Computational Stability and time Truncation of Coupled Nonlinear Equations with Exact Solutions

By
F. Baer and T.J. Simons

Department of Atmospheric Science
Colorado State University
Fort Collins, Colorado

Technical report to the National Science Foundation
Grant No. GA-761
August, 1968

Colorado
State
University

Department of
Atmospheric Science

Paper No. 131

COMPUTATIONAL STABILITY AND TIME TRUNCATION OF COUPLED NONLINEAR EQUATIONS WITH EXACT SOLUTIONS

by

F. Baer and T. J. Simons

Technical Report to
The National Science Foundation
Grant No. GA-761

Department of Atmospheric Science
Colorado State University
Fort Collins, Colorado

August 1968

Atmospheric Science Paper No. 131

CONTENTS

		Page
	Abstract	
1.	Introduction	1
2.	Truncation Schemes	4
3.	Linear Stability Properties	9
4.	Multi-Component System	15
5.	Numerical Calculations	25
6.	Conclusion	38
	Acknowledgements	41
	References	42
	Appendix	43

Computational Stability and Time Truncation of Coupled Nonlinear Equations with Exact Solutions

by

F. Baer and T.J. Simons
Colorado State University

ABSTRACT

A general numerical integration formula is presented which generates many of the commonly used one-dimensional finite-difference schemes. A number of these schemes are tested on a simple wave equation; three implicit and three explicit are chosen for further analysis with a nonlinear set of equations with known solutions. A seventh method of the implicit type not requiring iteration is also tested. A transformation is developed which allows the removal of linear terms from the nonlinear equations, thereby avoiding truncation of the linear terms. The results of the analysis show that energy components may have large errors when the total energy shows essentially none, and phase errors may be quite serious without indication from linear analysis. By treating the uncoupled linear terms exactly (no truncation), significant improvement in the numerical solutions ensues. The multi-level implicit schemes give superior results and are to be recommended if computing time is not a criterion. Great care must be taken in interpreting the linear stability criterion; to avoid significant truncation errors, especially for long time integrations, the critical truncation increment should be considerably reduced.

1. Introduction

The problems of computational stability and truncation errors are by no means recent in origin. Indeed, few physical problems are so simple as to yield mathematical representations which lend themselves to analytic solutions. More often than not, the appropriate equations are nonlinear and must be solved numerically with little insight into the exact solutions. To further complicate clarification of the errors arising from numerical computation, one is generally confronted with partial differential equations.

Despite these seemingly overwhelming obstacles, significant progress in studies of computational stability have been made, exemplified by the work of Richtmyer (1957). The traditional approach to such studies is to linearize the nonlinear equations and then compare the exact solutions of the linear system to the solution of the corresponding finite-difference equations. For different truncation procedures the approximations may be evaluated in terms of the true solution. For initial value problems where the linearizing assumption may not be valid for all time, little may be said except for the criterion of computational stability. Moreover, since finite-difference operations must generally be applied in both space and time, highly involved relationships between the truncation intervals evolve.

With reference to problems concerning atmospheric flow, the feasibility of converting the appropriate nonlinear partial differential equations to a finite set of ordinary nonlinear first-order differential equations in time (termed "spectral" equations) has been established. Such equations are generated by assuming the space dependence to be given by a series of known polynomials and solving for the time dependent coefficients through integration over the entire space domain. The technique seems to have been applied first by Silberman (1954) and discussed in detail by Platzman (1960). On the assumption that the series truncation does not create serious errors (a question not yet investigated in detail), or that the finite set of equations is an exact representation of the physical system, one is left with the considerably simpler problem of determining time truncation alone.

The investigation of ordinary differential equations by numerical methods has also not been neglected; see for example, Henrici (1962), or Hildebrand (1956). Again, however, when nonlinear equations are involved, little can be said about truncation errors of initial value problems. Moreover, if wave type solutions exist, error estimates of linear equations may be cast into doubt. Fortunately there exist some nonlinear systems of spectral equations which have analytic solutions. Such systems were first used to describe atmospheric flow by Lorenz (1960). Clearly a comparison between the finite-difference solution of the equations of such a system when compared to the analytic solution will give information on truncation errors as a function of time. Studies with various time differencing schemes have been made on this basis by Lilly (1965) and Young (1968).

A number of finite-differencing schemes have been utilized for integrating ordinary differential equations, and many are a composite of ingenious techniques which have occurred to various scientists and proved useful. In order to test the utility of such schemes, however, it seems worthwhile to generate them in some systematic fashion, thereby establishing a hierarchy of schemes with (we hope) increasing accuracy. One such systematic approach would be to assume that the function to be integrated can be represented by a polynomial which is exact at its known point values. The degree of accuracy of such a polynomial will then be established by the number of known points utilized. We shall show, moreover, that the most popular schemes can be represented by this approach.

To avoid the problem of being overwhelmed by an unmanageably large number of schemes, the schemes were tested by application to a first-order linear wave equation. If a scheme was not able to give good results for this equation, we assume it would not be satisfactory for a more complex system of equations. In this way we were able to reduce the number of schemes to a manageable size. It should be noted that if more points on the time axis are used to develop the interpolation polynomial than there are orders of derivatives in the differential equations, spurious solutions will result--frequently denoted as "parasitic" solutions--which must be handled with great care so as not to obscure the true physical solution.

The remaining schemes (those which gave satisfactory results with the wave equation) were then tested on a low-order spectral system of the type used by Lilly and Young. The system used here however (Baer, 1968) has the added flexibility of involving both linear and nonlinear terms in the first-order system of equations; it furthermore allows for time dependent phase changes which were constrained in previous experiments. Since linear contributions to differential equations may be determined without truncation, their influence has been investigated. Of the techniques which proved most accurate, multi-step methods were included, despite the presence of parasitic solutions. Previous calculations suggest that integral constraints of the system (say, energy or vorticity) were adequate indicators of truncation error when observed during calculation. This conclusion does not seem to be borne out. We shall see that slight phase errors will create amplitude errors in the individual dependent variables which have a tendency to cancel when the integral properties are evaluated. Thus, although the integral constraints will yield a good indication of computational stability (which is also available from linear theory), truncation errors can only be investigated from the detailed behavior of all the dependent variables in the system.

2. Truncation Schemes

As we have indicated, the spectral equations applicable to the atmosphere may be represented quite generally by a nonlinear set of first-order differential equations in time for which analytic solutions are not available unless the set is highly truncated. The dependent variables, which are the expansion coefficients of the space dependent polynomials, may be represented by a vector Ψ such that

$$\Psi = (\psi_i); \quad 1 \le i \le N$$

and the general set of equations may be written as

$$\dot{\Psi} = F(\Psi,t) = (f_i) \qquad (2.1)$$

where F is a vector operator and the dot notation signifies time differentiation. Suppressing indices, we may also state that the scalar equation for any expansion coefficient will be

$$\dot{\psi} = f(\Psi,t) \qquad (2.1a)$$

Because exact solutions in time are not available for (2.1), based on the complicated nature of the functions f, we may expect to know Ψ only at discrete points on the time axis. For simplicity let us assume Ψ (and therefore F) known at equal time increments

$$\begin{aligned} t &= t_0 + j\Delta t \\ j &= 0,1,2 \ldots,\tau \quad . \end{aligned} \qquad (2.2)$$

Over a given interval in time, we may establish by an interpolation formula a continuous function of time which corresponds to the known values at the discrete points given by (2.2). If we consider the continuous variable in time to be given as

$$t = t_0 + (\tau+s)\Delta t \; ; \quad \begin{matrix} -n \le s \le 1 \\ n \le \tau \end{matrix} \qquad (2.3)$$

then by Newton's backward interpolation polynomial (see Milne, 1949),

$$f(t) = \sum_{k=0}^{n} (-)^k \binom{-s}{k} \Delta^k f^\tau \tag{2.4}$$

In (2.4), the quantity in brackets is a binomial coefficient function of s, the superscript on f denotes the increment in time at which the function should be evaluated (from (2.2) the function is known at the time τ), and Δ^k represents the backward difference operator applied k times and has the value,

$$\begin{aligned} \Delta f^\tau &\equiv f^\tau - f^{\tau-1} \\ \Delta^k f^\tau &= \sum_{j=0}^{k} (-)^j \binom{k}{j} f^{\tau-j} \end{aligned} \tag{2.5}$$

Although we have specified that f is known at $\tau + 1$ points, we need not utilize all these values in establishing our polynomial (2.4), and hence we choose merely the last n point values. We may determine how the interpolation polynomial depends on the discrete point values by substituting (2.5) into (2.4) and noting the following identities;

$$\sum_{k=0}^{n} \sum_{j=0}^{k} = \sum_{j=0}^{n} \sum_{k=j}^{n}$$

and

$$\binom{-s}{j} \binom{-s-j}{k-j} = \binom{-s}{k} \binom{k}{j}$$

The polynomial becomes,

$$f(s) = \sum_{j=0}^{n} \alpha_{Ej}(s) f^{\tau-j}$$

$$\alpha_{Ej}(s) \equiv \binom{-s}{j} \sum_{k=0}^{n-j} (-)^{k} \binom{-s-j}{k} \tag{2.6}$$

If we wish to establish the value of f at the point $\tau + 1$, it would be necessary to extrapolate from (2.6); therefore we have used the subscript notation E. We could, however, assume the function known at $\tau + 1$ and write an equation similar to (2.6) which would then allow an interpolation to the point $\tau + 1$ and would read,

$$f(s) = \sum_{j=0}^{n} \alpha_{Ij}(s) f^{\tau+1-j}$$

$$\alpha_{Ij}(s) \equiv \binom{-s+1}{j} \sum_{k=0}^{n-j} (-)^{k} \binom{-s-j+1}{k} \tag{2.7}$$

To establish the value of ψ $(\tau+1)$ we may now substitute either (2.6) or (2.7) into (2.1a) and integrate. The integration may go over any sub-interval of the interpolation polynomial, but clearly not from a time preceding the point τ-n. Selecting the integer p $(p \leq n)$ at which point the function is known and integrating to $\tau+1$, we have,

$$\psi^{\tau+1} = \psi^{\tau-p} + \Delta t \int_{-p}^{1} f(s) ds \tag{2.8}$$

It is interesting to note that use of the extrapolating polynomial yields an Explicit solution for $\psi^{\tau+1}$, whereas the application of the interpolating polynomial leads to an Implicit solution, because the unknown function $f^{\tau+1}$ still exists on the right-hand side of the equation. If we define the integrals over s,

$$\int_{-p}^{1} \alpha_{Ej}(s)ds \equiv \bar{\alpha}_{Ej}(p)$$

$$\int_{-p}^{1} \alpha_{Ij}(s)ds \equiv \alpha_{Ij}(p) \qquad (2.9)$$

where the integrals may be evaluated by noting that the integrals are factorial polynomials in s which may be converted to polynomials in s by use of Sterling's numbers of the first kind (Milne, 1949), we may express the general finite-difference extrapolation formulas as follows:

Explicit: E_{pn}

$$\psi^{\tau+1} = \psi^{\tau-p} + \Delta t \sum_{j=0}^{n} \bar{\alpha}_{Ej} f^{\tau-j}$$

$$(2.10)$$

Implicit: I_{pn}

$$\psi^{\tau+1} = \psi^{\tau-p} + \Delta t \sum_{j=0}^{n} \bar{\alpha}_{Ij} f^{\tau+1-j}$$

We see from (2.10) that a wide variety of finite-difference integration schemes may be selected, and in a systematic fashion. As we increase p and n, we arrive at higher order schemes (more "steps") with the consequent expected increase in accuracy but also additional parasitic roots. Most of the standard numerical integration schemes fall into the classification given by (2.10). For example, the schemes E_{0n} and E_{1n} are generally associated with Adams-Bashforth and Nystrom respectively, whereas the schemes I_{0n}, I_{1n} are referred to as Adams-Moulton and Milne-Simpson respectively. The more involved predictor-corrector or multi-corrector

Table 1. Values of the coefficients $\bar{\alpha}_{Ij}(p)$, $\bar{\alpha}_{Ej}(p)$ for different integration schemes E_{pn}, I_{pn}, their names (if known) and the truncation error based on Taylor's series analysis.

Scheme (p,n)	Name	j=0	j=1	j=2	j=3	j=4	j=5	Truncation Error
E_{01}		3/2	-1/2					$\|5/12(\Delta t)^3\psi^{(3)}\|$
E_{02}	Adams-Bashforth	23/12	-4/3	5/12				$\|3/8(\Delta t)^4\psi^{(4)}\|$
E_{03}		55/24	-59/24	37/24	-9/24			$\|1/3(\Delta t)^5\psi^{(5)}\|$
E_{04}		1901/720	-2774/720	2616/720	-1274/720	251/720		$\|1/5(\Delta t)^6\psi^{(6)}\|$
E_{11}	Leapfrog	2	0					$\|1/3(\Delta t)^3\psi^{(3)}\|$
E_{12}		7/3	-2/3	1/3				$\|1/3(\Delta t)^4\psi^{(4)}\|$
E_{33}	Milne Predictor	8/3	-4/3	8/3	0			$\|1/3(\Delta t)^5\psi^{(5)}\|$
I_{01}	Euler Trapezoidal	1/2	1/2					$\|1/12(\Delta t)^3\psi^{(3)}\|$
I_{02}		5/12	8/12	-1/12				$\|1/24(\Delta t)^4\psi^{(4)}\|$
I_{03}	Moulton Corrector	9/24	19/24	-5/24	1/24			$\|1/36(\Delta t)^5\psi^{(5)}\|$
I_{04}		251/720	646/720	-244/720	106/720	-19/720		$\|1/53(\Delta t)^6\psi^{(6)}\|$
I_{13}	Milne Corrector	1/3	4/3	1/3	0			$\|1/90(\Delta t)^5\psi^{(5)}\|$
I_{14}		29/90	124/90	24/90	4/90	-1/90		$\|1/3(\Delta t)^6\psi^{(6)}\|$
I_{35}	Milne II Corrector	14/45	64/45	24/45	64/45	14/45	0	$\|1/120(\Delta t)^7\psi^{(7)}\|$

schemes would require a sequence of schemes described by (2.10).

A number of schemes whose properties will be investigated are listed in Table 1. Certain omissions will be noted. The E_{00} scheme, which is termed the "Euler forward" is always unstable in terms of fictitious amplification and is consequently of no interest. Similarly, the "Euler backward", I_{00} gives fictitious damping and is therefore ignored. Schemes with p=2 have been shown to yield results not appreciably superior to those for p=1 and their discussion would thus be redundant. For the implicit schemes, the coefficients $\bar{\alpha}_{I3}(1)$, $\bar{\alpha}_{I5}(3)$ vanish, and consequently the lower order forms I_{12}, I_{34} which require as much calculation as I_{13}, I_{35} have been ignored.

The schemes described by (2.10) may be subject to Taylor's series expansion about the point τ; for a given truncation (p,n) there will be an error of order $(\Delta t)^{n+2}$ times the same order of time derivative of ψ, listed in Table 1. When applying these techniques to wave type equations, however, such error estimates may be misleading.

3. Linear Stability Properties

If the schemes listed in Table 1 do not show adequate stability properties when applied to a linear differential equation, we may anticipate their failure with regard to nonlinear differential equations. We shall therefore test them on the simple linear wave equation,

$$\dot{\psi} = -i\rho\psi \qquad (3.1)$$

which could be generated from (2.10) by linearization and neglecting coupling terms. Note that ψ is a complex variable, but let us assume ρ to be real. The true solution of (3.1) shows only one mode which moves about the unit circle in the complex plane with period $2\pi/\rho$ beginning at unity when $t = \frac{2m\pi}{\rho}$. If we now define coefficients

$$\alpha_j = \begin{cases} \bar{\alpha}_{I,j+1} & \bar{\alpha}_{I,j+1} = 0 \quad \text{for } j=n \\ \bar{\alpha}_{Ej} \quad ; & \bar{\alpha}_{Ej} = 0 \quad \text{for } j = -1 \end{cases}$$

we may write both the implicit and explicit finite-difference schemes (2.10) after substitution of (3.1) for the values of the derivatives at the known discrete points by the single relation,

$$(1+i\alpha_{-1}\rho\Delta t)\psi^{\tau+1} = \psi^{\tau-p} - i\rho\Delta t \sum_{j=0}^{n} \alpha_j \psi^{\tau-j} \tag{3.2}$$

The solutions to (3.2) may be determined in a number of ways, but they must all satisfy the characteristic equation

$$(1+i\alpha_{-1}\rho\Delta t)\lambda^{n+1} = \lambda^{n-p} - i\rho\Delta t \sum_{j=0}^{n} \alpha_j \lambda^{n-j} \tag{3.3}$$

where the roots of (3.3) represent the solutions of (3.2). Since we have specified p≤n, there will be n+1 solutions to (3.2), only one of which corresponds to the real "physical" mode. The computational or parasitic modes (n of them) are distributed as follows at $\Delta t=0$; n-p roots begin at the origin, and p+1 roots are distributed equally about the unit circle with the physical mode at $\lambda=1$. As Δt is increased from zero, the roots will change from their initial points.

If the first root, λ_0, represents the "computed physical mode", we may compare it with the true solution. So long as its amplitude remains near unity, their will be no spurious damping or amplification. However, its phase, say θ_0, must also remain near the true phase for accuracy; i.e., we should observe that $-\theta_0/\rho\Delta t$ remains close to unity. The remaining n solutions are parasitic and enter only to disturb the physical solution. So long as their amplitudes remain less than unity (i.e., within the unit circle), they will be damped. If they go outside the unit circle, they will cause amplification and may be classified as "unstable" solutions. If they remain on the unit circle, by suitable choice of initial conditions their effects can be made innocuous.

All the schemes listed in Table 1 have been tested on (3.1). Their characteristic equations may be easily determined by substitution of the tabular coefficients together with the limits (p,n) into (3.3). The roots of these characteristic equations have been determined for various values of $\rho\Delta t$ and the amplitudes of all modes for each scheme have been plotted against $\rho\Delta t$ (abscissa) in Fig. 1. Pursuant to the previous discussion, wherever a mode exceeds unity on the ordinate, it will yield an unstable solution. Clearly the best schemes will be those for which all roots remain stable for the largest value of $\rho\Delta t$.

We may be considerably more precise about the behavior of these schemes by investigating the computational physical mode in more detail--both its amplitude and phase. On Fig. 2 we have plotted for all schemes the amplitude of the computational physical mode (and amplitudes of parasitic modes when they are within the ordinate scale) on the upper graph and the ratio $-\theta_0/\rho\Delta t$ on the lower graph against $\rho\Delta t$ on the abscissa. Here we may isolate the best schemes. Whereas from Fig. 1 we might have thought that scheme I_{01} was best because it is stable for all values of Δt, we see from Fig. 2 that this scheme (trapezoidal) has serious phase errors for reasonable values of $\rho\Delta t$.

Based on Fig. 2, we have selected three schemes in the explicit group and three in the implicit group for further study. Choosing $\rho\Delta t<.4$, we see that E_{03}, E_{11} and E_{33} are the best, whereas for the implicit schemes, the obvious choices are I_{01}, I_{13}, I_{35}. Scheme I_{01} was selected because of the strong stability property of its amplitude and also because of its general popularity, although its phase characteristics are less desirable.

An interesting sidelight to the selection of suitable finite-difference schemes is exemplified by Fig. 3. Suppose one would like a scheme no greater than two-step for which the coefficients

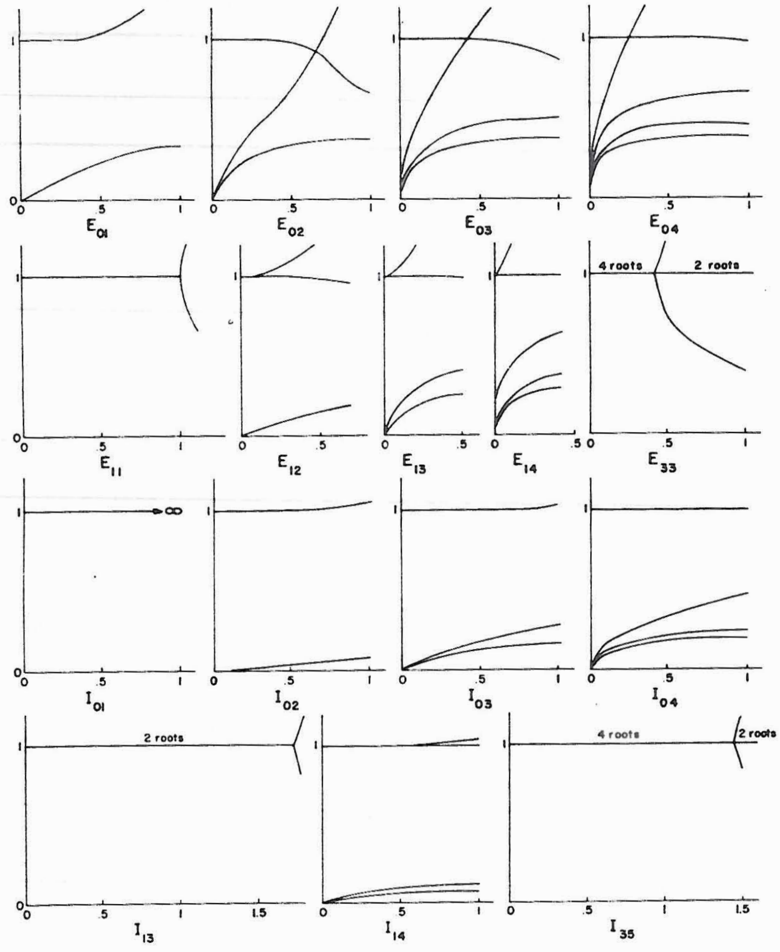

Fig. 1 Amplitudes of the roots of the simple linear wave equation $\dot{\psi} = i\rho\psi$, for various truncation schemes, plotted against $\rho\Delta t$ on the abscissa.

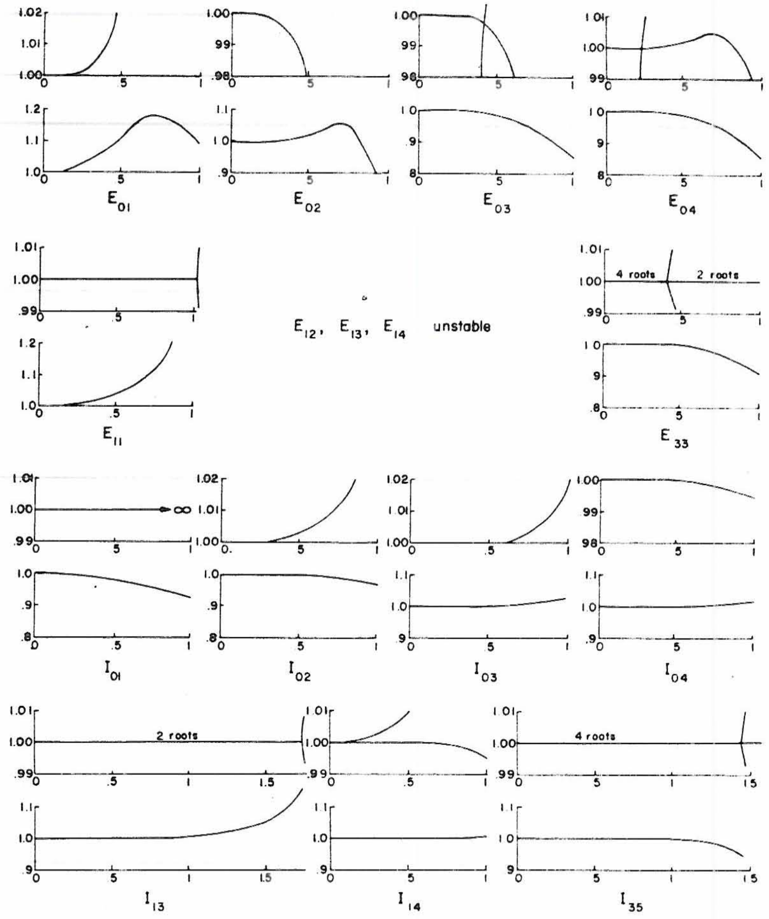

Fig. 2 Amplitude (upper) and phase (lower) of the physical root of the truncated linear wave equation plotted against $\rho\Delta t$. The phase is given proportional to $\rho\Delta t$.

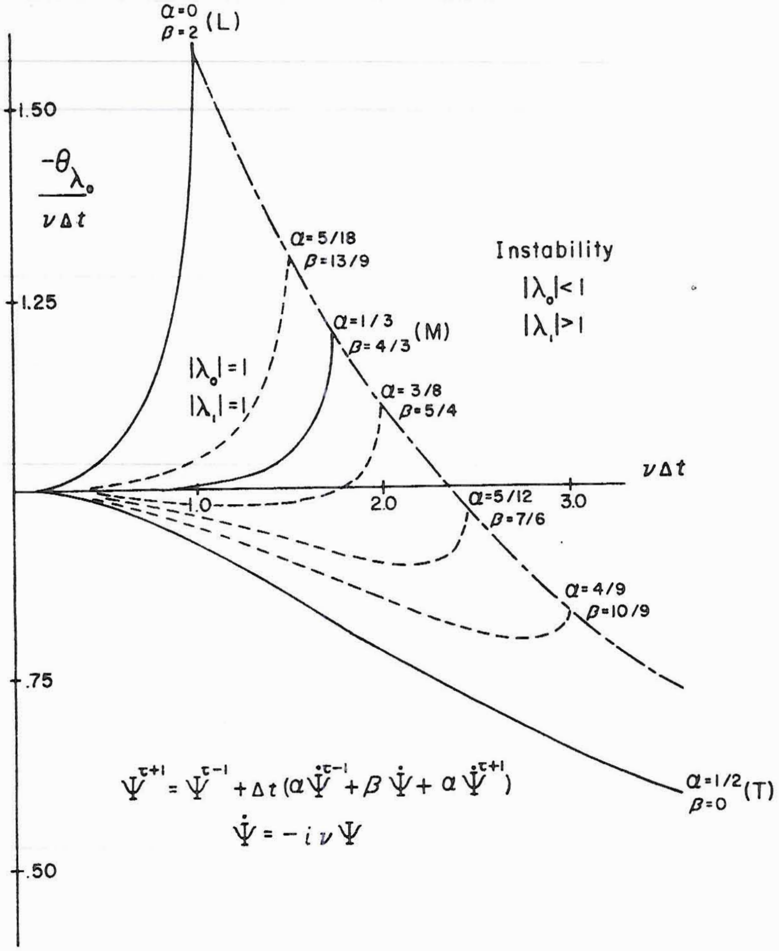

Fig. 3 Phase errors for a class of two-level truncation schemes in terms of the true phase including the leapfrog, trapezoidal and Milne schemes.

could be varied such that the most favorable properties may be chosen. Let the scheme be represented in terms of the arbitrary coefficients (α,β)

$$\begin{aligned} \psi^{\tau+1} &= \psi^{\tau-1} + \Delta t(\alpha\dot{\psi}^{\tau-1}+\beta\dot{\psi}^{\tau}+\alpha\dot{\psi}^{\tau+1}) \\ 2\alpha + \beta &= 2 \end{aligned} \qquad (3.4)$$

and test it on the equation $\dot{\psi} = -i\nu\psi$. Fig. 3 shows, for various combinations of α,β that the phase properties in the stable range (where both the real physical and the parasitic roots have amplitude unity) are effectively bounded by the error curves for the leapfrog (L-E_{11}) on the one hand, and the trapezoidal (T-I_{01}) on the other. The Milne scheme (M-I_{13}) is undoubtedly one of the best which satisfies the criteria of (3.4).

4. Multi-Component System

The coupled set of nonlinear first-order differential equations on which the six schemes which survived the linear analysis of the last section will be tested is part of the group of low-order spectral systems which were systematically developed by Platzman (1962) for the barotropic vorticity equation, but which also have applicability for baroclinic problems. The system under consideration involves an arbitrary zonal flow interacting with a single planetary wave composed of two complex components (describing its latitudinal variability) in a rotating atmosphere with spherical geometry. Details of this system, including the exact solutions (elliptic functions) have been presented by Baer (1968). If the zonal coefficients (real) are denoted as $\psi_\gamma(t)$ where $\gamma = 2m+1$, $m \leq M$, and the complex wave coefficients are described by the terms $\psi_\alpha(t)$, $\psi_\beta(t)$, the differential equation for the zonal terms may be written,

$$\dot{\psi}_{\gamma} = 2a_{\gamma} \text{ im } \psi_{\alpha}\psi_{\beta}^{*}$$

The zonal coefficients can be solved in terms of one coefficient ψ_n, by integration of the above equation. The time relationship thus developed between the zonal coefficients is unaltered if the integration is performed by numerical means, whereby we find that $\psi_{\gamma} = (a_{\gamma}/a_n)\psi_n + s_{\gamma}$. The system to be integrated therefore involves only three variables, ψ_n, ψ_{α}, ψ_{β} and is,

$$\dot{\psi}_n = 2a_n \text{ Im } \psi_{\alpha}\psi_{\beta}^{*}$$

$$\dot{\psi}_{\alpha} = -i\rho_{\alpha}\psi_{\alpha} + ih_{\alpha\beta}\psi_{\beta} + ig_{\alpha\alpha}\psi_n\psi_{\alpha} + ig_{\alpha\beta}\psi_n\psi_{\beta} \qquad (4.1)$$

$$\dot{\psi}_{\beta} = -i\rho_{\beta}\psi_{\beta} + ih_{\beta\alpha}\psi_{\alpha} + ig_{\beta\beta}\psi_n\psi_{\beta} + ig_{\beta\alpha}\psi_n\psi_{\alpha}$$

where $\rho_{\alpha,\beta} \equiv \nu_{\alpha,\beta} - h_{\alpha\alpha,\beta\beta}$. In matrix notation we find,

$$\dot{\psi}_n = \tilde{\psi}^{*}H\psi \qquad ; \quad \psi_n \text{ real scalar}$$

$$\dot{\psi} = (A + \psi_n D)\psi \qquad ; \quad \psi = \begin{pmatrix} \psi_{\alpha} \\ \psi_{\beta} \end{pmatrix}$$

$$A = A_1 + A_2 \qquad ; \quad \psi_{\alpha,\beta} = \frac{1}{\sqrt{2}} B_{\alpha,\beta}(t) e^{i\theta_{\alpha,\beta}(t)} \qquad (4.2)$$

$$A_1 \equiv -i\begin{pmatrix} \rho_{\alpha} & 0 \\ 0 & \rho_{\beta} \end{pmatrix} \quad ; \quad A_2 \equiv i\begin{pmatrix} 0 & h_{\alpha\beta} \\ h_{\beta\alpha} & 0 \end{pmatrix}; \quad D \equiv i\begin{pmatrix} g_{\alpha\alpha} & g_{\alpha\beta} \\ g_{\beta\alpha} & g_{\beta\beta} \end{pmatrix}$$

$$H \equiv ia_n \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = -\tilde{H} \quad ;$$

where the tilde denotes transposition. The physical significance of the constants which depend on spectrum truncation and initial conditions may be found in the paper by Baer (1968). Three different sets of conditions were used in this study, and the numerical values of the variables may be found in the Appendix.

System (4.2) involves both linear and nonlinear terms; the uncoupled linear effects are denoted by the matrix A_1, and the purely coupled linear terms by the matrix A_2. The term $\psi_n D\psi$ represents the nonlinear effect. The linear terms may be removed from truncation in a numerical integration by recognizing their exact influence. Let us therefore define the vector

$$\chi \equiv e^{-Gt}\psi \qquad (4.3)$$

such that the second differential equation of (4.2) becomes

$$\dot{\chi} \equiv e^{-Gt}(A-G+\psi_n D)e^{Gt}\chi \qquad (4.4)$$

where G is a matrix which we shall allow to take on the three possible values:

(a) $G = 0$: (TL)
(b) $G = A_1$: (TC)
(c) $G = A$: (EL)

If we now integrate (4.4) and the first of (4.2) by the different allowed schemes (see Section 3), we see that when $G = 0$ we truncate all linear terms, when $G = A$ we truncate the coupling terms, and when $G = A$ we deal with the exact linear solution.

We have seen from the last section that one level implicit schemes are always stable and do not add parasitic modes. They have the disadvantage, however, of requiring an iteration process for calculation with an unspecified convergence rate. We have consequently added another scheme to our set of six (3-implicit and 3-explicit) which is in effect an implicit scheme which can be solved without iteration. Consider the following two implicit techniques applied to the first of (4.2):

$$\psi_n^{t+\Delta t} - \psi_n^{t} = \tfrac{1}{2}\Delta t\left((\tilde{\psi}^*H\psi)^{t} + (\tilde{\psi}^*H\psi)^{t+\Delta t}\right)$$

$$\psi_n^{t+\Delta t} - \psi_n^{t} = \tfrac{1}{4}\Delta t\left(\tilde{\psi}^{*t+\Delta t} + \tilde{\psi}^{*t}\right)H\left(\psi^{t+\Delta t} + \psi^{t}\right)$$

If we now combine these two schemes by taking the first twice and subtracting the second, we find,

$$\psi_n^{t+\Delta t} - \psi_n^{t} = \frac{\Delta t}{2}\left(\tilde{\psi}^{*t}H\psi^{t+\Delta t} + \tilde{\psi}^{*t+\Delta t}H\psi^{t}\right) \qquad (4.5)$$

On the assumption that the quantities ψ_n, ψ, ψ^* are known at time t, (4.5) is linear in the terms at t+Δt and may thus, in combination with a difference form of the type (4.5) used on (4.4), be solved for the variables at t+Δt. In terms of the vector X (5 elements) defined as

$$X \equiv \begin{pmatrix} \psi_n \\ \psi \\ \psi^* \end{pmatrix} \qquad (4.6)$$

the non-iterative implicit scheme (IM) may be represented as follows,

$$X^{t+\Delta t} = R_1^{-1}R_2X^{t} \qquad (4.7)$$

where the matrices R_1 and R_2 are written,

$$R_1 = \begin{pmatrix} 1 & -\frac{\Delta t}{2}\bar{\psi}^{*t}H & \frac{\Delta t}{2}\bar{\psi}^{t}H \\ -\frac{\Delta t}{2}D\psi^{t} & e^{-G\frac{\Delta t}{2}} - \frac{\Delta t}{2}(A-G+\psi_n^{t}D) & 0 \\ -\frac{\Delta t}{2}D^*\psi^{*t} & 0 & e^{-G^*\frac{\Delta t}{2}} - \frac{\Delta t}{2}(A^*-G^*+\psi_n^{t}D^*) \end{pmatrix}$$

$$R_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{G\frac{\Delta t}{2}} + \frac{\Delta t}{2}(A-G) & 0 \\ 0 & 0 & e^{G^*\frac{\Delta t}{2}} + \frac{\Delta t}{2}(A^*-G^*) \end{pmatrix}$$

Before proceeding to a discussion of the numerical calculations of the different schemes, let us consider the linearized coupled equations and the finite-difference solution to these equations. The linearization of (4.2) may be accomplished by assuming that $\psi_n \rightarrow \bar{\psi}_n$=constant where it multiplies either ψ_α or ψ_β in the second equation of (4.2). The linearized equation may thus be expressed as,

$$\dot{\psi} = \bar{G}\psi$$

$$\bar{G} \equiv A+\bar{\psi}_n D = i\begin{pmatrix} -n_\alpha & \bar{G}_{\alpha\beta} \\ \bar{G}_{\beta\alpha} & -n_\beta \end{pmatrix} \qquad (4.8)$$

where the elements of $\bar{G}$ can be established from the definitions given in (4.2). The roots of $\bar{G}$ are listed in Table 2 together with the form of the model matrix $\bar{S}$, where

$$\bar{G} = \bar{S}i\bar{\Lambda}\bar{S}^{-1}$$

and $\bar{\Lambda}$ is the root matrix of $\bar{G}$. The solution to (4.8) may be written formally as,

$$\begin{aligned} \psi(t) &= e^{\bar{G}t}\psi(t=0) \\ &= \bar{S}e^{i\bar{\Lambda}t}\bar{S}^{-1}\psi(t=0) \end{aligned} \tag{4.9}$$

If the roots of $\bar{G}$ are pure imaginary, no physical amplification will take place and computational stability can be easily defined. The physical stability condition implied in $\bar{G}$ has been discussed by Baer (1968) and need not be repeated here. We shall concern ourselves with physically stable situations.

Table 2. Values of the roots and model matrices for various forms of G.

G	ζ	Λ_i	S
0	ψ	0	I
A_1	ψ	ρ_α, ρ_β	I
A	$S^{-1}\psi$	$-\dfrac{\rho_\alpha+\rho_\beta}{2} \pm \frac{1}{2}\left((\rho_\alpha-\rho_\beta)^2+4h_{\alpha\beta}h_{\beta\alpha}\right)^{\frac{1}{2}}$	$-i\begin{pmatrix} \rho_\beta+\Lambda_1 & \rho_\beta+\Lambda_2 \\ h_{\beta\alpha} & h_{\beta\alpha} \end{pmatrix}$
$\bar{G}$	---	$\nu_{1,2} = -\dfrac{\eta_\alpha+\eta_\beta}{2} \pm\left((\eta_\alpha-\eta_\beta)^2+4\bar{G}_{\alpha\beta}\bar{G}_{\beta\alpha}\right)^{\frac{1}{2}}$	$-i\begin{pmatrix} \eta_\beta+\nu_1 & \eta_\beta+\nu_2 \\ \bar{G}_{\beta\alpha} & \bar{G}_{\beta\alpha} \end{pmatrix}$

Let us now apply the leapfrog scheme to (4.8) with the option that some of the linear terms may be extracted, as we have done in (4.4). The appropriate form of (4.8), using the transformation (4.3), becomes

$$\dot{\chi} = e^{-Gt}(\bar{G}-G)\psi$$

$$\chi^{t+\Delta t} = \chi^{t-\Delta t} + 2\Delta t\ e^{-Gt}(\bar{G}-G)\psi^{t} \qquad (4.10)$$

Here G may take on any of the three values 0, A_1, A. Since we wish to compare the solutions of the second of (4.10) with (4.9), we return (4.10) to the variable ψ. Noting now that we may establish the roots of G and write a root matrix Λ,

$$G = Si\Lambda S^{-1} \qquad (4.11)$$

where the roots and the model matrix for different matrices G are listed in Table 2, we find for (4.10) using (4.11) and (4.3),

$$\zeta^{t+\Delta t} = e^{2i\Lambda\Delta t}\ \zeta^{t-\Delta t} + 2\Delta t(S^{-1}\bar{G}S-i\Lambda)\zeta^{t}$$

$$\zeta^{t} \equiv S^{-1}\psi^{t} \qquad (4.12)$$

Since the elements of ζ are a linear combination of the elements of ψ, they will have the same solutions (roots). By the usual method of establishing an amplification matrix for multi-step equations (Richtmyer, 1957), we define the vector ξ as

$$\xi^{t+\Delta t} = \zeta^{t}$$

and we get the solution to (4.12) in the form,

$$\begin{pmatrix} \zeta \\ \xi \end{pmatrix}^{t+\Delta t} = \begin{pmatrix} 2\Delta t(S^{-1}\bar{G}S)-i\Lambda & e^{2i\Lambda\Delta t} \\ I & 0 \end{pmatrix} \begin{pmatrix} \zeta \\ \xi \end{pmatrix}^{t} \tag{4.13}$$

where I is the unit matrix. The root equation for the amplification matrix in (4.13) is given as

$$\begin{vmatrix} 2i\Delta t(S^{-1}\bar{S}\bar{\Lambda}\bar{S}^{-1}S-\Lambda) - \lambda I & e^{2i\Lambda\Delta t} \\ I & -\lambda I \end{vmatrix} = 0 \tag{4.14}$$

which is, in general, a 4th order equation in the roots. Two of these roots are physically real, and the other two are parasitic. The real roots should be compared with the roots of the exact solution, $e^{i\bar{\Lambda}\Delta t}$. So long as Δt remains within the limits of computational stability, the roots will have amplitude of unity and we may therefore consider only the phase angles. Thus if the roots of (4.14) are,

$$\lambda_j = e^{i\theta_{\lambda j}} \tag{4.15}$$

we may compare the phase angles for the finite-difference solution to those of the exact solution by the ratio, $\theta_{\lambda j}/\nu_j\Delta t$. These ratios are shown for each of the approximations G = 0, A_1 , A on Fig. 4, plotted against the non-dimensional time unit Δt, where time has been non-dimensionalized by the earth's rotation rate. The data used in determining the roots was taken from case CA and is given in the appendix. The values of the frequencies $\nu_{1,2}$ are,

$\nu_1 = .310$ $\qquad\qquad$ $\nu_2 = .042$

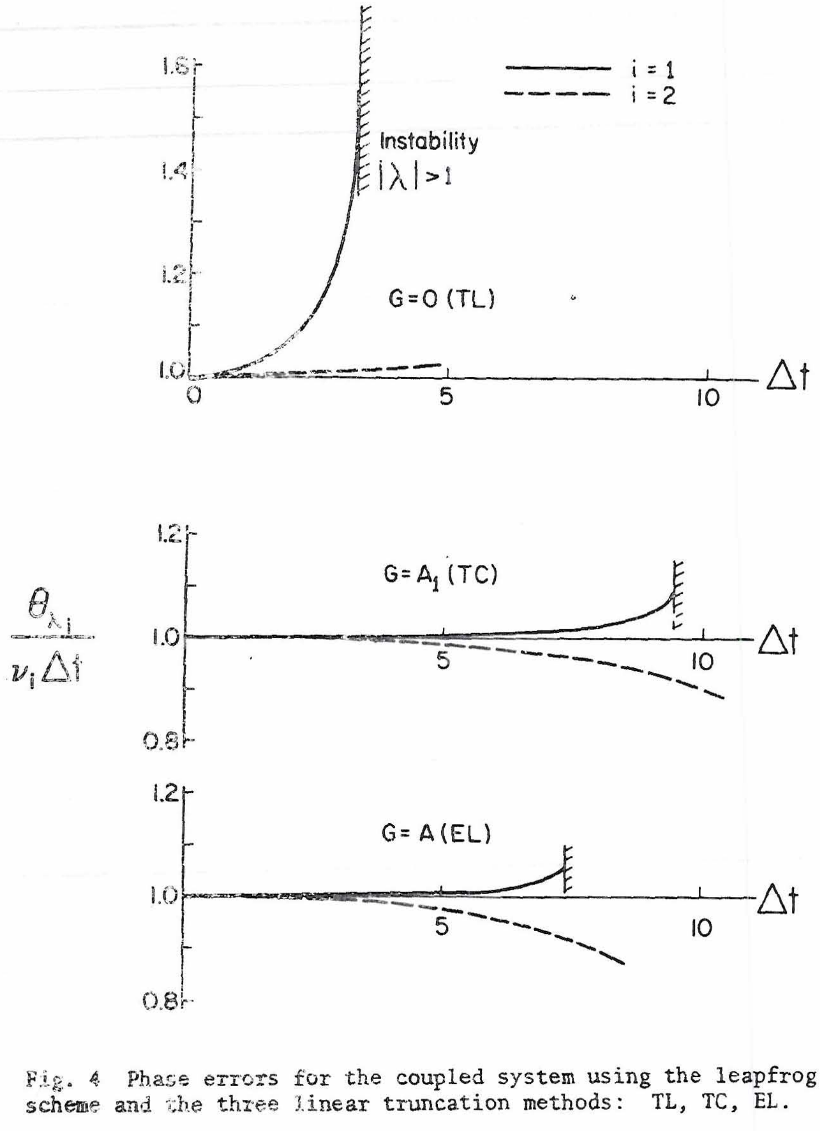

Fig. 4 Phase errors for the coupled system using the leapfrog scheme and the three linear truncation methods: TL, TC, EL.

We shall have occasion to compare these values to the exact frequencies of the nonlinear solution in the next section.

The phase errors for the approximation G = 0 may be readily determined since (4.14) reduces to the equation,

$$\lambda^2 - 2i\Delta t\nu_j\lambda - 1 = 0$$

from which we see that the phase angles $\theta_{\lambda j}$ (4.15) are given by

$$\theta_{\lambda j} = \sin^{-1} (\nu_j \Delta t)$$

a result identical to the one arrived at in Section 3 for uncoupled systems. The stability criterion and phase error at the stability point are

$$\Delta t = \frac{1}{\nu_j} \quad ; \quad \left(\frac{\theta_{\lambda j}}{\nu_j \Delta t} \right)_{\text{critical}} = \pi/2$$

Returning to the discussion of Fig. 4, we see that removal of the uncoupled terms ($G = A_1$) leads to a much more stable calculation with considerably lower errors in phase. Total truncation (G = 0) is clearly the worst case, whereas including an exact treatment of the coupling terms in A_2 does not improve the stability or phase errors; in fact, the extraction of more exact information in this case creates larger truncation errors. It should be noted that these results refer to a particular set of initial conditions, and are subject to change for different conditions. However, we may conclude with some confidence that the exact treatment of uncoupled linear terms will yield solutions to the nonlinear equations with less error for a given truncation element, Δt. We shall not consider the linearized equations for the other truncation schemes but proceed directly to the numerical calculation of the nonlinear equations.

5. Numerical Calculations

Three different sets of initial conditions were used to test the truncation schemes and they are listed in the appendix, denoted respectively as cases CA, CB, CC. Since the exact solutions to (4.2) are known, any variable determined from a numerical integration will be represented normalized by its exact value. For each case of initial data, the three methods for dealing with the linear terms were applied (TL, TC, EL) and three different time increments (Δt) were used; the scales of the time increments were determined from the characteristic frequencies of the cases. All seven truncation schemes discussed in Section 3 as having satisfactory linear properties were tested, and are listed in Table 3.

Let us now concentrate our attentions on the features of case CA, which was integrated numerically in excess of 51 days. Since this case has an exact nonlinear exchange period of 3.452 days, the integration period should be long enough to highlight important errors. The exact frequencies, and there are two because two wave components ψ_α, ψ_β exist, are $\nu_1 = .309$, $\nu_2 = .0192$. The first of the two frequencies calculated by linear theory (see Section 4) compares remarkably well with the first exact frequency, but the second is more than twice as large. However, because of the difference in magnitude of these frequencies, the first (larger) frequency will essentially determine the stability criterion. Using the first frequency and the linear (uncoupled) solutions for the different schemes developed in Section 3, we list in Table 3 the stability condition, $\Delta t_{\max}$, which the linear theory would indicate.

A common procedure for establishing stability and truncation errors is to investigate the development in time of some integral property of the system--generally conservative--as was done by both Lilly (1965) and Young (1968). For simple atmospheric flow problems, energy is the logical choice although Young also included the vorticity. To indicate the behavior of the total energy

Table 3. Total energy normalized by the exact value for case CA after 51.77 days for seven schemes three different time steps and different treatment of linear terms together with the linear stability criterion. In case of oscillation, range is tabulated.

Scheme	TL			TC			EL		
Stability Δt=(hrs.)→	2.07	4.14	8.28	2.07	4.14	8.28	2.07	4.14	8.28
E11 $\Delta t \le 12$ hrs.	.999 1.001	.991 1.006	.934 1.098	1.000	.998 1.001	.988 1.006	.999 1.003	.968 1.045	.889 81.881
E03 $\Delta t \le 5$ hrs.	.993	.822	overflow (11 days)	1.000	.989 .991	.855 .857	.999	.979 .980	.801 .812
E33 $\Delta t \le 5$ hrs.	1.000	.999 1.002	overflow (8 days)	.998 1.002	.988 1.013	overflow (32 days)	.878 1.409	overflow (50 days)	
IM $\Delta t \le \infty$	-	-	-	1.000	1.000 1.002	1.000 1.007	.999 1.000	.995 1.000	.982 1.000
I01 $\Delta t \le \infty$	1.000	1.000 1.001	.999 1.005	1.000	1.000 1.001	.999 1.003	.999 1.000	.998 1.001	.999 1.011
I13 $\Delta t \le 21$ hrs.	1.000	1.000	.999 1.001	1.000	1.000	1.000	1.000	1.000	.999 1.000
I35 $\Delta t \le 16$ hrs.	1.000	1.000	.999 1.003	1.000	1.000	.999 1.000	1.000	1.000	.999 1.001

of our solution (case CA) with time, we have prepared Table 3 in which we describe the total energy (conserved in the exact solution) for the different truncation schemes, different truncation intervals $\Delta t = 2.07, 4.14, 8.28$ hours, and different treatment of the linear terms. The energies have been listed after 51.77 days unless an oscillation occurs, in which case its range is tabulated. As indicated above, we have also listed the stability condition based on linear theory.

Unquestionably, the stability properties are well described by the total energy and correspond to those anticipated from linear theory. Where damping is predicted, as in scheme E03, the tabular values are in agreement. Where parasitic oscillations are anticipated (E11), they appear in the table. Further expected results show that the solutions deteriorate for increased Δt and that implicit schemes are generally superior (for given Δt) than explicit ones. A further observation, not previously investigated, is the improvement of the solution from TL to TC; i.e., when the uncoupled linear terms are treated exactly. If, however, one proceeds to treat all linear terms exactly (EL), the results appear somewhat less stable, as expected from the linear analysis (Section 4).

The above information is indeed valuable; however, it must be emphasized that the behavior of the total energy with time is not necessarily an indicator of the behavior of the detailed character of the solution. As we shall see, the individual amplitudes of the wave components may be seriously in error with no indication from the total energy. Moreover, the phase angles and wave velocities of the components from the truncated calculations may have no relation to the true solution, although the total energy is well conserved. To establish this fact , among others, we shall proceed to a detailed discussion of the calculations.

The component amplitudes which make up the total energy in our equations may be represented when we describe the truncated

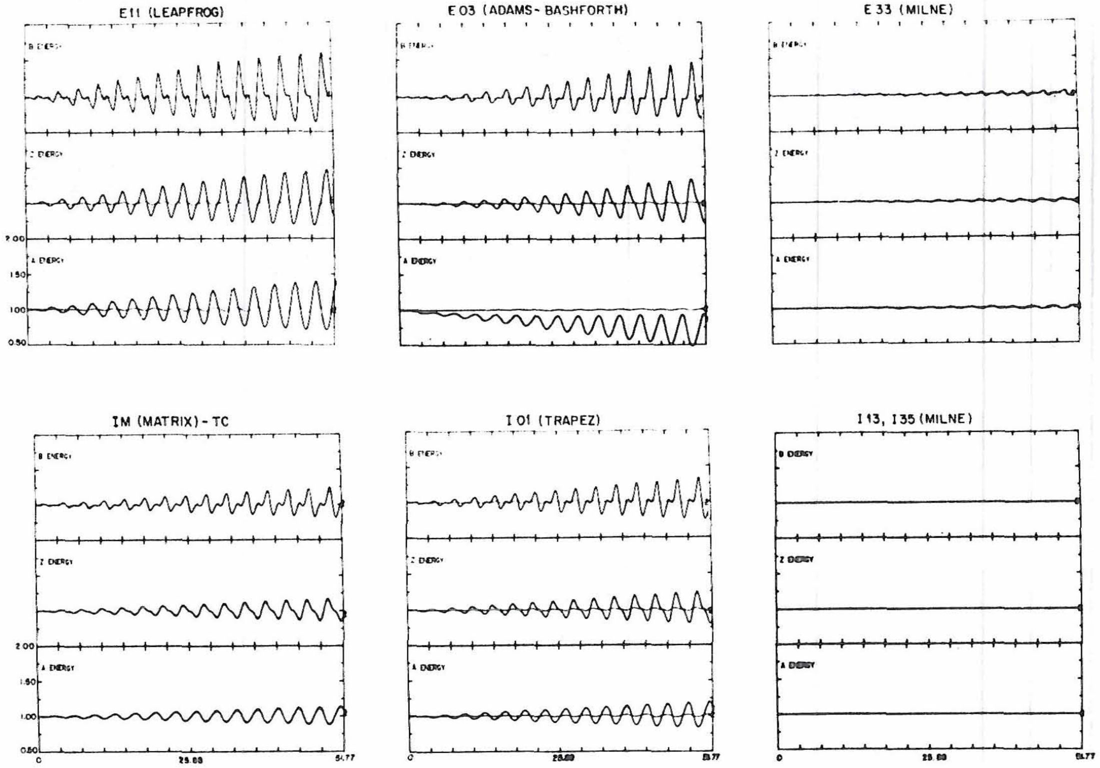

Fig. 5 Energy in the zonal and α- and β-waves as a function of time in days (abscissa) normalized by their exact values for the seven schemes which had favorable linear properties. Solid curves represent TL condition and dotted curves are for TC, both for Δt = 4 hrs.

value normalized by the exact value from (4.1), as

$$\begin{aligned}
&\text{A Energy} \equiv 2c_\alpha \psi_\alpha \psi^*_\alpha \text{ (truncated)}/2c_\alpha \psi_\alpha \psi^*_\alpha \text{(exact)} \\
&\text{B Energy} \equiv 2c_\beta \psi_\beta \psi^*_\beta \text{ (truncated)}/2c_\beta \psi_\beta \psi^*_\beta \text{(exact)} \\
&\text{Z Energy} \equiv \Sigma c_\gamma \psi^2_\gamma \text{ (truncated)}/\Sigma c_\gamma \psi^2_\gamma \text{ (exact)} \qquad (5.1) \\
&\text{T Energy} \equiv \text{Total Energy}
\end{aligned}$$

The time variation of the three energy components presented in (5.1) have been plotted for t = 4.14 hrs. for all seven schemes listed in Table 3, for both the TL and TC conditions based on data from case CA in Fig. 5. We have selected to discuss the TL condition because it is by far in most common usage, and the TC condition for comparison. From a superficial view of Fig. 5, one is immediately impressed with the sizeable errors in some of the schemes, a fact not established from Table 3. These errors have a regular period which is given by the first (largest) frequency, $\nu_1 = .309$. One must conclude, therefore, that the energy components cancel their errors on summation. A further observation is the remarkable improvement in the calculations (reduction in error) by use of the TC condition. Although this condition has been in computational use with higher-order systems for some time (Baer, 1964), its virtues had not been investigated in any detail.

Of the explicit schemes tested, E33 is by far the best with regard to truncation, showing almost no errors during the entire integration period for $\Delta t \simeq 4$ hrs. However, in terms of its utility as a computation scheme, we must refer back to Table 3 which elucidates its limited stability region ($\Delta t \leq 5$ hrs.). Scheme E03 shows errors in excess of 50% in the energy components and describes the anticipated damping with time, but only in the α-wave. The leapfrog scheme also shows large error excursions, but they are cut back dramatically by the TC condition.

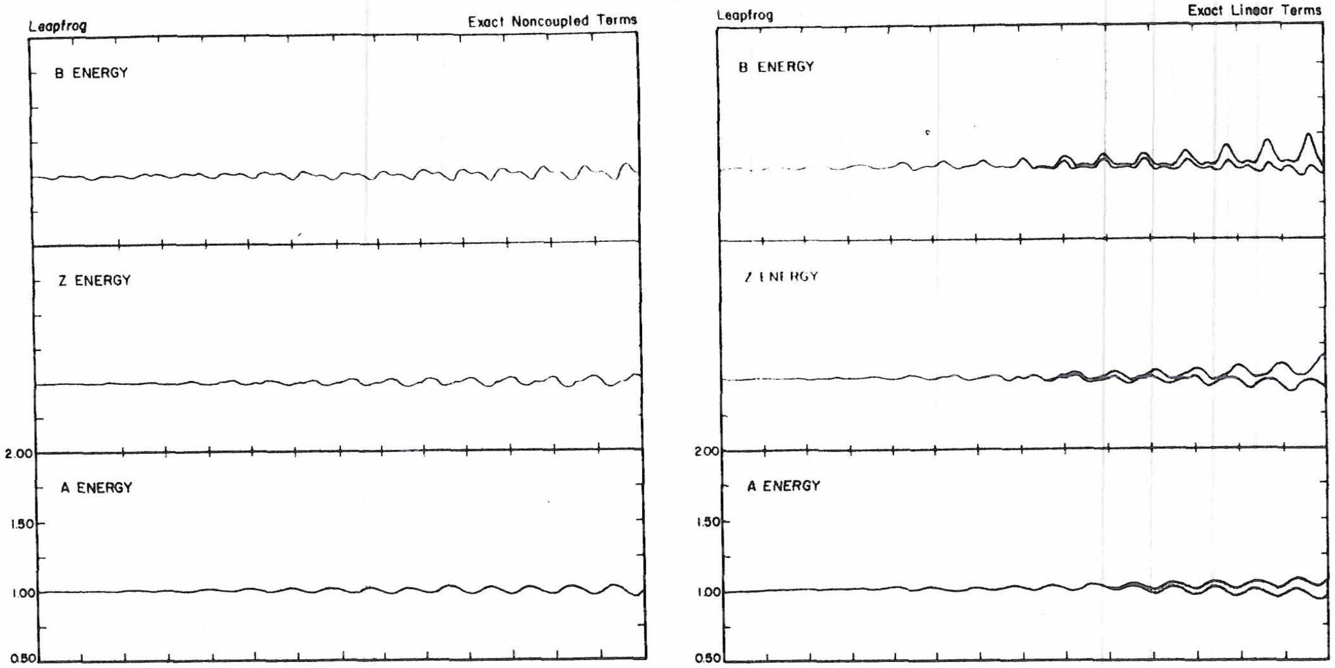

Fig. 6 Component energies calculated using the leapfrog scheme with Δt = 4 hrs., showing the difference between the TC (exact noncoupled) and EL (exact linear) methods.

Although Table 3 indicates no significant errors for the implicit schemes, Fig. 5 clearly does not corroborate this interpretation. Scheme I01 has errors as large as 50% in the components for the TL condition; they are, however, almost completely eliminated when the TC method is applied. An unfortunate and unexpected result of the tests is the poor quality of the IM computation. While the TL results are not available, the TC results suggest that this scheme is inferior to the others described on Fig. 5 (another observation not anticipated from the total energy information of Table 3). Schemes I13 and I35 have been plotted on the same chart since neither has any measurable error in the energy components over the total integration period for $\Delta t \approx 4$ hrs. They are clearly superior schemes, but I13 should be preferred, both because of its better stability condition (Table 3) and its ease of computation.

As indicated above, a striking feature described by Fig. 5 is the improved computation for the TC condition. Because this involves the exact treatment of part of the linear contribution, one might anticipate that the exact treatment of all the linear terms (EL) might further improve the calculated results. That this reasoning is incorrect has already been suggested by the deterioration of the stability criterion for the leapfrog scheme using the EL condition, seen from the total energy in Table 3. Since E11 shows this feature most strongly of all the schemes, we describe on Fig. 6 the different energy components with time for E11, $\Delta t = 4.14$ hrs. using both the TC and EL conditions; the comparison of TL to TC is evident from Fig. 5. None of the schemes show improved computation using the EL method, but most give results comparable to the TC calculation. Most remarkable is the instability which is set up in the E11 scheme using the EL method, a result not anticipated from the linear analysis of Section 4 (Fig. 4), wherein the stability condition for the EL calculation was superior to the TL method. We find here, therefore, a purely nonlinear phenomenon, not predictable by lineari-

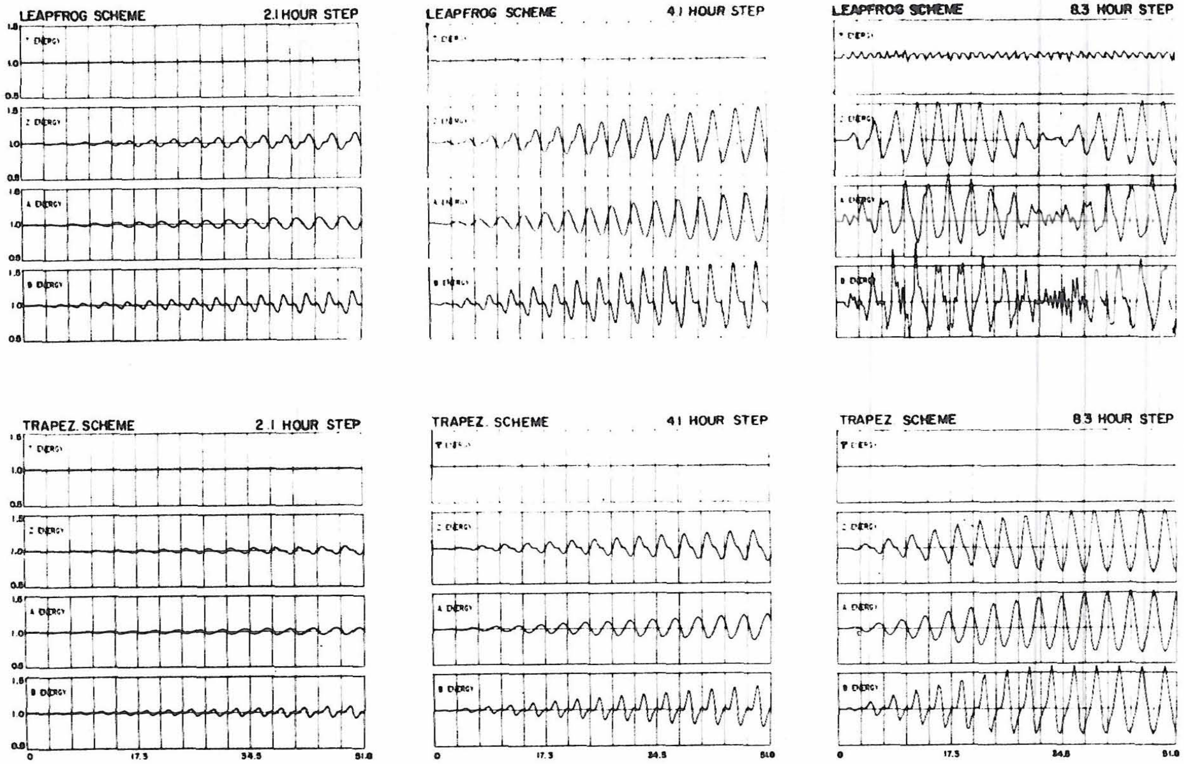

Fig. 7 Component energies calculated using the E11 and I01 schemes with the TL condition, showing the effect of increasing Δt.

zation. However, this observation is not systematic with regard to all the schemes, and does not appear for E03.

The error in the energy components as a function of Δt is described by Fig. 7. Here we show both E11 and I01 using the TL method for the three times, $\Delta t = 2.07, 4.14, 8.28$ hrs. We have chosen E11 and I01 because they are the most frequently used schemes in the explicit and implicit groups, respectively. Nevertheless, all schemes tend to show a similar deterioration of the result with increased Δt, although the higher level implicit schemes (I13 and I35) have extremely small errors for $\Delta t \simeq 8$ hrs. The failure of the total energy to indicate the errors in the components is plainly evident from this figure. An interesting feature of the leapfrog scheme which is apparent for $\Delta t \simeq 8$ is the larger error period, a modulation effect caused by the parasitic mode; this phenomenon has been observed and discussed in the past (see, for example, Baer, 1961). An indication of the component errors seen on Fig. 7 may be available from linear theory through the phase errors. Referring to Fig. 3, when $\Delta t \simeq 2$ the phase errors are almost indetectable, whereas when $\Delta t \simeq 8$ ($\nu\Delta t \simeq 2/3$), both the leapfrog and trapezoidal schemes show the sizeable phase errors. It is interesting to note that for the latter truncation the Milne scheme (I13) has almost no linear phase error and correspondingly no nonlinear computational errors.

Despite the appearance of large errors in the energy components, there exist periodic times at which the computed solutions describe the exact solution with great accuracy. One might thus be led to the conclusion that the numerical integrations will give satisfactory results at selected times (periodic) for all time, to be determined by the highest characteristic mode of oscillation (available from linear theory). Such reasoning, in analogy with the conclusions drawn from the behavior of the total energy only, is based on incomplete information and is unfortunately incorrect. The missing information are the phase angles of the α- and β-waves, both of which are time dependent; their time dependence may be

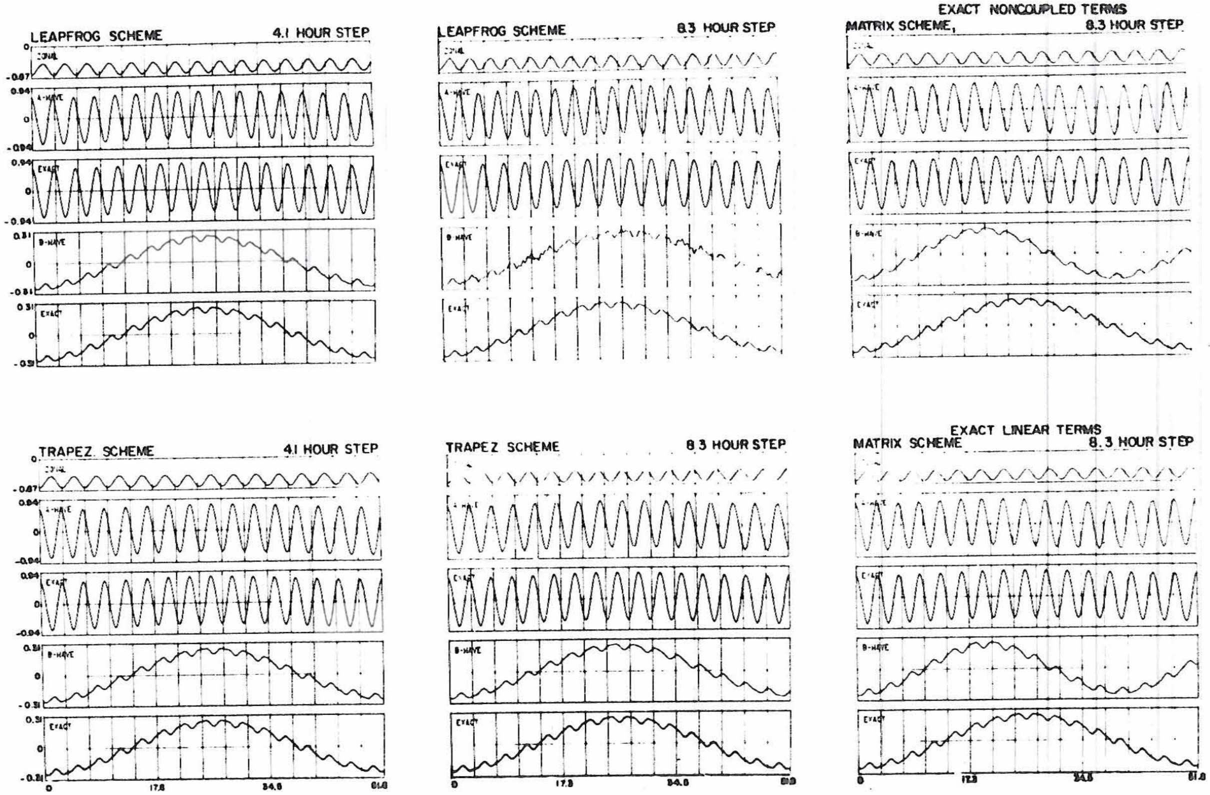

Fig. 8 Computed and exact values of the wave components Re ψ_α, Re ψ_β for the leapfrog, trapezoidal and matrix (IM) schemes, for various Δt and linear conditions, showing development of phase errors.

described by the real part of the stream components ψ_α, ψ_β and we present them as,

$$\begin{aligned} \text{A-wave} &\equiv \text{Re } \psi_\alpha \ (t) \\ \text{B-wave} &\equiv \text{RE } \psi_\beta \ (t) \end{aligned} \qquad (5.2)$$

In Fig. 8 we show the phase properties for the E11 and I01 schemes for the time steps Δt = 4.14, 8.28 hrs., using the TL method. By comparing the computed values of the two wave components as given in (5.2) to the exact values, we see that after 50 days the A-wave is significantly out of phase with the exact value. For Δt ≈ 8 hrs., the phase error is almost 180° in both schemes, whereas the error in the B-wave is negligible. This error grows with time, and the consequent solution therefore becomes less and less reliable. Having now established an almost insurmountable obstacle to these numerical integration schemes (the multi-step implicit schemes I13, I35 do not exhibit discernible phase errors for the time steps utilized), we observe that no apparent phase errors occur if we use the TC or EL condition. The interpretation of this correction must be based on the fact that the uncoupled linear terms include most of the high frequency phase properties and therefore cannot be successfully truncated. Although many of the schemes exhibit the phase characteristics outlined above, the implicit matrix scheme (IM) is nonconformist. With Δt = 8.28 hrs., Fig. 8 shows the phase properties for both the TC and EL methods of the IM scheme and highlights the phase errors, here primarily in the long period of the B-wave.

To lend some credence to generalizations from the above observations based only on case CA, Figs. 9 and 10 describe the behavior in time of the energy components for data from cases CB and CC respectively (numerical values to be found in the Appendix). The results described are based on the TL method and the time increments have been selected on the basis of the characteristic

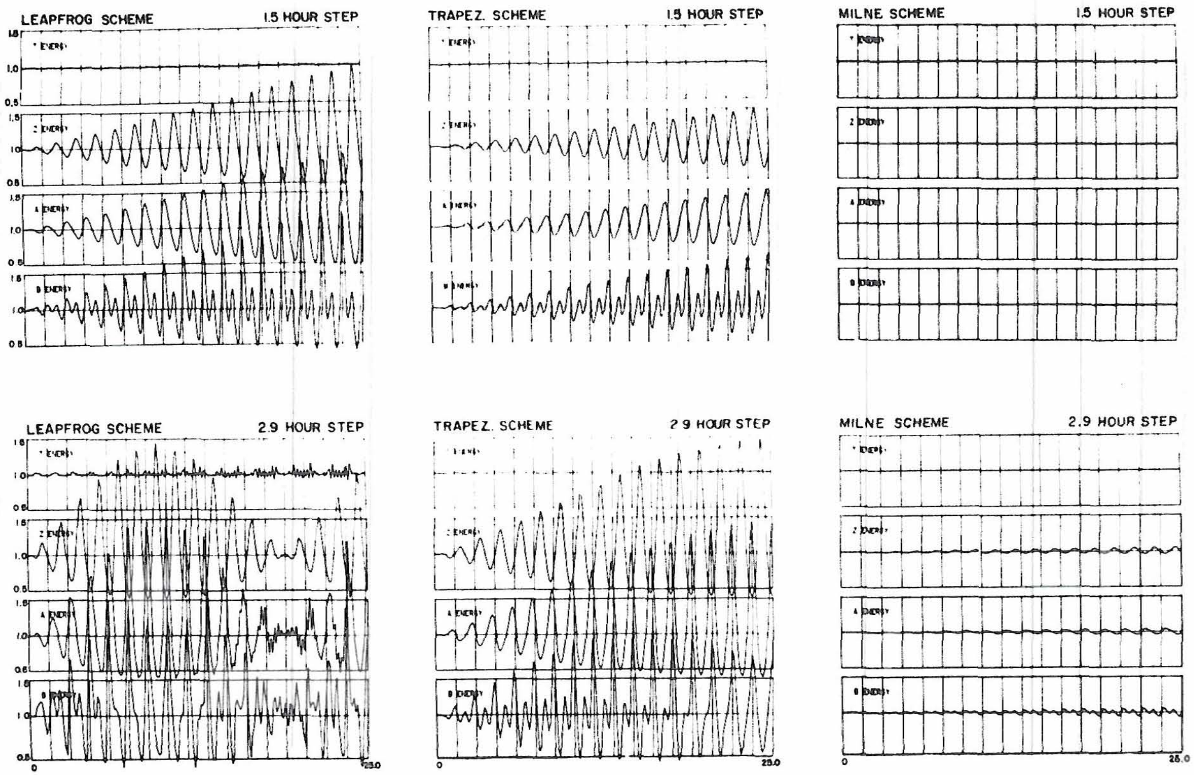

Fig. 9 Component energies calculated using the leapfrog, trapezoidal and Milne (I13) schemes using the TL condition for the two time steps Δt = 1.5, 2.9 hrs. for case CB (see Appendix).

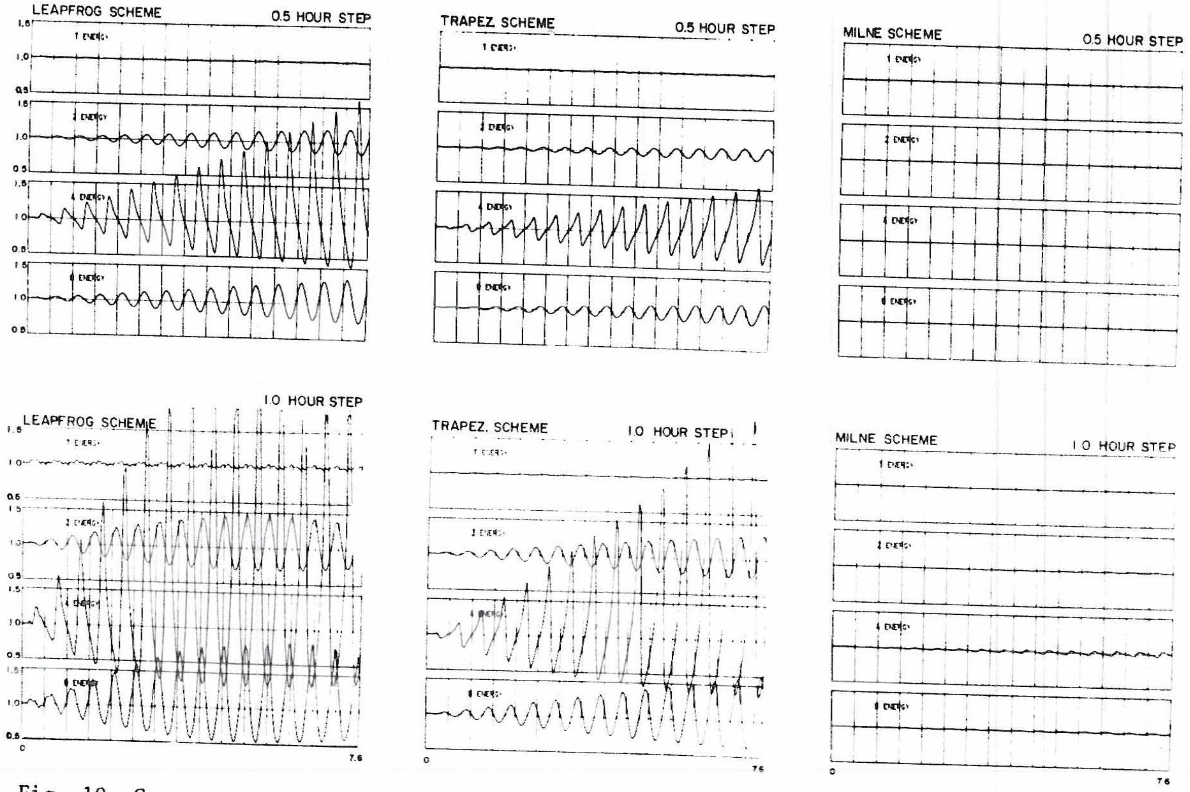

Fig. 10 Component energies calculated using the leapfrog, trapezoidal and Milne (I13) schemes using the TL condition for the two time steps Δt = 0.5, 1.0 hrs. for case CC (see Appendix).

frequencies (see Appendix). All the features which these figures can describe are similar to those discussed for case CA. Errors increase for increased Δt, total energy is well conserved whereas the component errors are large, a modulation period appears in the leapfrog scheme, and the Milne (I13) scheme is extremely accurate. We have found that the other features discussed in detail for case CA shows similar properties for cases CB and CC, and we shall consequently not reproduce these results here; we shall assert, however, based on Figs. 9 and 10, that the computational properties of the different schemes tested and discussed in this section are applicable to a wide variety of initial conditions.

6. Conclusion

The solution of the nonlinear equations which describe atmospheric flow (among others) by numerical means is today a commonplace event. Given a set of initial values, these equations are frequently integrated in time for long periods. It is therefore imperative that an integration scheme be chosen which is not only stable, but also has negligible truncation errors, so that the true solution is not obscured. The development of the "spectral" approach allows this solution to be carried out in time alone, thereby bypassing the space truncation influence. Moreover, the reduction of the spectral equations to low-order form, with their known solutions, enables us to test directly the validity and accuracy of any truncation technique.

Since a wide variety of schemes exist and have been applied, it is desirable to find a general method whereby such schemes may be systematically presented for testing. We have developed such a method based on finite-difference polynomial interpolation, and have shown that many of the more common schemes--both implicit and explicit--are incorporated in our presentation. A number of the lower level schemes have been tested on a simple linear wave equation and those with the most favorable qualities (best stability condition and least truncation) have been selected for testing with a low-order nonlinear spectral system. Included in this group is an implicit method which is not a member of the general set, but is interesting because it does not require iteration.

The low-order system is of particular interest as it involves both linear (coupled and uncoupled) and nonlinear effects. Linear terms may be handled without truncation, and a procedure whereby these terms are removed from the equations may have some impact on the numerical solution of the remaining purely nonlinear equations. An indication that the truncation errors are modified by such elimination is suggested from the solution of the linearized low-order equations, both exactly and with finite-difference methods.

The comparison of the truncated solutions to the exact ones yields some interesting observations. Whereas it has been common to estimate truncation errors of an integration from the behavior of conservative integral properties, our results indicate that only stability can be discussed in this way. The amplitudes of functional variables in our nonlinear system showed wild deviations (errors) at times during the numerical integration, but the conservative property (energy) was well conserved; this was caused by a cancellation of the individual amplitude errors. One must conclude that the conservation of integral constants in a numerical calculation is not sufficient to justify confidence in the results. Furthermore, the satisfactory prediction of amplitudes is also not sufficient; one must also assure the accurate calculation of the phase angles.

Linear theory seems to yield satisfactory information about the computational stability of our nonlinear system, as may be seen from the development of the conservative property, and the linear phase errors (for any scheme) are indication of errors in the amplitudes of the dependent variables. Nonlinear phase errors, which are pronounced for the explicit schemes, may be removed by the exact consideration of the uncoupled linear terms of the nonlinear equations; the latter technique also reduces the amplitude errors significantly. As might have been anticipated, reduction of the truncation interval, Δt, will yield improved solutions.

As a consequence of our calculations, it would be most advisable to select a truncation increment (Δt) substantially less than the critical one determined from linear analysis, if truncation errors are to be minimized. Moreover, to avoid phase errors, any uncoupled linear terms should be removed from the equations by a linear transformation involving the exact solution of such terms. Finally, if computation time is not a serious consideration, an implicit method should be selected in preference to an explicit one. Multi-step methods, although they involve more parasitic solutions, seem to yield superior results. If, for reasons of economy and speed, an explicit scheme is chosen, a technique denoted as "restart", which begins a new calculation periodically from the mean data at the restart time, appears to reduce high frequency amplifying parasitic oscillations, but other truncation properties of this procedure have not been evaluated.

ACKNOWLEDGEMENTS

Computer time for this study was made available by the computer facility of the National Center for Atmospheric Research and by the computer center at Colorado State University. Mrs. Margaret Stollar typed the manuscript.

REFERENCES

Baer, F., 1961: The spectral vorticity equation. Ph.D. Thesis, Dept. of Geophysical Sciences, The University of Chicago, 97 pp.

______, 1964: Integration with the spectral vorticity equation. J. Atmos. Sci., 21, 260-276.

______, 1968: Studies in low-order spectral systems. Atmos. Sci. Paper No. 129, Dept. of Atmos. Sci., Colorado State University, 77 pp.

Henrici, Peter, 1962: Discrete Variable Methods in Ordinary Differential Equations. John Wiley and Sons, New York, 407 pp

Hildebrand, F.B., 1956: Introduction to Numerical Analysis. McGraw Hill Company, New York, 510 pp.

Lilly, Douglas K., 1965: On the computational stability of numerical solutions of time-dependent non-linear geophysical fluid dynamics problems. Mon. Wea. Rev., 93, 11-26.

Lorenz, E.N., 1960: Maximum simplification of the dynamic equation Tellus, 12, 243-254.

Milne, William Edmund, 1949: Numerical Calculus. Princeton University Press, Princeton, New Jersey, 393 pp.

Platzman, G.W., 1960: The spectral form of the vorticity equation. J. Meteor., 17, 635-644.

__________, 1962: The analytical dynamics of the spectral vorticity equation. J. Atmos. Sci., 19, 313-328.

Richtmyer, R.D., 1957: Difference Methods for Initial Value Problems. Interscience Publishers, New York, 238 pp.

Silberman, I., 1954: Planetary waves in the atmosphere. J. Meteo 11, 27-34.

Young, John A., 1968: Comparative properties of some time differencing schemes for linear and nonlinear oscillations. Mon. Wea. Rev., 96, 357-364.

APPENDIX

Constants in Eq. (4.1)		Case CA	Case CB	Case CC
	a_n	-.09788005	-.69019528	-4.8903939
	ρ_α	-.26691770	.21502294	.79837156
	ρ_β	-.03839707	.86136911	.38763314
	$h_{\alpha\beta}$	-.09899117	.27432320	1.1105865
	$h_{\beta\alpha}$	-.07127364	.23888073	.59288305
	$g_{\alpha\alpha}$	-.08220212	.19160456	-1.0310522
	$g_{\alpha\beta}$	-.03050807	.43151307	7.2731965
	$g_{\beta\alpha}$	-.12767626	-.41340017	4.1460202
	$g_{\beta\beta}$	-.01396861	.34083072	3.2491569
Initial Values	ψ_n	-.60497847	.80465985	-.28630513
	ψ_α	.63421748	.42852365	.10114551
	ψ_β	-.25891820	.10713091	.07680246
Solutions of Eq. (4.1)				
Energy variations	zonal	.374→.244	.700→.271	.200→.292
(normalized)	α-wave	.402→.613	.245→.627	.464→.076
	β-wave	.223→.143	.055→.102	.336→.632
Energy Exchange				
Period (days)		3.452	1.470	.508
Wave periods observed in		3.24	1.29	.527
exact solutions (days)		52.	10.3	-
Wave frequencies from		.3097	-.6407	-1.903
linearized equations: $\nu_{1,2}$		.0421	-.1040	.0404
Corresponding wave periods		3.23	1.56	.526
(days)		23.8	9.6	24.75