfy the boundary conditions exactly. The boundary conditions are included in Eq. (15) by overwriting the top and bottom rows of the matrices and the vector $\{\theta_N^{old}\}$. More precisely, the discrete equations may be written after boundary condition application in the form

$$[B]\{a_k^{(2)}\} + [C]\{a_k^{(1)}\} + [D]\{a_k\} = (-1/\alpha\Delta t)\{\theta_N^{old}\}_{mod} \quad (17)$$

where

$$[B] = \begin{bmatrix} \cdots & 0 & \cdots \\ & \cdots & \\ & T_k(x_j) & \\ & \cdots & \\ \cdots & 0 & \cdots \end{bmatrix}$$

$$[C] = \begin{bmatrix} \beta T_0(x_0) & \beta T_1(x_0)\cdots & & \beta T_N(x_0) \\ & & \cdots & \\ & & 0 & \\ & & \cdots & \\ \delta T_0(x_N) & \delta T_1(x_N)\cdots & & \delta T_N(x_N) \end{bmatrix}$$

$$[D] = \begin{bmatrix} \zeta T_0(x_0) & \zeta T_1(x_0)\cdots & & \zeta T_N(x_0) \\ & & \cdots & \\ & & (-1/\alpha\Delta t)T_k(x_j) & \\ & & \cdots & \\ \gamma T_0(x_N) & \gamma T_1(x_N)\cdots & & \gamma T_N(x_N) \end{bmatrix}$$

$$\{\theta_N^{old}\}_{mod} = \begin{Bmatrix} -\alpha\Delta t a \\ \vdots \\ \theta_N^{old} \\ \vdots \\ -\alpha\Delta t b \end{Bmatrix}$$

The solution at the new time step is completely determined if $\{a_k\}$ is found. To solve for $\{a_k\}$ in Eq. (17), we write both $\{a_k^{(1)}\}$ and $\{a_k^{(2)}\}$ in terms of a known matrix times $\{a_k\}$:

$$\{a_k^{(2)}\} = [B^*]\{a_k\}$$
$$\{a_k^{(1)}\} = [C^*]\{a_k\} \quad (18)$$

The matrices $[B^*]$ and $[C^*]$ can be found by using the recursive relationships given by Eqs. (8) and (9). It can be shown that

$$[B^*] = [I - A^{11}]^{-1}[A^{01}][I - A^{11}]^{-1}[A^{01}]$$
$$[C^*] = [I - A^{11}]^{-1}[A^{01}] \quad (19)$$

where

$$[A^{11}] = \begin{bmatrix} 0 & 0 & \frac{1}{c_0} & & & & \\ & 0 & 0 & \frac{1}{c_1} & & 0 & \\ & & 0 & 0 & \frac{1}{c_2} & & \\ & & & 0 & 0 & \ddots & \\ & & & & 0 & \ddots & \frac{1}{c_{N-2}} \\ & 0 & & & & \ddots & 0 \\ & & & & & & 0 \end{bmatrix}$$

$$[A^{01}] = \begin{bmatrix} 0 & \frac{2}{c_0} & & & & & \\ & 0 & \frac{4}{c_1} & & & 0 & \\ & & 0 & \frac{6}{c_2} & & & \\ & & & 0 & \ddots & & \\ & 0 & & & \ddots & \frac{2N}{c_{N-1}} & \\ & & & & & 0 & \end{bmatrix}$$

Substituting Eq. (18) into Eq. (17) and collecting terms gives

$$[[B][B^*] + [C][C^*] + [D]]\{a_k\} = (-1/\alpha\Delta t)\{\theta_N^{old}\}_{mod} \quad (20)$$

By matrix inversion we can find $\{a_k\}$

$$\{a_k\} = [[B][B^*] + [C][C^*] + [D]]^{-1}(-1/\alpha\Delta t)\{\theta_N^{old}\}_{mod} \quad (21)$$

Finally the desired solution at the new time step is found by using Eq. (5)

$$\theta_N^{new}(x_j) = \sum_{k=0}^{N} a_k T_k(x_j) \quad (22)$$

The above is easily modified to implement the Crank-Nicolson scheme for which the results are reported in this article.

Fortunately, the inversion in Eq. (21) needs only to be done once in the time evolution of the solution. In fact, the inverted matrix here only depends on the number of grid points N, if the collocation points are always defined by Eq. (3) and on the boundary condition parameters.

Convergence Studies

The model equation being considered, it is recalled, is the one-dimensional, transient heat equation with general Robin boundary conditions. The specific boundary condition cases of homogeneous, Dirichlet conditions on both surfaces and of adiabatic conditions on both surfaces will be considered for purposes of these convergence studies. In addition, three specific initial conditions will be considered: 1) a unit top hat profile; 2) a unit saw-tooth profile homogeneous at the boundaries and peaking at the centerline; and 3) a unit cosine distribution having one complete cycle across the width and being $+1$ at the centerline and -1 at both boundaries. These problem definitions represent typical initial conditions encountered in single and multidimensional applications, and permit

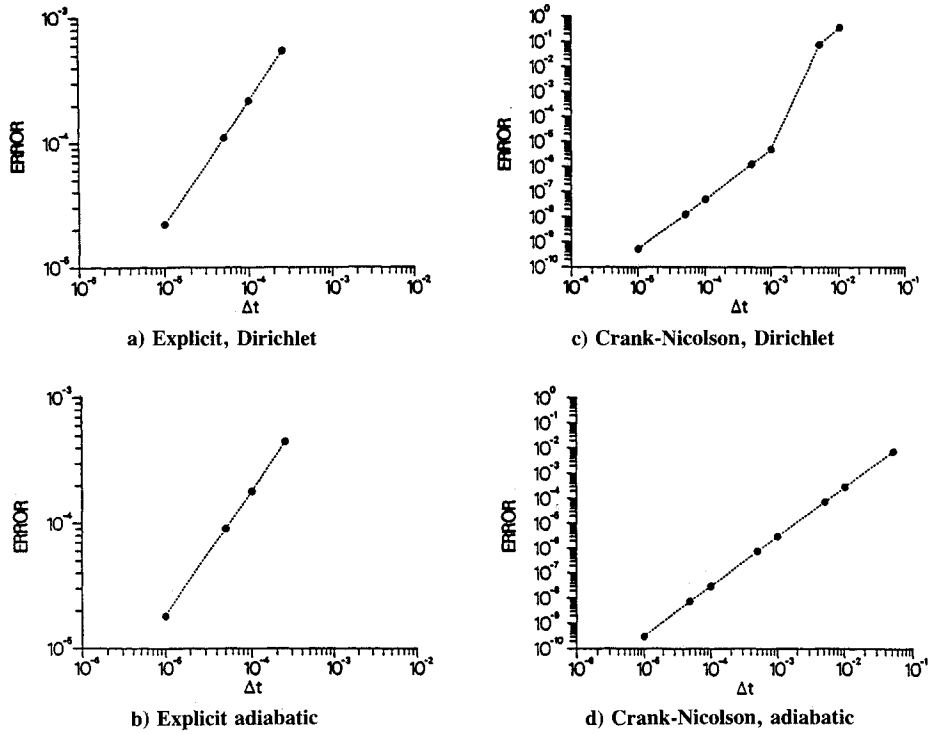


Fig. 1 Temporal convergence for $N = 20$.

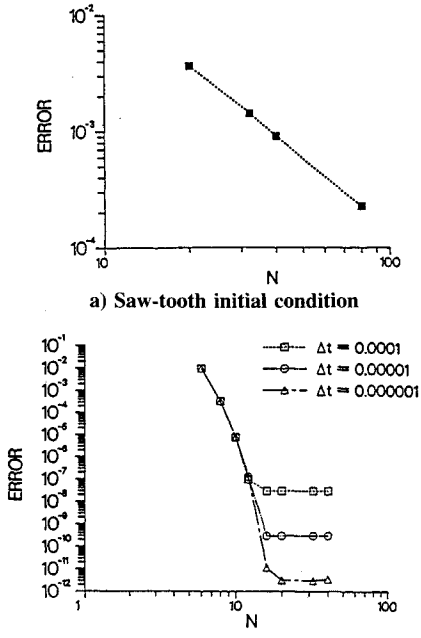


Fig. 2 Spatial convergence.

the examination of initial conditions that contain singularities, as well as those that are smooth throughout the entire spatial domain.

For 20 collocation points, the convergence results are shown in Figs. 1 and 2. The error presented here is the absolute error in the spectral approximate solution as determined by comparison with the exact solution. The exact solution was obtained using separation of variables procedures. Numerical values of the exact solution at the collocation points were obtained by summing the resultant series expansion; in certain cases, several thousand terms were required in order to obtain values accurate to machine round-off limits. Figures 1a and 1b correspond to the explicit procedure and indicate the expected first-order convergence with time step. To facilitate reasonable boundary condition scenarios, Fig. 1a corresponds to the top hat initial condition with homogeneous Dirichlet boundary conditions, while Fig. 1b corresponds to the cosine distribution initial condition with adiabatic boundary conditions on both sides.

Figures 1c and 1d, also for 20 collocation points, present the temporal convergence characteristics for the Crank-Nicolson procedure. These figures indicate the expected second-order convergence with time-step size. Figure 1c corresponds to the homogeneous Dirichlet boundary conditions with unit top hat initial condition, and Fig. 1d corresponds to the cosine initial distribution with adiabatic boundary conditions. The jump in error for Fig. 1c is attributed to the coarse temporal mesh for the largest two time-step intervals. Other test conditions which confirm proper code operation and convergence include both homogeneous Dirichlet and adiabatic conditions with the saw-tooth initial condition.

For sufficiently small time-step sizes, the spatial discretization error will dominate over the temporal discretization error. The magnitude of this spatial error, as determined from the theory of discrete Chebyshev expansions, is dependent on the number of times the function being represented can be continuously differentiated. To examine this, and the resulting convergence, the saw-tooth initial condition and the cosine initial distribution are examined; in both cases both boundaries are adiabatic. The results are presented in Figs. 2a and 2b, respectively.

In Fig. 2a, using $\Delta t = 0.0001$, the spatial convergence is approximately second order. This is consistent with the expectation that spectral convergence will not be achieved due to the discontinuity in first derivative that results from the saw-tooth initial condition. Conversely, however, for the cosine initial distribution, for which results are provided in Fig. 2b, spectral convergence is expected. This is demonstrated by the results of that figure, wherein the convergence shows an ever increasing (in magnitude) negative slope until the temporal discretization error becomes dominant, at which point the error becomes independent of the number of collocation points. The three different time-step sizes, with their differing asymptotic errors demonstrate this very effectively.

Explicit Time-Step Determination

In order to perform cost comparisons between the explicit formulation and the Crank-Nicholson scheme, it is useful to provide a time-step reference frame. Such a reference is provided by the explicit time-step limit since the explicit scheme is unstable for time steps larger than this value, and since accuracy requirements alone do not usually require such small time steps.

For Chebyshev collocation spectral methods applied to the one-dimensional transient heat equation, it has been shown by Gottlieb and Orzag[5] that the maximum stable time step for the explicit formulation has the following dependence:

$$\Delta t_{\text{lim}} \alpha (1/N^4) \tag{23}$$

This restriction is more severe than the finite-difference counterpart, and is partly due to the increased closeness of the collocation points that occurs near the boundaries. Since the exact dependence of the stable time step on N is not known from analytic considerations, this was determined by performing numerical experiments.

While instability leads to computed values of the dependent variable that are outside the physically realistic bounds determined by initial and boundary conditions, numerical studies indicated that, near the stability limit, such simple observation is unreliable in determining the stability limit to acceptable accuracy. For the determination of this limiting time step, a flat, null, initial profile with a centerline perturbation of 10^{-4}, and with boundary conditions as appropriate, was used. Growth in the magnitude of the perturbation was detected and used to establish the maximum stable time step using the procedure described below.

Estimation of the time-step limit depends on how long in the time domain detection efforts for instability continue. With unstable formulations, as the ratio of the time-step size to the explicit limit becomes larger, instability is manifested earlier in the time domain. For universal application, however, it is desirable to know the explicit limit independent of an ending nondimensional time, t_{end}. To determine this, estimates of the time-step limit found using several nondimensional time extents were extrapolated to an infinite ($1/t_{\text{end}} = 0$) nondimensional ending time. For a specified t_{end} experiment, bisection was used to determine Δt_{lim} within an interval range of 10^{-7} for $N \leq 20$ and 10^{-8} for $N > 20$. The results of this study for homogeneous Dirichlet boundary conditions and for adiabatic boundary conditions are presented in Table 1. It is estimated that the values in Table 1 are accurate to 0.15%.

Inspection of Table 1 reveals that the explicit time-step limit depends on the boundary condition application. The dependence of the explicit time-step limit on the Biot modulus, with $N = 20$, is illustrated in Fig. 3; the data in the figure were obtained using the above-described procedure. The complete transition from adiabatic dependence to Dirichlet dependence essentially takes place in the range $1 \leq Bi \leq 10^3$. The curve presented in the figure represents a correlation of the data and is given by

$$\Delta t_{\text{lim}} = A_1 + A_2 \tanh \left\{ A_3 \log \left(\frac{A_4}{Bi} \right) \right\} \tag{24}$$

where the coefficients are given by

$$\begin{aligned} A_1 &= 0.55849 \times 10^{-3} \\ A_2 &= 0.29839 \times 10^{-3} \\ A_3 &= 1.25 \\ A_4 &= 50.0 \end{aligned} \tag{25}$$

Table 1 Explicit time-step limits

	$\Delta t_{\text{lim}} \times 10^3$	
N	Dirichlet	Adiabatic
16	0.6294	2.083
20	0.2601	0.8533
32	0.04050	0.1329
40	0.01643	0.05466

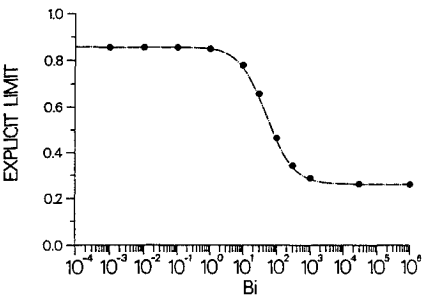


Fig. 3 Explicit time step limit dependence on boundary condition parameter values for $N = 20$.

For collocation point distributions different from $N = 20$ the equation

$$\Delta t_{\text{lim}} = (C/N^4) \tag{26}$$

was used where C is found by calibrating with one recorded value of the explicit limit data where $N = 20$. This procedure introduces error of interpolation/extrapolation of less than 2.5%.

Cost Assessment and Comparison

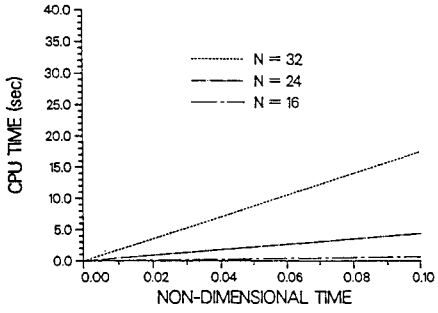
The explicit scheme has a low cost per time step, with the major cost being that of the two Fourier cosine transforms (FCT) required; however, the time step size is limited by stability considerations. Conversely, the Crank-Nicolson scheme is unconditionally stable, but requires a significant effort in matrix preparation and inversion, at the very first time step, as well as matrix multiply operations for each time step. In order to perform a cost comparison, then, these costs need to be determined. In this study, timing tests were performed, the resulting data correlated, and graphical presentation made of the time-step ratio (Crank-Nicolson/explicit limit) that results in equal costs for both schemes. It can then be determined, for prescribed applications, what time-step ratio is required to provide more economical solutions using the Crank-Nicolson scheme.

Explicit Scheme Costs

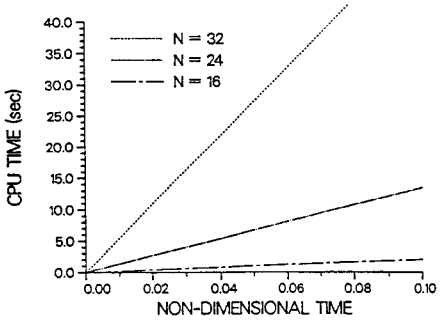
With the exception of a very small amount of initial overhead, the explicit scheme computational cost is directly proportional to the number of time steps. Since the maximum stable explicit time step is used in all experiments, the CPU time is given by

$$\text{CPU}^e = \text{A}_1^e t/\Delta t_{\text{lim}} = A_1^e N_t \tag{27}$$

where t is the time duration of the numerical simulation and N_t is the corresponding number of time steps. The calibrating coefficient A_1^e was found by numerical experiments on a VAX 11/785 computer and varies with boundary condition type and number of collocation points, N. This is shown in Fig. 4 for the Dirichlet and adiabatic boundary conditions. The slope of the curve increases as N increases for two reasons; the cost increases with N because the size of the FCT which need to be computed increases with N, and because the maximum stable time step-size decreases with N. Also, for a given N, the slope of the curve for the Dirichlet case is greater than



a) Adiabatic boundaries



b) Dirichlet boundaries

Fig. 4 Explicit scheme costs.

Table 2 Correlation coefficients for CPU times

Case	Coefficient	Value
Explicit		
Homogeneous	C_1^e	1.16262×10^{-4}
Dirichlet	C_0^e	4.87048×10^{-3}
Adiabatic	C_1^e	1.28679×10^{-4}
	C_0^e	4.43716×10^{-4}
Implicit		
Fixed	C_3^f	1.15547×10^{-4}
	C_2^f	5.82188×10^{-4}
	C_1^f	1.27375×10^{-2}
Variable	C_2^t	1.46875×10^{-5}
	C_1^t	9.85625×10^{-4}
	C_0^t	1.0×10^{-5}

that for the adiabatic case; this is due to the influence that boundary condition type has on the maximum stable time-step size.

The dependence of A_1^e on N is correlated in the form

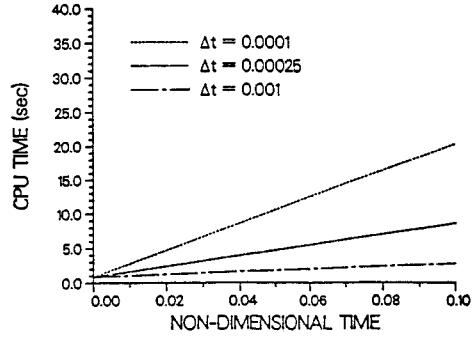
$$A_1^e = C_1^e N \log_2 N + C_0^e \tag{28}$$

to reflect the $N \log_2 N$ cost of the FCT computation and the fixed costs per time step. The coefficients C_1^e and C_2^e depend on the boundary condition type and are given in Table 2 for the Dirichlet and adiabatic cases. Using these coefficients in the above correlation leads to errors of less than 4% for $N = 16$ and less than 7.5% for $N = 32$, both with $N_t = 400$. Although a linear (with N) term is not included, its inclusion did not enhance the correlation; in fact, the correlation was degraded by such inclusion.

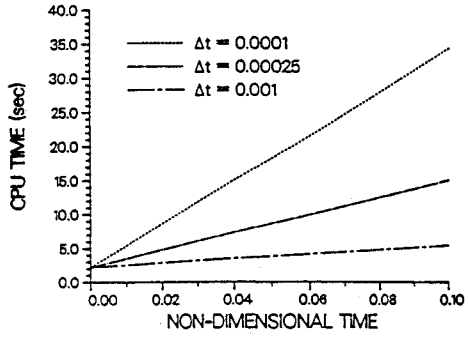
Crank-Nicolson Scheme Costs

The Crank-Nicolson scheme has a substantial overhead, initial cost as well as a time-step-dependent cost, and so the CPU time correlation takes the form

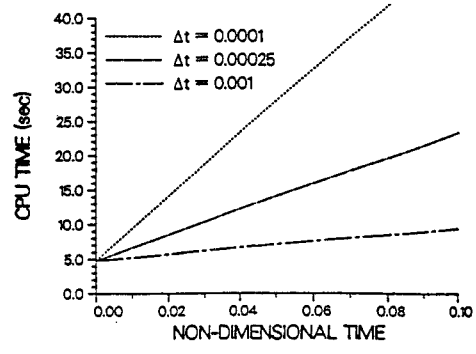
$$\text{CPU} = A_1^{CN}(t/\Delta t^{CN}) + A_0^{CN} \tag{29}$$



a) $N = 16$



b) $N = 24$



c) $N = 32$

Fig. 5 Crank-Nicolson scheme costs.

where Δt^{CN} is the time step used in the Crank-Nicolson scheme which will, in general, be different than that used in the explicit scheme.

Figure 5 presents the correlation results for $N = 16$, 24, and 32 for several time-step sizes. Unlike the explicit scheme, CPU costs for the Crank-Nicolson scheme do not depend on the type of boundary condition. Correlations for the Crank-Nicolson scheme CPU time as a function of the number of grid points, N, and the number of time steps, N_t, were made up of two parts: 1) the one-time fixed cost associated with preparing and inverting the coefficient matrix; and 2) the costs associated with preparing the right sides, solving for expansion function coefficient vector, and transforming it to real space.

A three-term polynomial correlation was used to approximate CPU costs of the fixed part of the Crank-Nicolson scheme

$$\text{CPU}^f = C_3^f N^3 + C_2^f N^2 + C_1^f N \tag{30}$$

Data for $N = 20$, 32, and 40 were used for calibration. The N^3 term reflects part of cost of the operations involved in matrix inversion. For the CPU cost incurred at each time step of the Crank-Nicolson scheme, a three-term polynomial was used

$$\text{CPU}^t = (C_2^t N^2 + C_1^t N + C_0^t) N_t \tag{31}$$

This part of the correlation was calibrated with CPU timings for $N = 16$, 24, and 32. Finally, the CPU time correlation for Crank-Nicolson time integration is

$$\text{CPU} = C_3^f N^3 + C_2^f N^2 + C_1^f N + (C_2^t N^2 + C_1^t N + C_0^t) N_t \tag{32}$$

This correlation predicted the CPU time for $N = 40$ for $N_t = 400$ with an error of only -2.8%. The correlation coefficients resulting from this study are also presented in Table 2.

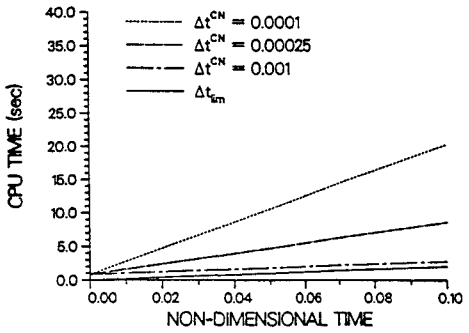
Explicit/Crank-Nicolson Comparison

In this section, the computational times for the explicit and Crank-Nicolson schemes are compared using the correlations developed in the previous sections. In making this comparison, six plots are first constructed to compare CPU time vs nondimensional time for both schemes. These are presented in Fig. 6 for both the adiabatic and homogeneous Dirichlet boundary conditions, presenting three different Crank-Nicolson time steps in each, and comprising the $N = 16$, 24, and 32 cases.

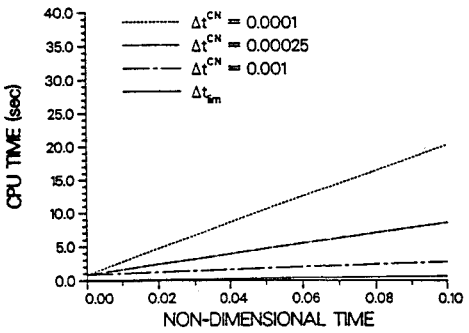
The component figures of Fig. 6 indicate that for several combinations of number of collocation points, boundary condition types, and nondimensional ending time, the Crank-Nicolson scheme can be considerably less costly than the explicit scheme. For example, the adiabatic case with $N = 32$, and the homogeneous Dirichlet case with $N = 24$ and 32, illustrate this well. However, for certain combinations of N and boundary condition type, a very long nondimensional ending time will be required in order for the Crank-Nicolson scheme to become economically competitive. This occurs, for example for the adiabatic case, with $N = 16$.

It is desirable to have an aid for estimating what size of Crank-Nicolson time step retains adequate accuracy in the solution. The accuracy of the Crank-Nicolson scheme using a time-step size equal to the maximum explicit time step is usually more than adequate. Crank-Nicolson solutions found using time steps in excess of 10 times the explicit limit often have more than acceptable accuracy. For example, for homogeneous Dirichlet boundary conditions and a unit top hat initial profile, the Crank-Nicolson solution found using $N = 20$ and $\Delta t = 10\Delta t_{\text{lim}} \approx 0.0025$ had a maximum absolute error of 0.141×10^{-3}.

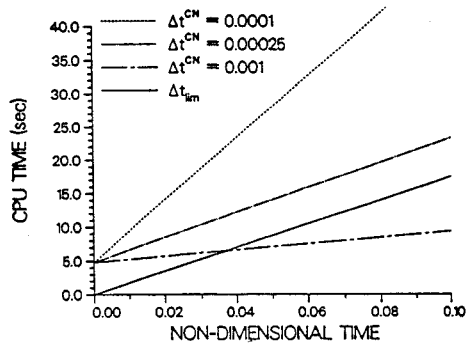
In determining the economics of Crank-Nicolson vs explicit, then, in view of the above use of a multiplier of the maximum



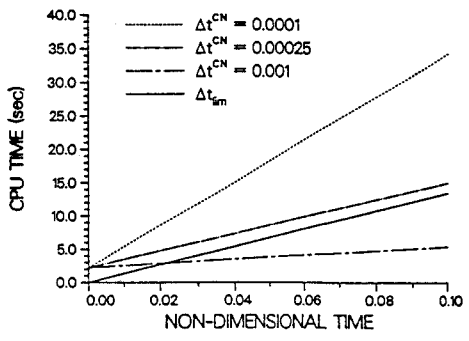
a) Adiabatic, $N = 16$



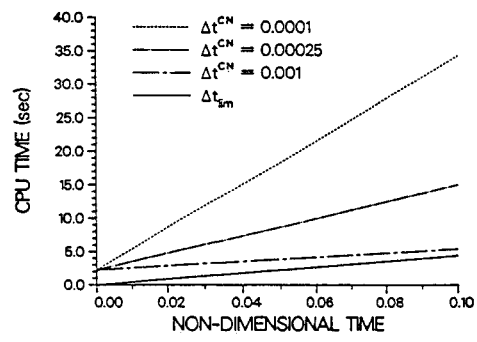
d) Dirichlet, $N = 16$



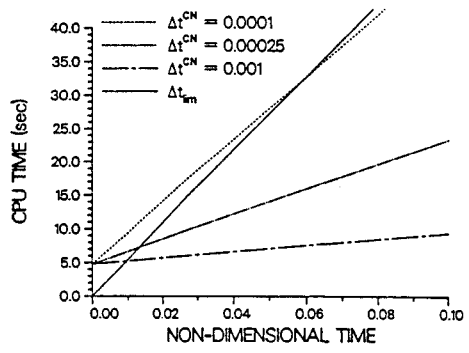
b) Adiabatic, $N = 24$



e) Dirichlet, $N = 24$



c) Adiabatic, $N = 32$



f) Dirichlet, $N = 32$

Fig. 6 Explicit/Crank-Nicolson cost comparison.

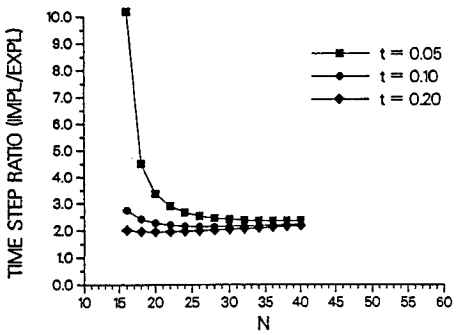


Fig. 7 Time-step ratio for equal cost; Dirichlet boundary conditions.

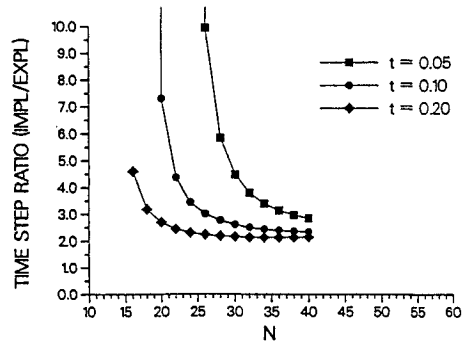


Fig. 8 Time-step ratio for equal cost; adiabatic boundary conditions.

explicit time step, it would appear that the determination of that multiple of the explicit limit time step, which when used in the Crank-Nicolson scheme, results in a computational cost identical to that of the explicit scheme, would be of value in assessing the relative economics of the two schemes. This approach is adopted in this work and is accomplished by equating the two expressions for CPU time as determined earlier by correlation. This results in an expression for the time-step size ratio requiring equal cost for the two schemes given by

$$\left.\frac{\Delta t^{CN}}{\Delta t_{\lim}}\right|_{\text{equal cost}} = \frac{(C_2^i N^2 + C_1^i N + C_0^i)t/\Delta t_{\lim}}{(C_1^e N \log_2 N + C_0^e)t/\Delta t_{\lim} - (C_3^f N^3 + C_2^f N^2 + C_1^f N)} \tag{33}$$

It is noted that this ratio depends on $\Delta t_{\lim}$ itself, and therefore will depend on the boundary condition type. Equation (33) does, however, provide a useful measure of the relative cost of the two schemes within an accuracy context. For example, if one were to assume that a time-step ratio of 10 provides for acceptable accuracy in the Crank-Nicolson scheme, then from Fig. 8, for the adiabatic boundary condition with $N = 18$ and $t = 0.20$, the Crank-Nicolson scheme will be more economical and provide for acceptable accuracy in the time-step ratio range from 3 to 10; the Crank-Nicolson scheme will be more economical for all time-step ratios greater than 3.

As noted in the above example, the time-step ratios determined by Eq. (33) are presented in Figs. 7 and 8 for the homogeneous Dirichlet and adiabatic boundary conditions, respectively. The curves on these figures correspond to different simulation end times and represent the locus of parameter combinations for which the computational costs for the explicit scheme and the Crank-Nicolson scheme are equal; for time-step ratios above each curve, the Crank-Nicolson costs are lower than the explicit scheme costs, and vice versa for the region below the curves. These two figures, Figs. 7 and 8, can then be used to determine under which conditions the explicit or Crank-Nicolson schemes should be used.

Conclusions

The Chebyshev spectral method has been formulated to include the application of general Robin boundary conditions in both an explicit and Crank-Nicolson time advance formulation. Convergence of the method was demonstrated for two examples, one of which is suitable for providing spectral convergence. Both time step and grid size convergence were examined.

Timing studies were performed for both the explicit scheme and the Crank-Nicolson scheme, and the results correlated with number of collocation points, time-step size, and the length of the simulation. Studies were conducted to determine the maximum allowable time-step size for the explicit scheme, and it was observed that this time-step size depends on the boundary condition type. In fact, for the general Robin boundary condition, this time-step size displays a continuous dependence on the Biot modulus; this dependence has been correlated and that correlation is presented in this article.

As a result of the timing studies, it is possible to ascertain under what conditions the Crank-Nicolson scheme will be economically superior to the explicit scheme. An expression that provides the time-step ratio for equal cost of the two schemes has been provided and graphs illustrating this dependence have been provided for the cases of homogeneous Dirichlet and adiabatic boundary conditions. It would appear that the Crank-Nicolson scheme generally provides for significantly lower cost than the explicit scheme while still providing more than acceptable accuracy.

Acknowledgments

The authors acknowledge the financial support of the Natural Science and Engineering Research Council of Canada for their financial support of this project in the form of an operating grant to G. E. Schneider and a postgraduate scholarship to K. Wittich.

References

[1]Schneider, G. E., "Discrete Simulation of Solid/Liquid Phase Change Energy Transfer Including Free Convective Motion," Final Rept. to Communications Research Center, Dept. of Communications, Ottawa, Ontario, Canada, 1984.

[2]Canuto, C., Hussaini, M. Y., Quarteroni, A., and Zang, T. A., *Spectral Methods in Fluid Dynamics*, Springer-Verlag, New York, 1988.

[3]Gottlieb, D., Hussaini, M. Y., and Orszag, S. A., "Theory and Application of Spectral Methods," *Spectral Methods for Partial Differential Equations*, edited by R. G. Voigt, D. Gottlieb, and M. Y. Hussaini, SIAM, Philadelphia, PA, 1984, pp. 1–54.

[4]Schneider, G. E., and Wittich, W., "Assessment of Boundary Condition Implementation Schemes Using the Spectral Method on the Heat Equation," AIAA 24th Thermophysics Conf., AIAA Paper 89-1685, Buffalo, New York, June 12–14, 1989.

[5]Gottlieb, D., and Orzag, S. A., "Numerical Analysis of Spectral Methods: Theory and Applications," SIAM-CBMS, Philadelphia, PA, 1977.

[6]Wittich, W., "The Chebyshev Collocation Spectral Method Applied to the Heat Equation," M.S. Thesis, Univ. of Waterloo, Waterloo, Ontario, Canada, 1990.