THE FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

SOME INACCURACIES IN FINITE DIFFERENCING

HYPERBOLIC EQUATIONS

By

RICHARD GROTJAHN

A thesis submitted to the
Department of Meteorology
in partial fulfillment of
the requirements for the
degree of Master of Science.

Approved:

James L. O'Brien
Professor Directing Thesis

S. J. Blumsack

T. J. V.-Branschn

T. N. Krishnamurti

August, 1975

ABSTRACT

The errors introduced by the use of various numerical
schemes for solving mathematical models have generally been
only vaguely determined previously by numerical modelers.
A method for a more quantitative analysis of the inaccuracies
is outlined. The error associated with some simple schemes
is analyzed for several linear hyperbolic systems representa-
tive of typical problems in meteorology and oceanography.
Results of previous studies of phase velocity inaccuracies
are confirmed and form a basis for an extension of the anal-
ysis to group velocities. Significant angular and magnitude
errors are found in the group velocity. Directional errors
of 180° are found for some waves. Since the group velocity
is the propagation speed of the energy, such errors may
have severe consequences in a numerical model. When anal-
ysis was made of complex systems of equations, results found
for simple systems reappeared. Thus, studies of simple
systems may provide useful indications of behavior in more
complex problems where the analysis may have to be limited.
Only the long waves, i.e., those resolved by many grid
points, are represented with any reasonable accuracy.

points, are represented with any reasonable accuracy.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF SYMBOLS

A	subscript "analytical"
C_g	group velocity
E	subscript "explicit"
f	Coriolis parameter (variable)
f_0	Coriolis parameter (constant)
I	subscript "implicit"
j	grid index, x direction
k	wavenumber, x direction
ℓ	wavenumber, y direction
m	grid index, y direction
n	time level index
SI	subscript "semi-implicit"
u	x velocity
U	mean state velocity, x direction
v	y velocity
V	mean state velocity, y direction
β	first meridional derivative of f
Δt	time step
Δx	grid increment, x direction
Δy	grid increment, y direction
η	mode number in y direction
Δy	grid increment, y direction
η	mode number in y direction

λ wavelength

Λ dimensionless CFL parameter

ϕ geopotential

Φ mean state geopotential

ψ streamfunction

ω phase frequency

LIST OF ILLUSTRATIONS

1. INTRODUCTION

In order to study some real physical system, a scientist will often represent that system in a mathematical sense. This may be a collection of numbers which he labels observations or an equation or system of equations which he employs as a model. In proceeding from the real system to the mathematical model, assumptions and approximations were made which introduced a certain amount of error. In many geophysical problems, analytical tools are incapable of solving the equations in the model. Therefore, numerical techniques may have to be invoked which will introduce additional distortion into the representation. If the numerical techniques are carried out on a computer, there is the round off error inherent to the truncated arithmetic. The first and last aspects of the total error are generally well recognized and documented. However, the error created by approximating continuous differential equations with discrete algebraic ones is generally not known in detail by the numerical modeler. This lack of knowledge concerning the error introduced by finite differencing provided the impetus for this work.

impetus for this work.

This report will examine the error in finite differencing, which we call truncation error, in specific detail

1

for several linear systems derived from the "shallow water" primitive equations. It will be demonstrated that the true nature of the truncation error is not revealed in nebulous statements concerning the order of a Taylor series expansion; that descriptions such as "$O(\Delta t^2)$" are grossly deficient and that more precise analysis is imperative.

The method of analysis used here has its roots in the classical stability analyses of Courant et al. (1928) and its extension by von Neumann (e.g., Charney et al., 1950). From the continuous equations the form of the analytic solution is determined. From this analytic form an appropriate discrete form of the solution is derived. These are substituted into the continuous and finite differenced equations, respectively, and the physically correspondent solutions are compared. For wave type solutions (which form the bulk of the solutions derived here) this results in a constraint upon the phase frequencies. Any phase or amplitude disparities between the continous and discrete sets of equations can be deduced by comparing the respective frequencies.

Most previous studies of the error have been more or less empirical in nature (e.g., Orzag, 1971), however, a few quantitative analyses have been made. Those of Kurihara (1965) and Kwizak (1970) are of particular interest here. Kurihara was concerned with testing several implicit and iterative formulations on rather simplified equations. He Kurihara was concerned with testing several implicit and iterative formulations on rather simplified equations. He also considered the computational modes which appear in

three time level schemes. Kwizak applied Kurihara's method
to a semi-implicit formulation and used an approach similar
to the one used here, to analyze some common space and time
differencing formulations. Elvius and Sundström (1973)
wrote the f-plane shallow water primitive equations in matrix
form, determined the eigenvectors and eigenfrequencies, and
compared them with the continuous solutions. This method
is more complete in that boundary conditions could be ex-
plicitly included but, as we shall see, in more complex
systems there is some ambiguity introduced when the eigen-
values can only be found numerically. Their main concern
was the proper choice of boundary conditions for various
efficient semi-implicit staggered grid formulations. The
work of Baer and Simmons (1970) for nonlinear, essentially
spectral, equations involved testing several schemes for
stability and also conservation of integral constraints.
They found that there may be large amplitude errors which
cancel out in the computation of the integral properties.
Thus truncation error could not be necessarily determined
from a scheme's conservative properties. Young (1968)
tested many schemes on the spectral equations and also con-
cluded that truncation error bore little relation to the
order of the scheme chosen. However, he chose not to pursue
the details of the truncation error.

It was decided that the emphasis here would not be
the details of the truncation error.

It was decided that the emphasis here would not be
upon testing many schemes (e.g., Grammeltvedt, 1969).

Neither would various boundary or staggered grid formulations be emphasized - there are simply too many. The problem would become too specialized. Instead, some simple commonly used schemes are applied to several hyperbolic systems and analyzed away from the boundaries. More specific applications are left to the reader. The purpose here is to outline a way to determine quantitatively the error and to present what appear to be some general properties of it.

Some of the results presented here can be found in previous work (Kwizak, 1970). They are included for completeness. What represents new work is the study of more complex systems and comparisons of group velocities. The group velocity is perhaps a more important quantity than the phase velocity, in that energy is propagated at the group velocity.

It will be shown that explicit time differencing tends to speed up oscillations whereas implicit formulations slow them down. Space differencing makes non-dispersive waves dispersive. It also uniformly slows down the phase velocities, with the error increasing for shorter wavelengths. The group velocity for some wavelengths can be in the opposite direction of the correct group velocity. However, long waves, i.e. those resolved by many (> 10) grid points, are generally handled reasonably well. Properties evident in simple systems reappear when the schemes are points, are generally handled reasonably well. Properties evident in simple systems reappear when the schemes are applied to complex systems of equations with similar types

of solutions.   This may be a useful approximation to a scheme's behavior in more complicated systems, since the method is found to have limitations when applied to those more complicated systems.

It is the author's belief that more precise know-ledge of the error will lead to improvements upon it and that more elegant schemes can be formulated which are based on explicit knowledge of the error.   At the least, a more judicious choice of a scheme can be made (e.g. Kreiss and Oliger, 1972).   It may even be possible to employ the error itself in improving the scheme.

2. METHOD

The continuous differential equations are approximated by discrete algebraic (finite differenced) ones. The solution to the continuous equations is given by a function. The solution to the discrete equations is also given by a function. It is these two functions, both analytically derived, which are compared. Requiring an analytical solution to the differential equations means that we are dealing with equations which are never solved numerically! However, one chooses a solvable model which is similar to the more realistic model of interest, and the analysis of the former is applied, with some care, to the latter. In this report all the equations have been linearized so as to study the error in the scheme without the added complication of non-linear effects. One might reasonably expect that one formulation judged superior to another via linear analysis would perform better than the other in a similar problem containing some degree of nonlinearity.

Many of the solutions derived here are propagating waves. The continuous solutions for a one dimensional problem have the form

waves. The continuous solutions for a one dimensional problem have the form

$$S_A = e^{i(kx - \omega t)} \tag{1}$$

6

where k is wavenumber in the x direction and ω is the phase frequency. The appropriate general solution for the differenced system is easily found

$$S_j = e^{i(kj\Delta x - \omega n\Delta t)} \tag{2}$$

where j is an index in the x direction, Δx is the spatial grid interval, n is a time step index and Δt is the time step. The form (2) is appropriate because it will have the same functional form when substituted into the discrete equations as (1) has when substituted into the differential set. To see this more clearly, consider the first derivative of each solution:

$$\frac{\partial}{\partial x}(S_A) = ikS_A \tag{3}$$

$$\frac{1}{2\Delta x}(S_{j+1} - S_{j-1}) = i\frac{\sin k\Delta x}{\Delta x} S_j \tag{4}$$

where subscripts in (4) refer to a grid point. Just as there is no x dependence in the coefficient multiplying S_A on the right hand side of (3), there is no x dependence in the coefficient multiplying S_j in (4). That is, (3) and (4) have the same functional form.

One cannot always form the discrete analog this easily. For example, one of the problems treated later has a solution of the form

$$S_A = e^{x^2}$$

then

$$S_A = e^{x^2}$$

then

$$\frac{\partial}{\partial x}(S_A) = 2xe^{x^2} \tag{5}$$

Note that finite differencing a solution

$$S_j = e^{(j\Delta x)^2} \tag{6}$$

leads to

$$\frac{1}{2\Delta x}(S_{j+1} - S_{j-1}) = \frac{1}{\Delta x}\sinh(2j\Delta x^2)\, e^{\Delta x^2} S_j$$

This does not have the same functional form as (5). To
be strictly correct one must resort to an eigenvector
analysis in such a case. In the particular problem con-
sidered later, it was allowable to use the form (6) despite
being inexact, because (6) was close to the correct solution.

Substitution of (1) and (2) into their respective
systems yields dispersion relations from which the phase
frequencies of the two systems are derived as a function
of relevant parameters. The phase error is thereby deduced
as a function of those parameters.

For solutions whose temporal dependence is
exp (iωnΔt), the function for ω will involve either the
arcsine or arctangent of some quantity. This presents an
infinite number of frequencies, each shifted by some incre-
ment of π. Only those values of the inverse transcendentals
between -π/2 and π/2 will be considered. This is justified
since as the time and space increments approach zero, the
finite differenced solution for ω will approach the true
solution, not one shifted by some increment of π.
finite differenced solution for ω will approach the true
solution, not one shifted by some increment of π.

The group velocities are obtained by differentiating
the phase frequencies with respect to wavenumber. This

will be used as an indication of how the energy contained in different wavenumbers is propagating. An excellent review of the physics of the group velocity is provided by Whitham (1974).

It should be noted that the imaginary part of the phase frequency will imply growth (or decay). If the analytical frequency is real and the finite differenced frequency complex, then there will be amplitude errors as well, in the numerical scheme chosen.

In this study various combinations of the "shallow water" primitive equations will be used. Some simple systems with only one distinct solution will be considered first. Many of these results will reappear when systems with multiple roots and non-constant coefficients are examined.

3. SIMPLE SYSTEMS

There are two basic approaches to time differencing, either explicit or implicit formulations. We will treat both. In the more complex systems they will be mixed.

As an example of an ordinary differential equation in time, consider inertial oscillations. If we define

$$w \equiv u + iv$$

then our model is

$$w_t = - ifw \tag{7}$$

The solution to (7) is

$$w = e^{i\omega t}$$

where

$$\omega_A = - f \tag{8}$$

An explicit formulation of (7) is

$$w^{n+1} = w^{n-1} - 2if \, \Delta t \, w^n \tag{9}$$

where superscripts refer to the time level. Substitution into (9) of a solution of the form

$$w = D \, e^{i\omega n\Delta t}$$

yields

yields

$$\omega_E = - \frac{1}{\Delta t} \arcsin (f\Delta t) \tag{10}$$

10

For an implicit formulation

$$w^{n+1} = w^n - \frac{i\Delta t}{2} (w^n + w^{n+1})$$

one obtains a phase frequency

$$\omega_I = - \frac{2}{\Delta t} \arctan \left(\frac{f\Delta t}{2}\right) \tag{11}$$

The ratios of the physical solutions in (10) and (11), normalized by (8), are plotted in Fig. 1 as a function of the nondimensional parameter $f\Delta t$. The explicit formulation is tending to overestimate the oscillation whereas the implicit treatment is underestimating it. The explicit formulation is unstable for $f\Delta t > 1$. As the time step increases the two schemes diverge from the correct value.

It is interesting to note that in such a linear problem one could achieve quite high accuracy, while retaining large time increments, through a judicious mixture of explicit and implicit integrations. That is, one could nullify the overestimation of an explicit formulation by following with an implicit scheme. For a nonlinear model, such as the spectral equations, this technique may be useful, though two problems remain. The time steps would probably still need to be relatively small so as not to miss too much of the nonlinear effects during any given time step. Also a useful implicit scheme which avoids iteration would be preferred.

Before proceeding to problems with mixed time and space derivatives, consider space differentiation singly.
Before proceeding to problems with mixed time and space derivatives, consider space differentiation singly.
For a wave type solution,
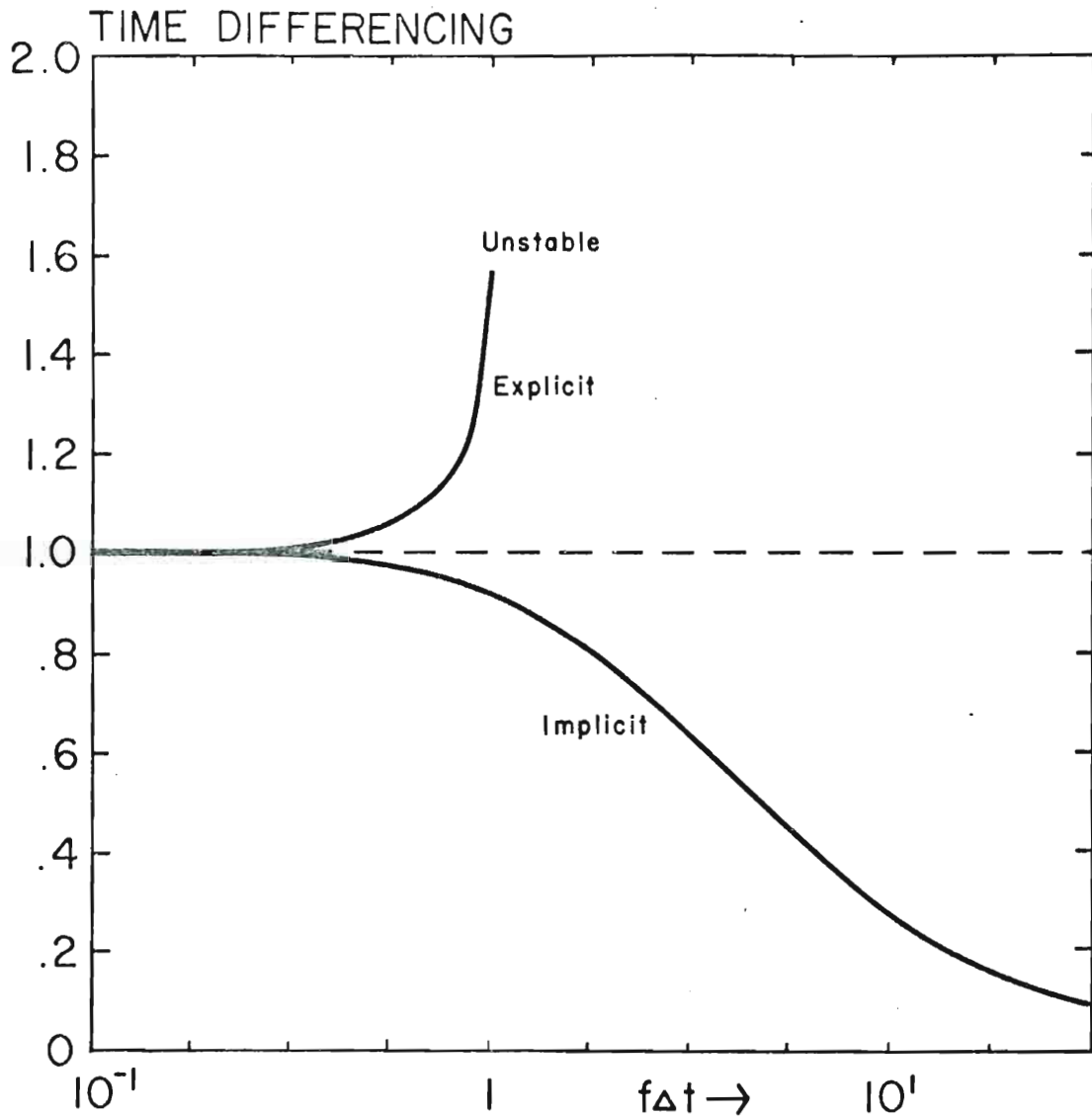
Fig. 1. Frequencies for the explicit and implicit schemes normalized by the correct frequency as a function of the nondimensional parameter fΔt. The explicit scheme overestimates the oscillation, whereas the implicit scheme underestimates it.

$$\frac{\partial}{\partial x} e^{(ikx)} = ik \, e^{ikx}$$

and for a centered-in-space finite difference approximation,

$$\frac{1}{2\Delta x} \left(e^{ik(j+1)\Delta x} - e^{ik(j-1)\Delta x} \right) = \frac{i \, \sin k\Delta x}{\Delta x} e^{ikj\Delta x} \qquad (12)$$

In a sense there is a "computational wavenumber", $\sin k\Delta x/\Delta x$, corresponding to the true wavenumber, k. This implies that the representation is fairly good for long waves (i.e., those resolved by many grid points) but quite poor for shorter waves, particularly those less than $4\Delta x$. (Note: a $4\Delta x$ wavelength corresponds to $k\Delta x = \pi/2$). A fourth order scheme is

$$\frac{1}{\Delta x} \left[\frac{2}{3} \left(e^{ik(j+1)\Delta x} - e^{ik(j-1)\Delta x} \right) - \frac{1}{12} \left(e^{ik(j+2)\Delta x} - e^{ik(j-2)\Delta x} \right) \right]$$

$$= i \left(\frac{8 \, \sin k\Delta x - \sin 2k\Delta x}{6\Delta x} \right) e^{ikj\Delta x} \qquad (13)$$

The relations (12) and (13), normalized by the correct value k are both plotted in Fig. 2. The five point scheme is clearly superior to the centered-in-space formulation for all wavenumbers.

Consider the simple advection equation

$$\phi_t + U\phi_x + V\phi_y = 0 \qquad (14)$$

where U and V are constant basic state velocities in the x and y directions, respectively. For a solution of the form

$$\phi = e^{i(kx+\ell y-\omega_A t)} \qquad (15)$$

$$\phi = e^{i(kx+\ell y-\omega_A t)} \qquad (15)$$

one obtains the constraint

$$\omega_A = Uk + V\ell \qquad (16)$$
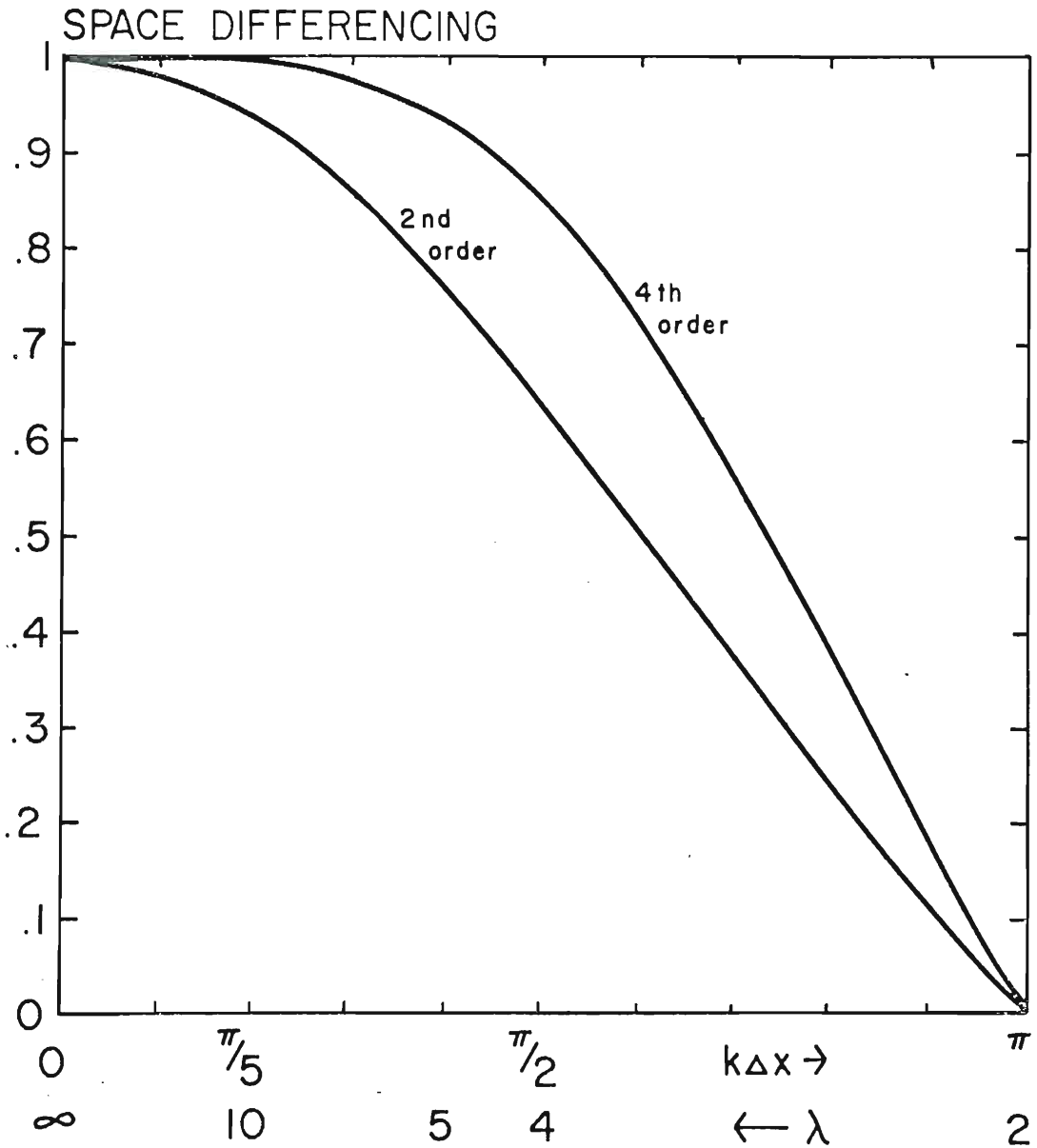
Fig. 2. Comparison of the two finite difference approximations to a first derivative in space, normalized by the correct value, as a function of wavenumber. λ is the corresponding wavelength in terms of grid intervals (i.e., where kΔx = π/2 wavelength is 4Δx).

lized by the correct value, as a function of wavenumber. λ is the corresponding wavelength in terms of grid intervals (i.e., where kΔx = π/2 wavelength is 4Δx).

An explicit formulation of (14) is

$$\phi^{n+1} = \phi^{n-1} - U \frac{\Delta t}{\Delta x}\left(\phi^n_{j+1} - \phi^n_{j-1}\right) - V \frac{\Delta t}{\Delta y}\left(\phi^n_{m+1} - \phi^n_{m-1}\right)$$

which, when the solution

$$\phi^n = e^{i(kj\Delta x - \ell m\Delta y - \omega_E n\Delta t)} \tag{17}$$

is inserted, reveals

$$\omega_E = \frac{1}{\Delta t} \arcsin\left[U \frac{\Delta t}{\Delta x} \sin k\Delta x + V \frac{\Delta t}{\Delta y} \sin \ell\Delta y\right] \tag{18}$$

Substitution of (17) into an implicit formulation

$$\phi^{n+1} = \phi^n - U \frac{\Delta t}{4\Delta x}\left(\phi^n_{j+1} + \phi^{n+1}_{j+1} - \phi^n_{j-1} - \phi^{n+1}_{j-1}\right)$$

$$- V \frac{\Delta t}{4\Delta y}\left(\phi^n_{m+1} + \phi^{n+1}_{m+1} - \phi^n_{m-1} - \phi^{n+1}_{m-1}\right)$$

yields

$$\omega_I = \frac{2}{\Delta t} \arctan\left[U \frac{\Delta t}{2\Delta x} \sin k\Delta x + V \frac{\Delta t}{2\Delta y} \sin \ell\Delta y\right] \tag{19}$$

Appendix 1 outlines these derivations in detail.

The explicit phase frequency (18) is compared to the continuous one, (16), Fig. 3, for various values of Λ. Here Λ is the nondimensional ratio

$$\Lambda = U \frac{\Delta t}{\Delta x}$$

$\Lambda = 1$ is the Courant-Friedrichs-Lewy (CFL) condition for linear stability. There are two aspects of the curves worth noting. The characteristic shape is similar to that in Fig. linear stability. There are two aspects of the curves worth noting. The characteristic shape is similar to that in Fig. 2. The waves have been made dispersive (the phase velocity depends on the wavenumber) by the finite differencing process,

with the representation poorer for shorter waves. At the
CFL condition the phase frequency is exact for waves longer
than $4\Delta x$. This is due to the characteristics of the contin-
uous system intersecting the grid points as indicated by
the $\Lambda = 1$ curve in Fig. 4. The true value at the point A''
is that at the point A'. If one writes the centered-in-
space difference as

$$A'' = B - \Lambda (B' - A')$$

then for $\Lambda = 1$, B' is the value at B and the difference
reduces to A'' = A' which is the true solution. The fact
that waves smaller than $4\Delta x$ appear to be treated poorly
is due in part to our use of only the first harmonic of
the arcsine (values between $-\pi/2$ and $\pi/2$). Since at
$\Lambda = 1$ we are computing

$$\text{arcsine (sin } k\Delta x),$$

we are therefore not retrieving $k\Delta x$ for $k\Delta x > \pi/2$. This
slight ambiguity concerning the arcsine in deriving the
phase frequencies was apparently not recognized by previous
investigators. However, in actual computer tests, the
choice of arcsine values between $-\pi/2$ and $\pi/2$ appears to
be proper. Note that there is no such ambiguity in the com-
putation of the group velocities.

For $\Lambda = 1$, the implicit formulation (see Fig. 3) is
very close to the explicit scheme in terms of accuracy (but

very close to the explicit scheme in terms of accuracy (but
just slightly worse). For $\Lambda > 1$, the explicit scheme is
unstable, whereas the implicit scheme maintains stability

(a)

(b)
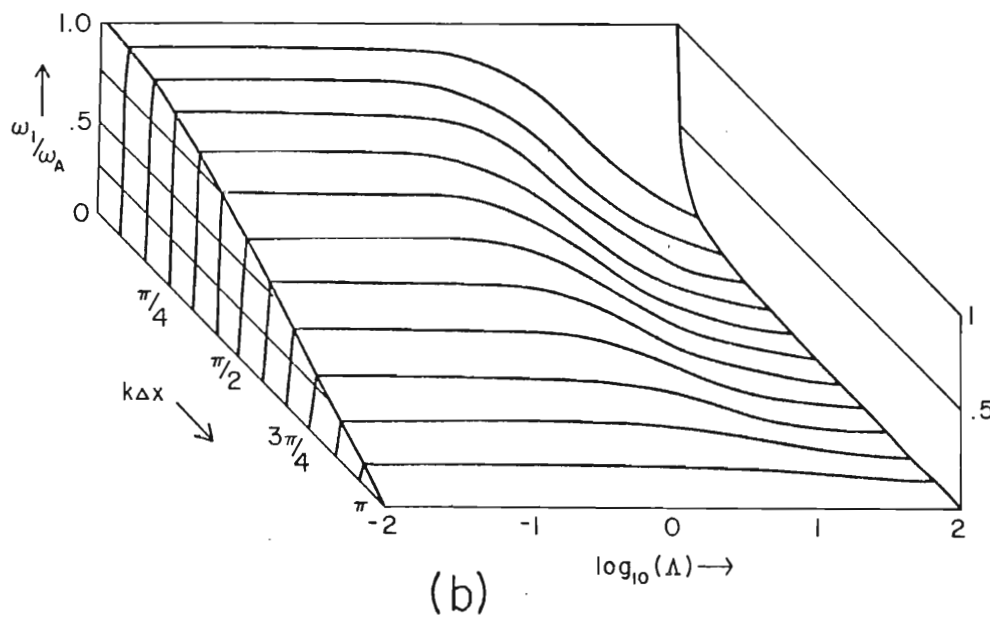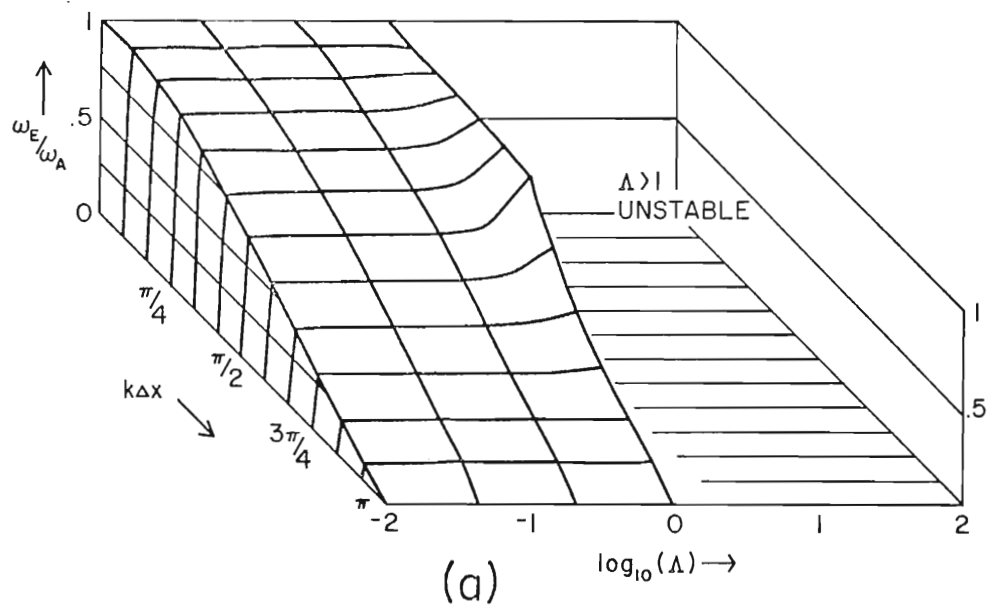
Fig. 3. Accuracy of the (a) explicit and (b) implicit

Fig. 3. Accuracy of the (a) explicit and (b) implicit phase frequencies for the advection equation as a function of the CFL parameter Λ. The limit for linear stability is $\Lambda = 1$. For $\Lambda < 1$ the two schemes are nearly the same; for $\Lambda > 1$ the implicit scheme maintains stability by slowing down the waves.
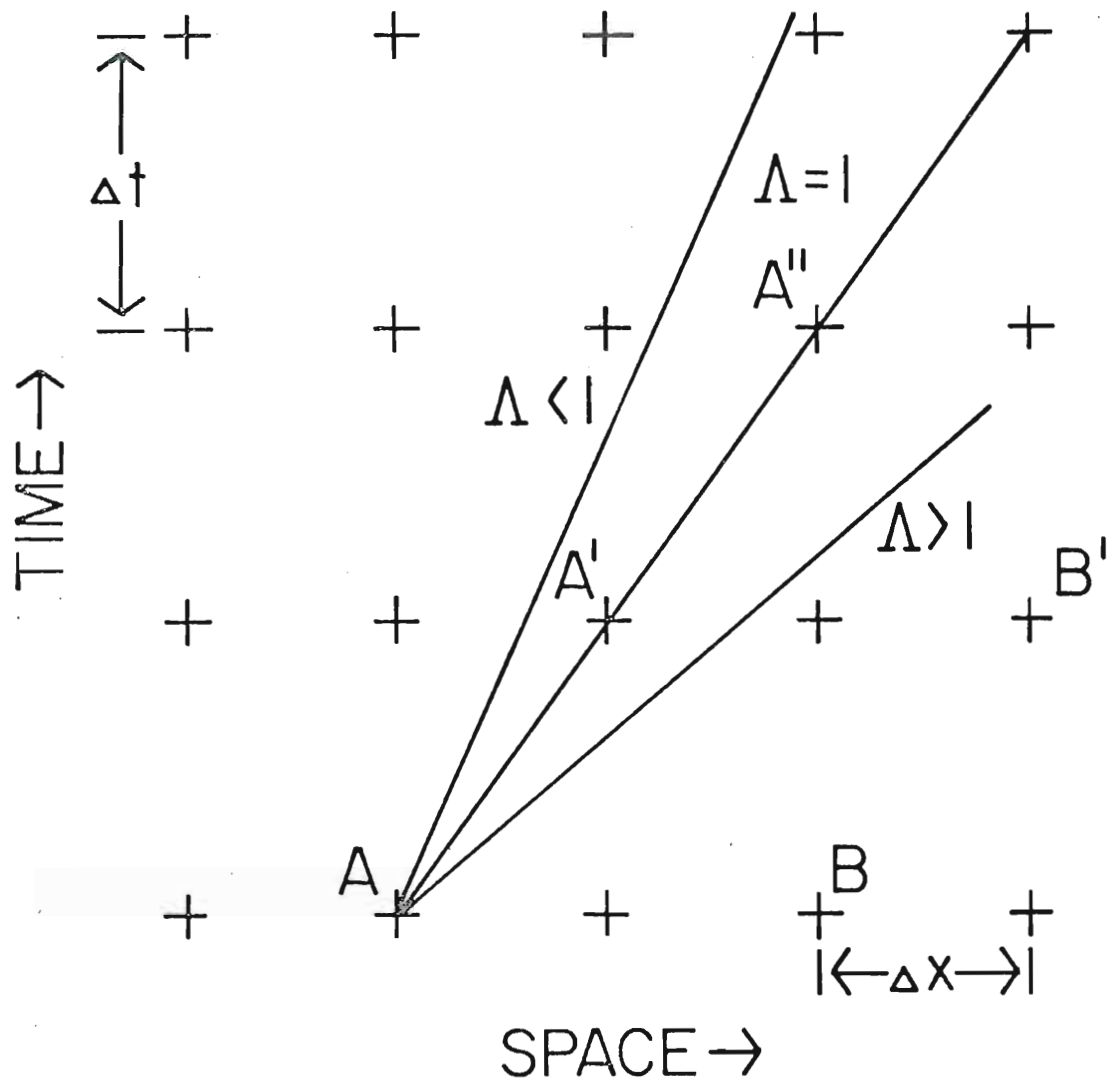
Fig. 4. Three characteristic curves originating from the point A, superimposed on a grid point lattice. The $\Lambda < 1$ curve is stable, the $\Lambda > 1$ curve is linearly unstable for an explicit formulation. For the explicit formulation used, the characteristic $\Lambda = 1$ is predicted exactly.

by slowing down the phase speeds. The rate of slowing down is not uniform. Figure 5 compares the velocities of the waves at various values of Λ to a quantity which will be termed the "CFL velocity"

$$C_{FL} = \frac{\pi \Delta x}{2 \Delta t \ k \Delta x}$$

If the velocity of any wavenumber exceeds this quantity, then the CFL condition is violated by that wave. We see that at $\Lambda = 1$ all the phase speeds are well below the CFL velocity, and that this velocity is only <u>asymptotically</u> approached for large Λ.

The group velocities for the one-dimensional case $\ell = 0$ are given by

$$C_{gA} = \frac{d\omega_A}{dk} = U \tag{20}$$

for the analytical, for the explicit

$$C_{gE} = \frac{d\omega_E}{dk} = U \frac{\cos k\Delta x}{(1 - (U \frac{\Delta t}{\Delta x} \sin k\Delta x)^2)^{1/2}} \tag{21}$$

and for the implicit,

$$C_{gI} = \frac{d\omega_I}{dk} = U \frac{\cos k\Delta x}{1 + (U \frac{\Delta t}{\Delta x} \sin k\Delta x)^2} \tag{22}$$

The relations (21) and (22), normalized by (20) are plotted in Fig. 6 for various values of Λ. The dispersive property

in Fig. 6 for various values of Λ. The dispersive property introduced by the finite differencing has important effects upon the transport of energy. The group velocity for the
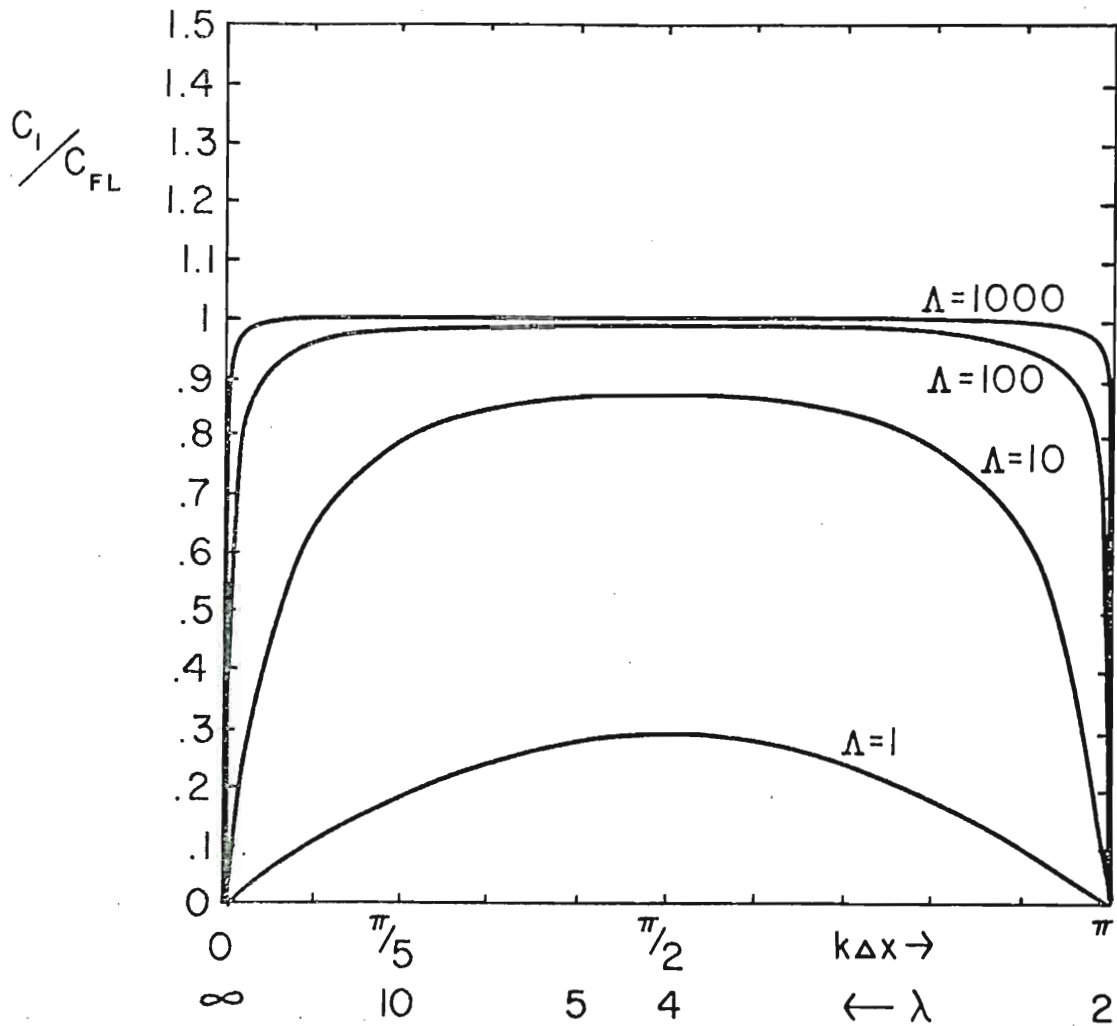
Fig. 5. Comparison of the phase velocity for the implicit formulation to the largest possible phase velocity for stability as a function of wavenumber for various values of Λ. The implicit scheme slows the waves down more than would be required to maintain stability, the limit is only asymptotically approached.
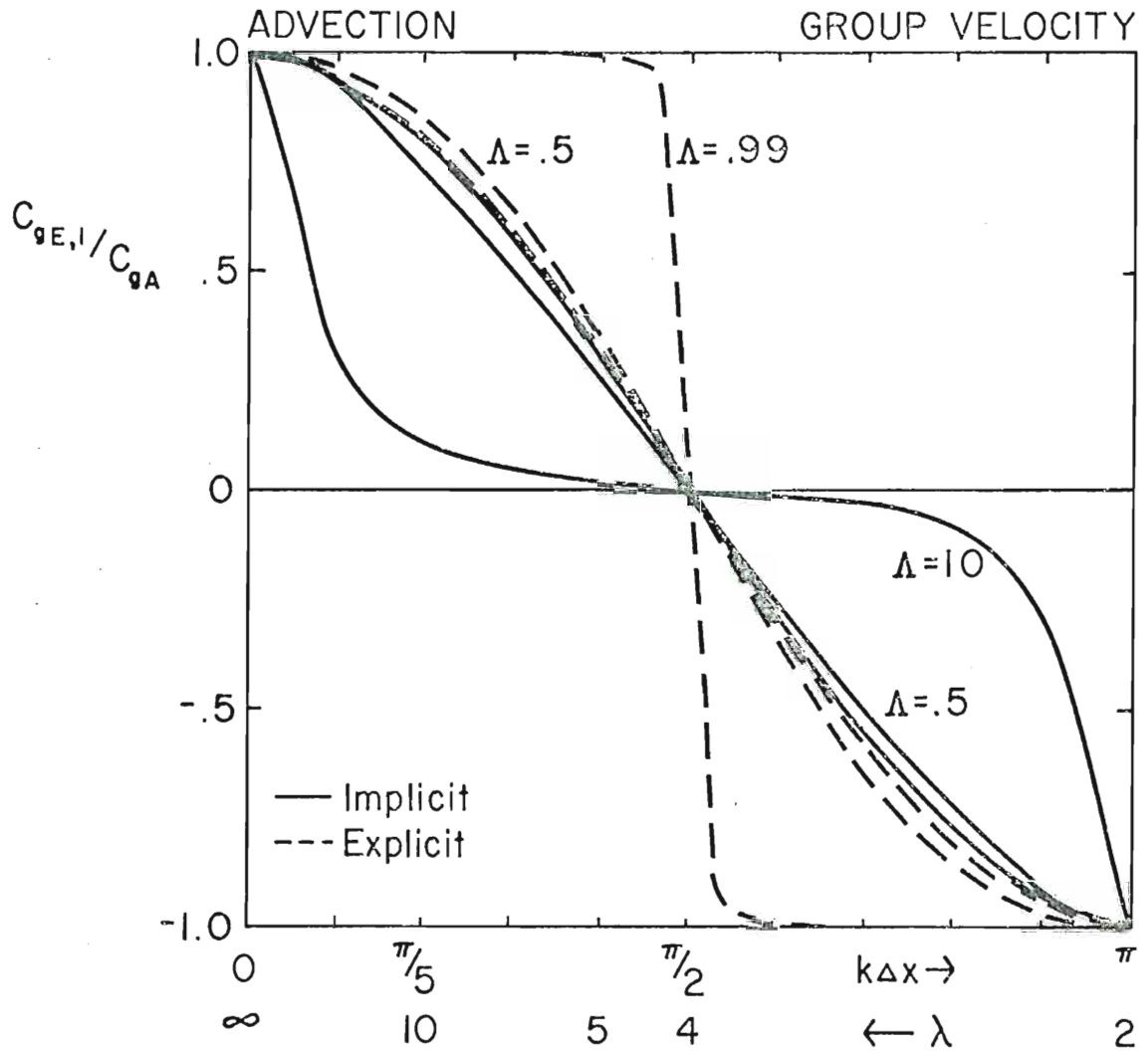
Fig. 6. Group velocities for the one-dimensional advection equation. Velocities of the two numerical schemes are normalized by the analytical value. The explicit scheme is always more accurate than the implicit for waves longer than 4Δx. For waves shorter than 4Δx, the energy in those wavelengths is propagated by both schemes in the opposite direction to the correct group velocity.

long waves (>> 4Δx) is approximated reasonably well. The dramatic effect is that energy in waves shorter than 4Δx is propagated in the opposite direction to the correct velocity. Waves 4Δx long have zero group velocity. The implicit scheme, for large (typical) values of Λ does a poor job even for the long waves. It is also apparent that the explicit formulation, for low wavenumbers, is always more accurate than the implicit scheme. Finally, Fig. 7 compares the explicit version of the second order scheme to the fourth order scheme. Where the group velocity for the latter is given by

$$U \left[\frac{\frac{4}{3} \cos k\Delta x - \frac{1}{3} \cos 2k\Delta x}{(1 - (\frac{4}{3} U \frac{\Delta t}{\Delta x} \sin k\Delta x - U \frac{\Delta t}{6\Delta x} \sin 2k\Delta x)^2)^{1/2}} \right]$$

The fourth order scheme actually does worse than the second order for the shortest waves. However, the long waves are handled quite well, though their phase speed is overestimated.

Gravity waves can be modelled by the following set of equations

$$u_t = -\phi_x \qquad v_t = -\phi_y \qquad \phi_t + \Phi(u_x + v_y) = 0$$

where Φ is a mean state geopotential. A dispersion relation can be obtained using solutions for u, v and ϕ of the form (15)

$$\omega_A = \pm (\Phi (k^2 + \ell^2))^{1/2} \tag{23}$$

An explicit formulation

$$\omega_A = \pm (\Phi (k^2 + \ell^2)) \tag{23}$$

An explicit formulation

$$u^{n+1} = u^{n-1} - \frac{\Delta t}{\Delta x} (\phi^n_{j+1} - \phi^n_{j-1})$$
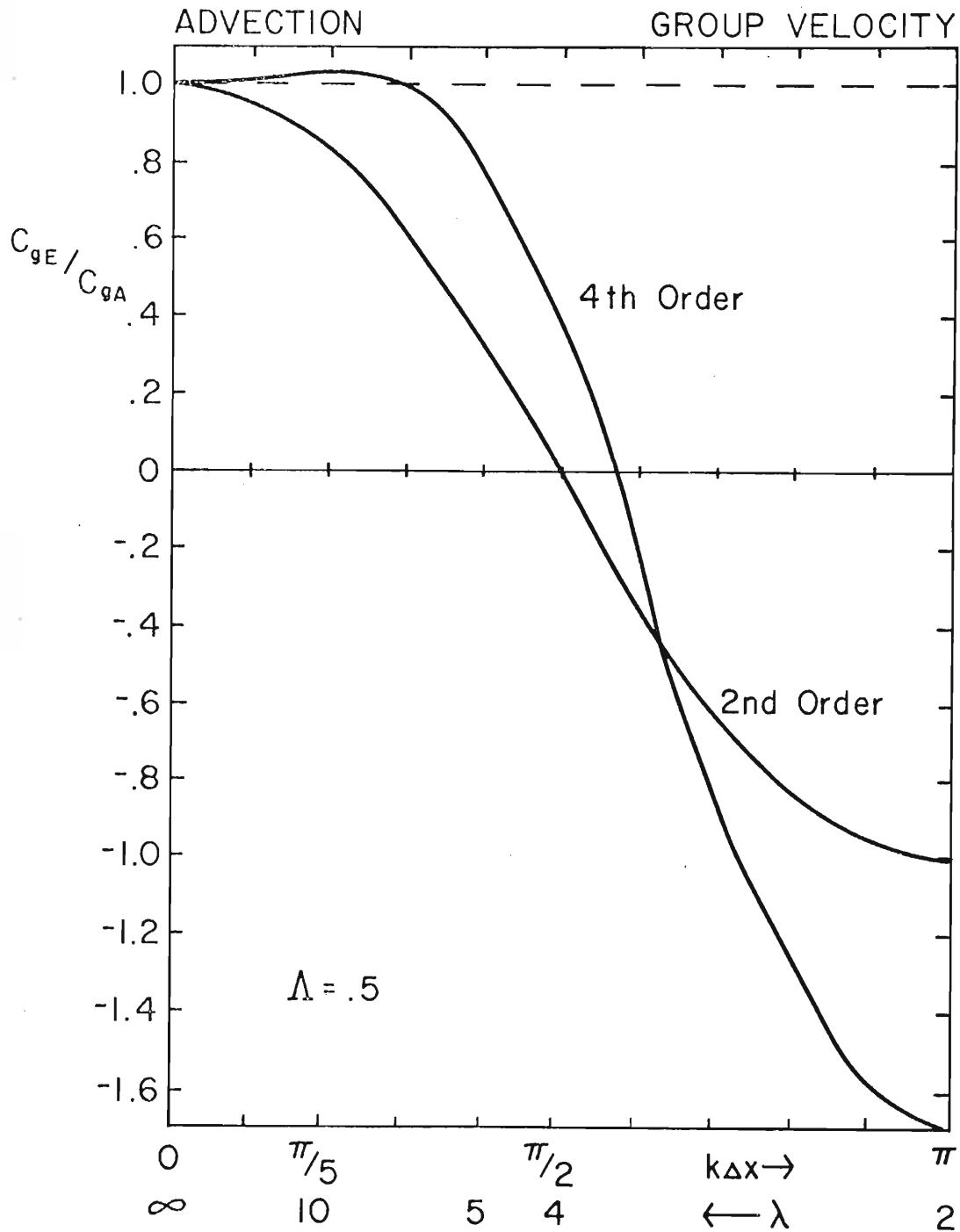
ADVECTION   GROUP VELOCITY

Fig. 7. Comparison of the group velocities between
a so-called "fourth order" scheme and the "second order",
centered-in-space scheme as a function of wavenumber.

Fig. 7. Comparison of the group velocities between
a so-called "fourth order" scheme and the "second order",
centered-in-space scheme as a function of wavenumber.
Long waves are treated quite well by the higher order scheme,
however its treatment of the shortest waves is inferior to
the second order scheme.

$$v^{n+1} = v^{n-1} - \frac{\Delta t}{\Delta y} \left(\phi_{m+1}^n - \phi_{m-1}^n \right)$$

$$\phi^{n+1} = \phi^{n-1} - \Phi \left[\frac{\Delta t}{\Delta x} (u_{j+1}^n - u_{j-1}^n) + \frac{\Delta t}{\Delta y} (v_{m+1}^n - v_{m-1}^n) \right]$$

after substitution of solutions of the form (17) implies that

$$\omega_E = \frac{1}{\Delta t} \arcsin \left\{ \pm \left[\Phi \, \Delta t^2 \left(\frac{\sin^2 k\Delta x}{\Delta x^2} + \frac{\sin^2 \ell\Delta y}{\Delta y^2} \right) \right]^{1/2} \right\} \tag{24}$$

An implicit formulation

$$u^{n+1} = u^n - \frac{\Delta t}{4\Delta x} \left(\phi_{j+1}^n + \phi_{j+1}^{n+1} - \phi_{j-1}^n - \phi_{j-1}^{n+1} \right)$$

$$v^{n+1} = v^n - \frac{\Delta t}{4\Delta y} \left(\phi_{n+1}^n + \phi_{m+1}^{n+1} - \phi_{m+1}^n - \phi_{m-1}^{n+1} \right)$$

$$\phi^{n+1} = \phi^n - \Phi \left[\frac{\Delta t}{4\Delta x} (u_{j+1}^n + u_{j+1}^{n+1} - u_{j-1}^n - u_{j-1}^{n+1}) \right.$$
$$\left. + \frac{\Delta t}{4\Delta y} (v_{m+1}^n + v_{m+1}^{n+1} - v_{m-1}^n - v_{m-1}^{n+1}) \right]$$

yields in a similar fashion

$$\omega_I = \frac{2}{\Delta t} \arctan \left\{ \pm \left[\Phi \, \frac{\Delta t^2}{4} \left(\frac{\sin^2 k\Delta x}{\Delta x^2} + \frac{\sin^2 \ell\Delta y}{\Delta y^2} \right) \right]^{1/2} \right\} \tag{25}$$

Because of the symmetry of the relations (23)-(25) we need consider only the one-dimensional case $\ell = 0$. The solutions behave quite similarly to the case of the one-dimensional advection equation; both are shown in Fig. 8.

The components of the vector group velocities are symmetric so we will consider only one of them. Here,
The components of the vector group velocities are symmetric so we will consider only one of them. Here, however, ℓ will be treated as a parameter and allowed to vary instead of setting it to zero. The true group
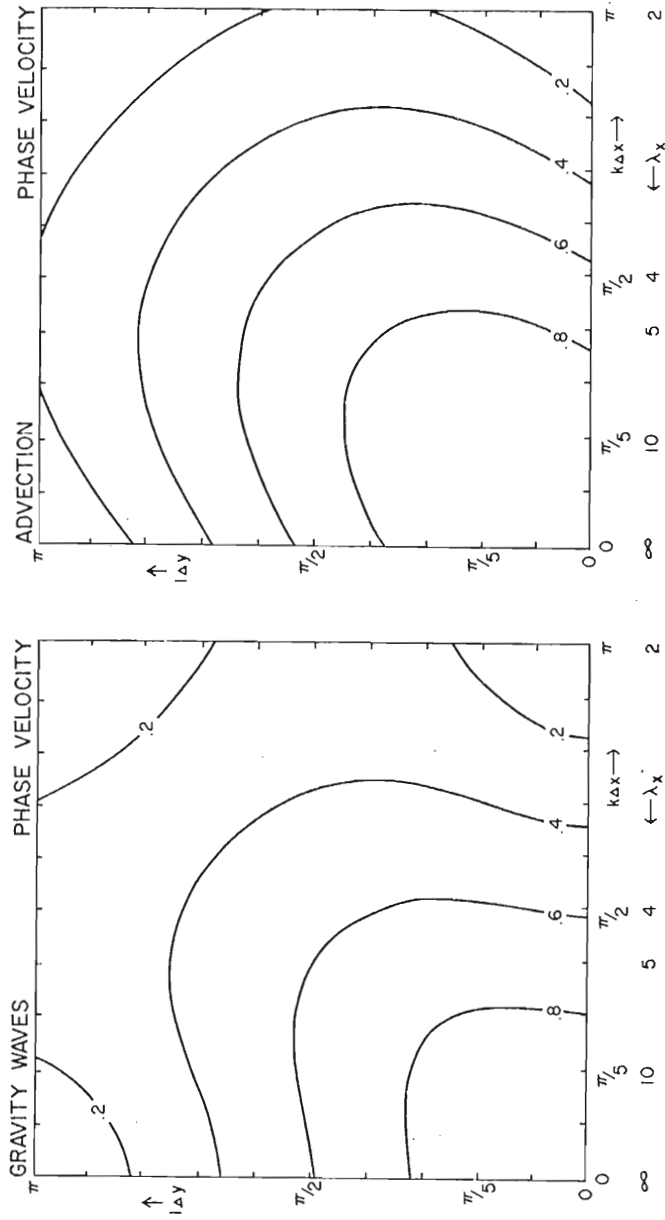
Fig. 8. Contour plots of the phase frequency for the two-dimensional (a) graa) gravity waves system and (b) advection equation. The frequencies are normaliormalized by the analytical solutions for various values of the x and y wavenumbersumbers, k and ℓ.

velocity is

$$C_{gA} = \frac{\pm \ \Phi^{1/2} \ k}{(k^2 + \ell^2)^{1/2}} \tag{26}$$

The explicit version is

$$C_{gE} = \frac{\pm \ \Phi^{1/2} \ \sin k\Delta x \ \cos k\Delta x}{\Delta x \ (1 - \Delta t^2 \ A)^{1/2} \ A^{1/2}} \tag{27}$$

where

$$A = \frac{\sin^2 k\Delta x}{\Delta x^2} + \frac{\sin^2 \ell\Delta y}{\Delta y^2}$$

and the implicit version is

$$C_{gI} = \pm \ \frac{\Phi^{1/2} \ \sin k\Delta x \ \cos k\Delta x}{\Delta x \left[1 + \Phi \ \frac{\Delta t^2}{4} \ A\right] A^{1/2}} \tag{28}$$

Contour plots of these relations appear in Figs. 9 and 10. Figure 9 compares the explicit formulation (27) to the analytical, (26), for a $\Lambda = \Phi \ \Delta t^2/\Delta x^2$ ratio near the maximum allowable for stability ($\Delta x = \Delta y$ for simplicity here). A graph of the implicit formulation for the same value of Λ is very similar to Fig. 9, and is not reproduced, instead a plot is presented for a value of Λ considerably in excess of that used in Fig. 9.

The solutions to the governing equations are plane waves and relations (26)-(28) are merely the component of the group velocity in the x direction. There will be errors in the direction of the vector group velocity as well as in the group velocity in the x direction. There will be errors in the direction of the vector group velocity as well as in the magnitude. In some cases these errors can be quite large, particularly for waves which are traveling nearly
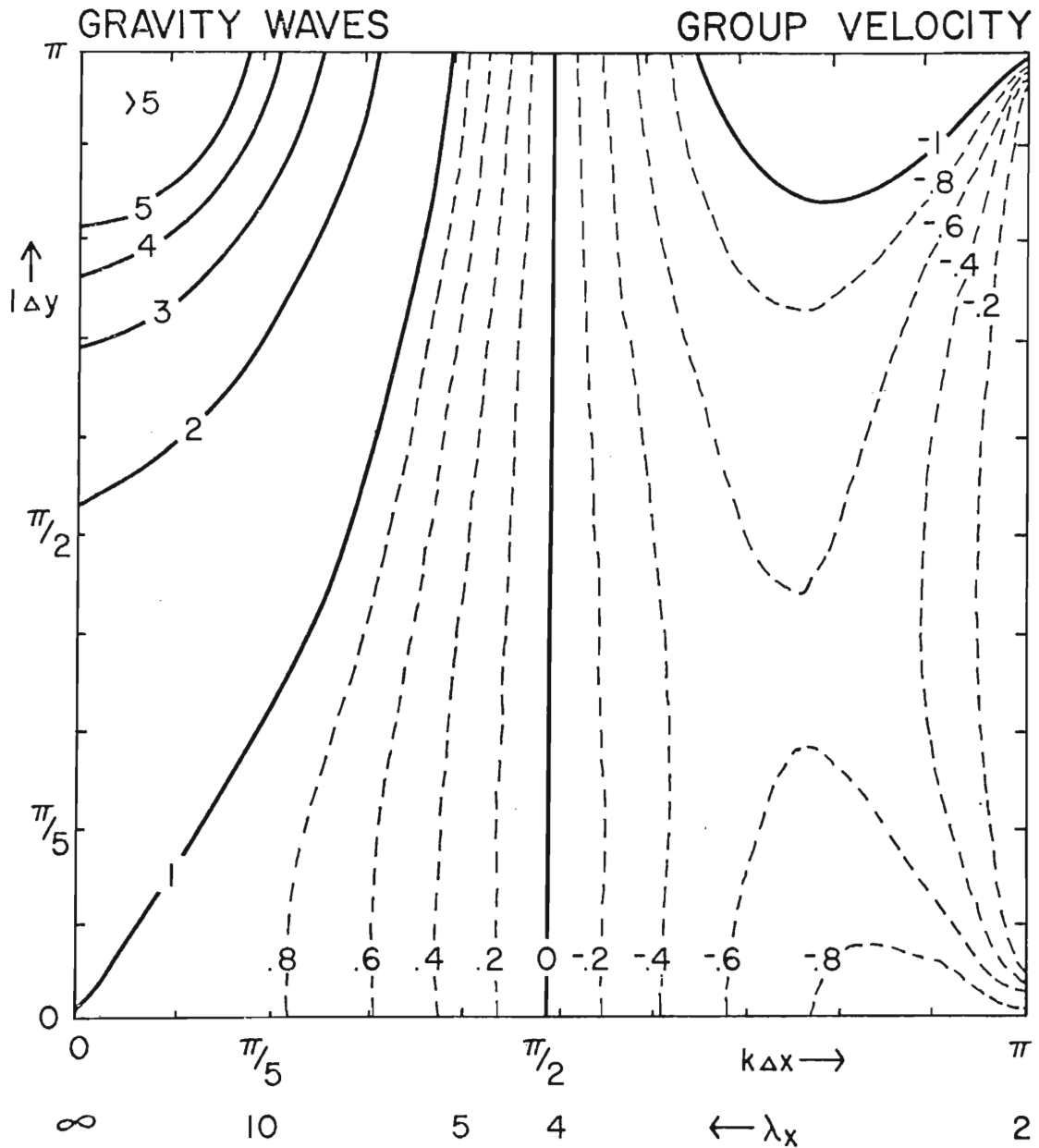
Fig. 9. x component of the vector group velocity of the gravity waves for the explicit scheme, at Λ = 0.5. The contours are normalized by the correct group velocity.
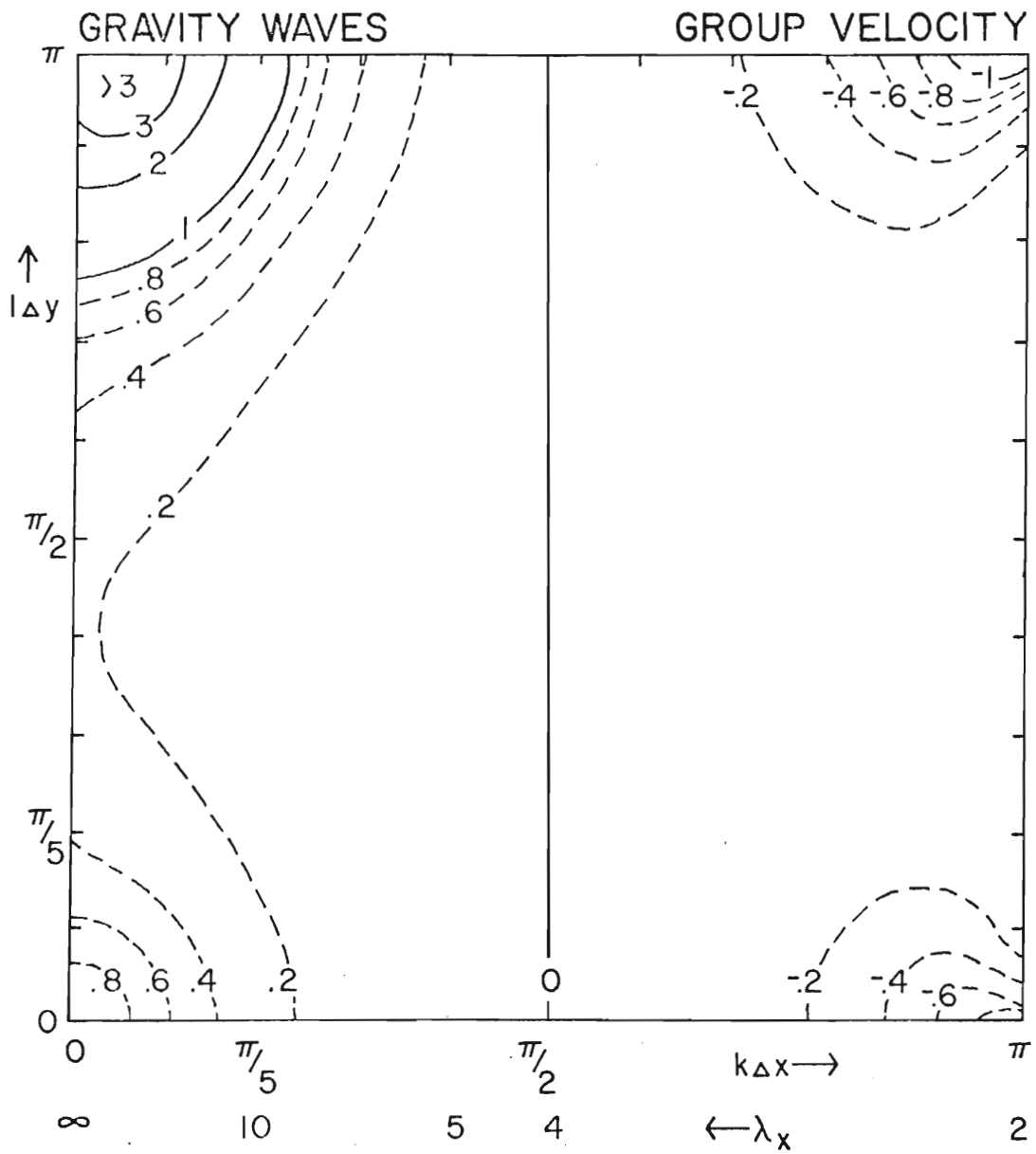
Fig. 10. Same as Fig. 9 except for the implicit scheme for $\Lambda = 10$.

perpendicular to the direction of the derivative. Figure
11 is a contour plot of the angular error in the group
velocity for the gravity waves and for the two-dimensional
advection solution. The lines are of constant angle meas-
ured positive counterclockwise from the correct angle
(note that a 180° error is equivalent to a -180° error).
This error is the same for all values of Δt and Δx, and
for the implicit and explicit schemes. The contours do not
intersect the axis for the advection plot since the problem
becomes one-dimensional and thus the angular error is
either 0 or 180°.

30

GRAVITY WAVES
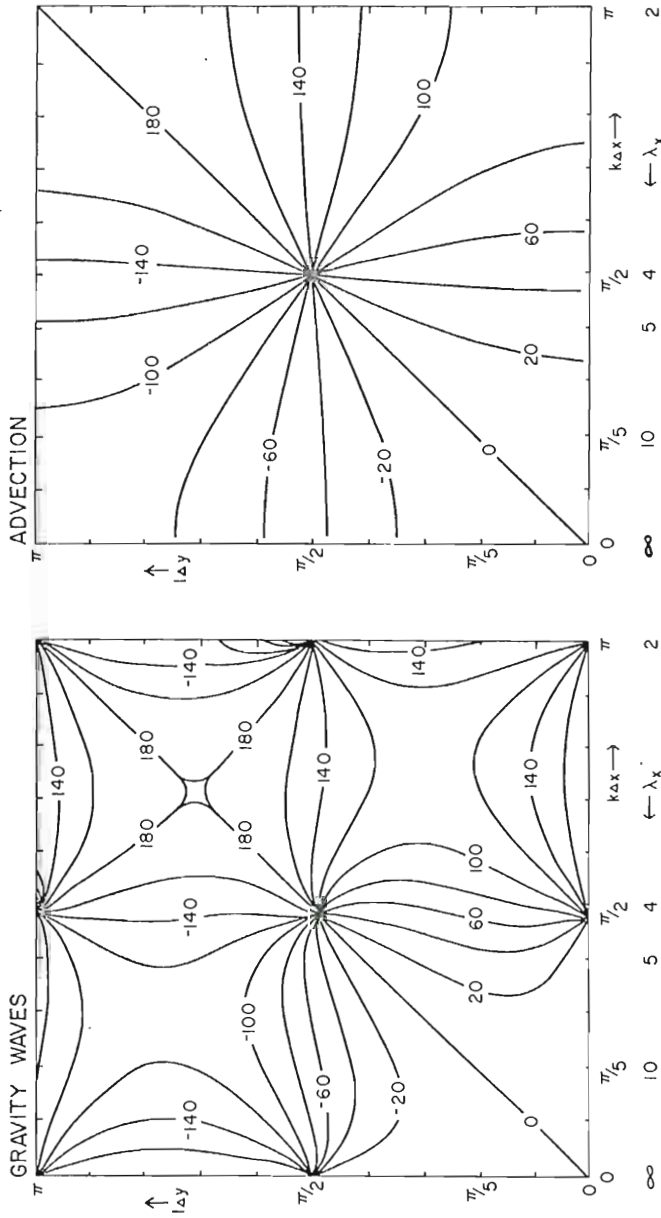
ADVECTION

(a)

(b)

Fig. 11. Contour plots of the angular error in the vector group velocity of the f the numerical schemes measured positive counterclockwise from the correct direction foion for the two-dimensional (a) gravity waves system and (b) advection equation. Lines iines in (b) do not intersect the axes since the problem changes character there.

4.   COMPLEX SYSTEMS

In the previous section we considered three simple systems, each with one distinct mode. More typical problems with more than one mode and non-constant coefficients will now be considered. The explicit and implicit formulations will be combined into a semi-implicit approach for some of the analysis.

The full shallow water primitive equations on an f-plane are

$$u_t + Uu_x + Vu_y = f_0 v - \phi_x$$
$$v_t + Uv_x + Vv_y = - f_0 u - \phi_y \tag{29}$$
$$\phi_t + U\phi_x + V\phi_y + \Phi(u_x + v_y) = 0$$

where f_0 is a constant Coriolis parameter. Assuming solutions for u, v and ϕ of the form (15) one can show that there are three phase frequencies given by

$$\omega_A = Uk + V\ell$$
$$\omega_A = Uk + V\ell \pm (f_0^2 + \Phi(k^2 + \ell^2))^{1/2} \tag{30}$$

These are merely the advection solution and two inertia-gravity waves doppler shifted by the mean state velocity.

A simple semi-implicit scheme for (29) is
gravity waves doppler shifted by the mean state velocity.

A simple semi-implicit scheme for (29) is

31

$$u^{n+1} = u^{n-1} - U \frac{\Delta t}{\Delta x} \left(u^n_{j+1} - u^n_{j-1} \right) - V \frac{\Delta t}{\Delta y} \left(v^n_{m+1} - u^n_{m-1} \right)$$

$$+ 2\Delta t \, f_0 \, v^n - \frac{\Delta t}{2\Delta t} \left[\phi^{n+1}_{j+1} + \phi^{n-1}_{j+1} - \phi^{n+1}_{j-1} - \phi^{n-1}_{j-1} \right]$$

$$v^{n+1} = v^{n-1} - U \frac{\Delta t}{\Delta x} \left(v^n_{j+1} - v^n_{j-1} \right) - V \frac{\Delta t}{\Delta y} \left(v^n_{m+1} - v^n_{m-1} \right)$$

$$- 2\Delta t \, f_0 \, u^n - \frac{\Delta t}{2\Delta y} \left[\phi^{n+1}_{m+1} + \phi^{n-1}_{m+1} - \phi^{n+1}_{m-1} - \phi^{n-1}_{m-1} \right]$$

$$\phi^{n+1} = \phi^{n-1} - U \frac{\Delta t}{\Delta x} \left[\phi^n_{j+1} - \phi^n_{j-1} \right] - V \frac{\Delta t}{\Delta y} \left(\phi^n_{m+1} - \phi^n_{m-1} \right)$$

$$- \Phi \frac{\Delta t}{2\Delta x} \left[u^{n+1}_{j+1} + u^{n-1}_{j+1} - u^{n+1}_{j-1} - u^{n-1}_{j-1} \right]$$

$$- \Phi \frac{\Delta t}{2\Delta y} \left[v^{n+1}_{m+1} + v^{n-1}_{m+1} - v^{n+1}_{m-1} - v^{n-1}_{m-1} \right] \tag{31}$$

Now, substitute solutions of the form (17) into (31). After considerable algebraic manipulation, the two roots corresponding to the inertia-gravity waves (30) are

$$\omega_{SI} = \frac{1}{\Delta t} \arcsin \left\{ \left[\pm R^{1/2} + A \right] / P \right\} \tag{32}$$

where

$$A = U \frac{\Delta t}{\Delta x} \sin k\Delta x + V \frac{\Delta t}{\Delta y} \sin \ell\Delta y$$

$$P = \Phi \, \Delta t^2 \left[\frac{\sin^2 k\Delta x}{\Delta x^2} + \frac{\sin^2 \ell\Delta y}{\Delta y^2} \right] + 1$$

$$R = \left[\Delta t^2 \, f_0^2 - 1 - A^2 \right] / P + P^2 + A^2$$

The advection root of this finite differenced set is merely Eq. (18). However, the eigenvector for this set will be different from that for the explicit version of (14). Equa-

different from that for the explicit version of (14). Equation (32) is compared with (30) in Fig. 12. The nature of the variation appears suspiciously similar to a pure

INERTIA-GRAVITY WAVES   PHASE VELOCITY
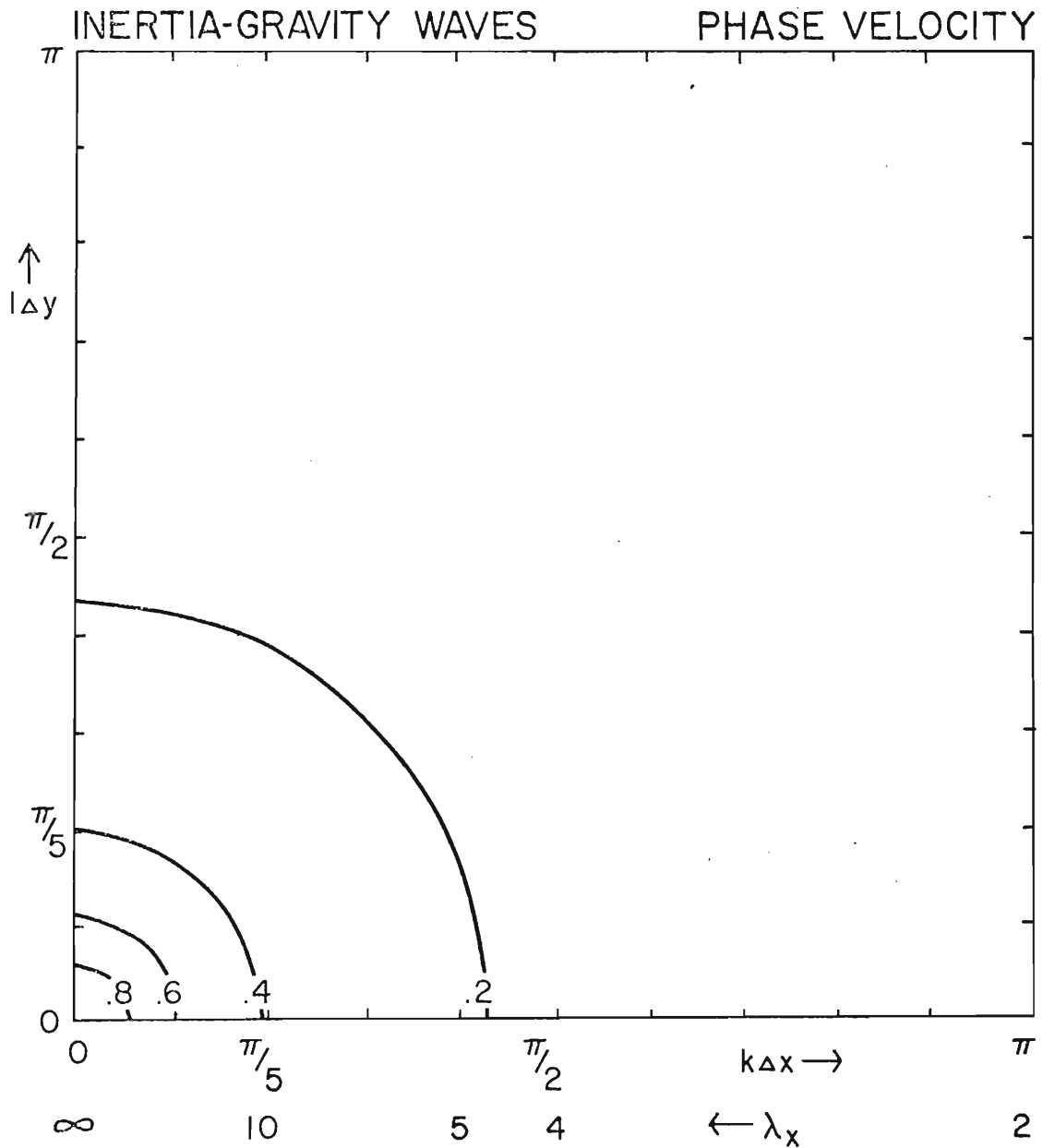
Fig. 12. Contours of the phase frequency of the
inertia-gravity waves in the semi-implicit scheme normalized
by the analytical frequency for Λ = 0.5. This figure is
consistent with the results from the gravity wave system
treated in Section 3.

implicit treatment of gravity waves; indeed, that is essentially what takes place. The advection or "quasi-geostrophic" modes were compared in Section 3. In this linear problem, the mixture of the time schemes for different modes affects those modes essentially individually.

The group velocity components are again symmetric; the x component is given by

$$C_{gA} = U \pm \Phi k \left[f_0^2 + \Phi(k^2 + \ell^2) \right]^{-1/2}$$

for the analytical, and for the semi-implicit:

$$C_{gSI} = \frac{1}{\Delta t} \left(1 - \{[\pm R^{1/2} + A]/P\}^2 \right)^{-1/2} \frac{dB}{dk}$$

where

$$\frac{dB}{dk} = \frac{-(\pm R^{1/2} + A)}{P^2} \frac{dP}{dk} + P^{-1} \left\{ \pm \frac{1}{2} R^{-1/2} \frac{dR}{dk} + \frac{dA}{dk} \right\}$$

$$\frac{dA}{dk} = U\Delta t \cos k\Delta x$$

$$\frac{dP}{dk} = 2\Phi \frac{\Delta t^2}{\Delta x} \cos k\Delta x \sin k\Delta x$$

$$\frac{dR}{dk} = [\Delta t^2 f_0^2 - 1 - A^2] \frac{dp}{dk} - 2AP \frac{dA}{dk} + 2P \frac{dP}{dk} + 2A \frac{dA}{dk}$$

One sees in Fig. 13, that near the stability limit, determined mainly by the advection speed, the inertia-gravity waves are transporting their energy far below the correct speed. This is consistent with the results of one-dimensional gravity waves presented in Fig. 7 for an implicit

sional gravity waves presented in Fig. 7 for an implicit

treatment at high values of Λ.
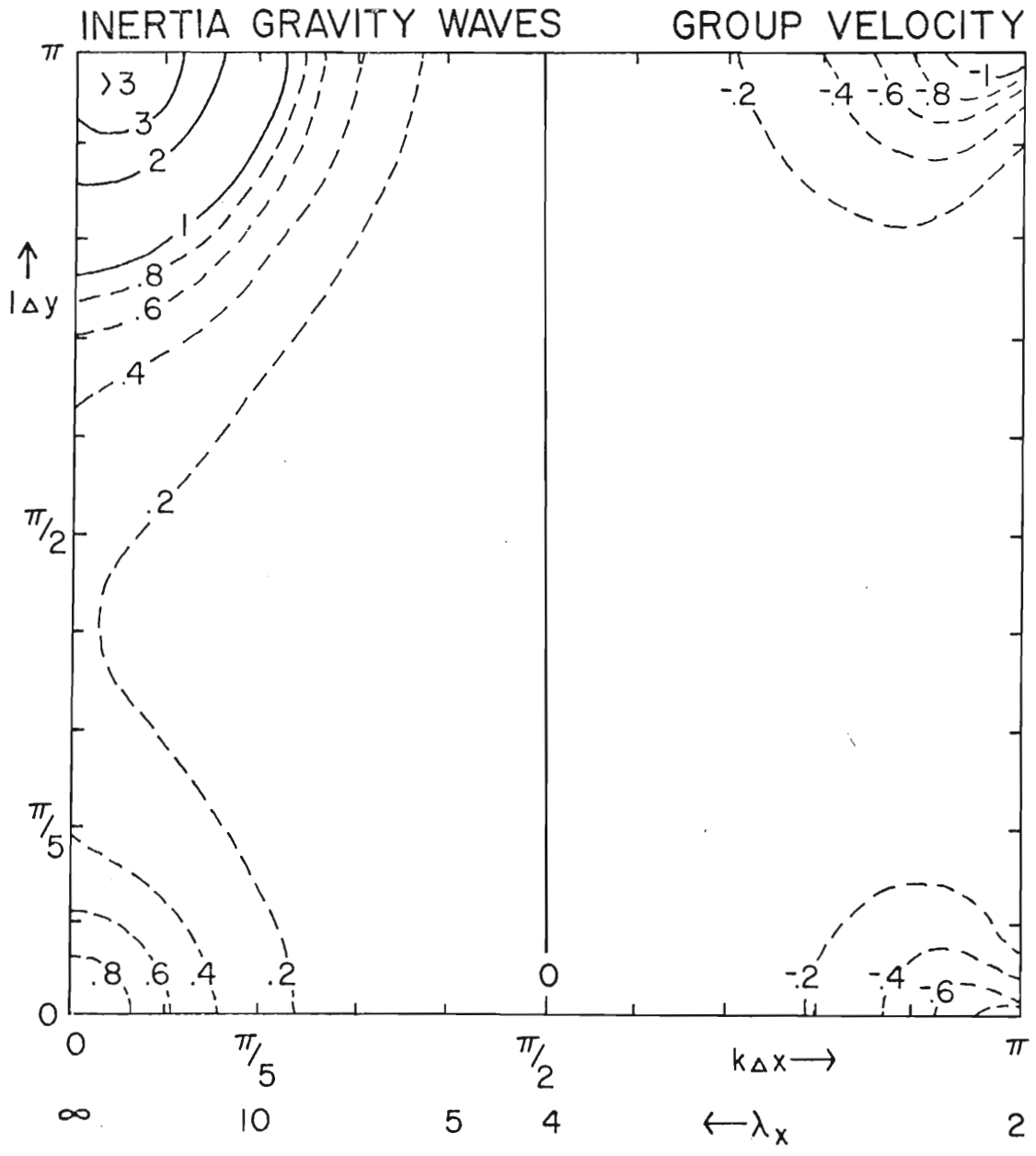
Fig. 13. x component of the vector group velocity of
the semi-implicit case normalized by the correct value for
Λ ~ 10. Note similarity to the pure gravity wave case pre-
sented in Fig. 10.

Results similar to those from the f-plane primitive
equations are obtained for a semi-implicit two-layer set.
The internal gravity waves are handled like slow moving
gravity waves, the external gravity waves are handled as
in the previous case. The treatment of the different
components of the modes, either explicitly or implicitly,
again separates out. This is in part a consequence of the
CFL condition being exceeded (that is Λ large) only for the
external gravity waves, so that the explicit and implicit
treatments of the other modes are nearly equivalent. It is
therefore reasonable to infer that the mixture of time
schemes when the modes are widely separated, affects those
modes essentially individually. Since semi-implicit schemes
are used when one has waves which move much faster than
others, the behavior of either set of waves can be closely
approximated by an implicit or explicit treatment of the
whole model. This is not too surprising for a linear model
and is fortunate, since the analytical derivation of the
semi-implicit phase frequencies can be intractable when
either an explicit or implicit scheme is solvable.

The quasi-geostrophic mode is generally of more
interest than inertia-gravity waves for the atmosphere
since it contains the major part of the energy in the large-
scale mid-latitude flows. However, inertia-gravity waves
are of major interest in oceanography, both for tidal prob-
scale mid-latitude flows. However, inertia-gravity waves
are of major interest in oceanography, both for tidal prob-
lems and on a smaller scale for internal waves. The very

small inertia-gravity waves are generally of less interest. The inertia-gravity waves are the apparent mechanism for bringing the flow and pressure fields into geostrophic balance for both the atmosphere and ocean (e.g., Blumen, 1972). Thus, errors in their estimation may affect the adjustment time scales of numerical models.

Next we will introduce a variable Coriolis parameter, f, by considering the nondivergent barotropic vorticity equation. This model was commonly used for large-scale atmospheric prediction, until more sophisticated, primitive equation models were developed. However, it remains an important tool for representing many important dynamic processes. Gates has studied the errors caused by numerical integration of a one-dimensional form of this equation. The results to be presented here for phase speeds are consistent with his analytical studies (Gates, 1959) and comparisons of actual integrations with known harmonic solutions (Gates and Riegel, 1962).

If we define a streamfunction ψ:

$$v = \frac{\partial \psi}{\partial x}, \qquad u = -\frac{\partial \psi}{\partial y}$$

then our vorticity equation, for two dimensions, can be written

$$\frac{\partial}{\partial t} \nabla^2 \psi + \beta \frac{\partial \psi}{\partial x} = 0 \qquad (33)$$

where ∇^2 is the horizontal Laplacian operator, and $\beta = df/dy$.

Assume that the solution for ψ has the form (15), then (33) yields a dispersion relation

$$\omega_A = \frac{-\beta k}{k^2 + \ell^2} \tag{34}$$

If we finite difference (33) in the following explicit manner, using a five point, "second order" Laplacian operator

$$\frac{1}{\Delta x^2}\left(\psi_{j+1}^{n+1} + \psi_{j-1}^{n+1} - 2\psi^{n+1}\right) + \frac{1}{\Delta y^2}\left(\psi_{m+1}^{n+1} + \psi_{m-1}^{n+1} - 2\psi^{n+1}\right) =$$

$$\frac{1}{\Delta x^2}\left(\psi_{j+1}^{n-1} + \psi_{j-1}^{n-1} - 2\psi^{n-1}\right) + \frac{1}{\Delta y^2}\left(\psi_{m+1}^{n-1} + \psi_{m-1}^{n-1} - 2\psi^{n-1}\right) -$$

$$- \frac{\Delta t \beta}{\Delta x}\left(\psi_{j+1}^{n} - \psi_{j-1}^{n}\right)$$

the corresponding phase frequency is obtained

$$\omega_E = \frac{1}{\Delta t} \arcsin\left\{\frac{-\beta \Delta t \sin k\Delta x}{4\Delta x\ A}\right\}$$

where

$$A = \frac{\sin^2\left(\frac{k\Delta x}{2}\right)}{\Delta x^2} + \frac{\sin^2\left(\frac{\ell\Delta y}{2}\right)}{\Delta y^2}$$

An implicit scheme, where all the terms are at the n and n+1 time levels dictates a phase frequency satisfying

$$\omega_I = \frac{2}{\Delta t} \arctan\left\{\frac{-\beta \Delta t \sin k\Delta x}{8\ \Delta x\ A}\right\}$$

In a typical problem where Rossby waves are being studied, there will be other modes which may require a much lower ratio of $\Delta t/\Delta x$ for stability than that for pure Rossby waves. Therefore a value of $\Delta t/\Delta x = 10^{-1}$ s m^{-1} was chosen. Since β is so small (around 10^{-11} m^{-1} s^{-1} for the earth) and the arctangent of a small number approximates the arcsine of that number, then the two treatments are essentially identi-
the arctangent of a small number approximates the arcsine of that number, then the two treatments are essentially identical. The phase velocity of the finite differenced equations

is compared with the continuous solution (34) in Fig. 14. When the motion is generally directed more in the x direction, the numerical schemes underestimate the phase speed. When the motion is mostly in the y direction the velocity is overestimated for small k.

The components of the group velocity here are not symmetric; consider the x component first. The analytical and numerical components are, respectively

$$C_{gAx} = \beta \left(\frac{k^2 - \ell^2}{(k^2 + \ell^2)^2} \right)$$

$$C_{gNx} \simeq \frac{\beta}{4} \left| \frac{\sin k\Delta x \, \sin \frac{k\Delta x}{2} \, \cos \frac{k\Delta x}{2}}{\Delta x^2 \, A^2} - \frac{\cos k\Delta x}{A} \right|$$

The x component is generally overestimated by the numerical scheme, particularly for large values of k and ℓ. Note that the energy in waves which propagate at a 45° angle will have a nonzero x component in the numerical scheme. In contrast, the y component is handled reasonably well. The y component functions of the group velocity are

$$C_{gAy} = \frac{2\beta\ell k}{(k^2 + \ell^2)^2}$$

and

$$C_{gNy} \simeq \frac{\beta}{4\Delta x} \left(\frac{\sin k\Delta x \, \sin \frac{\ell\Delta y}{2} \, \cos \frac{\ell\Delta y}{2}}{\Delta y \, A^2} \right)$$

for the analytical and numerical models. Again there will be significant errors in the direction as well as the magnitude of the group velocity in the numerical model. These
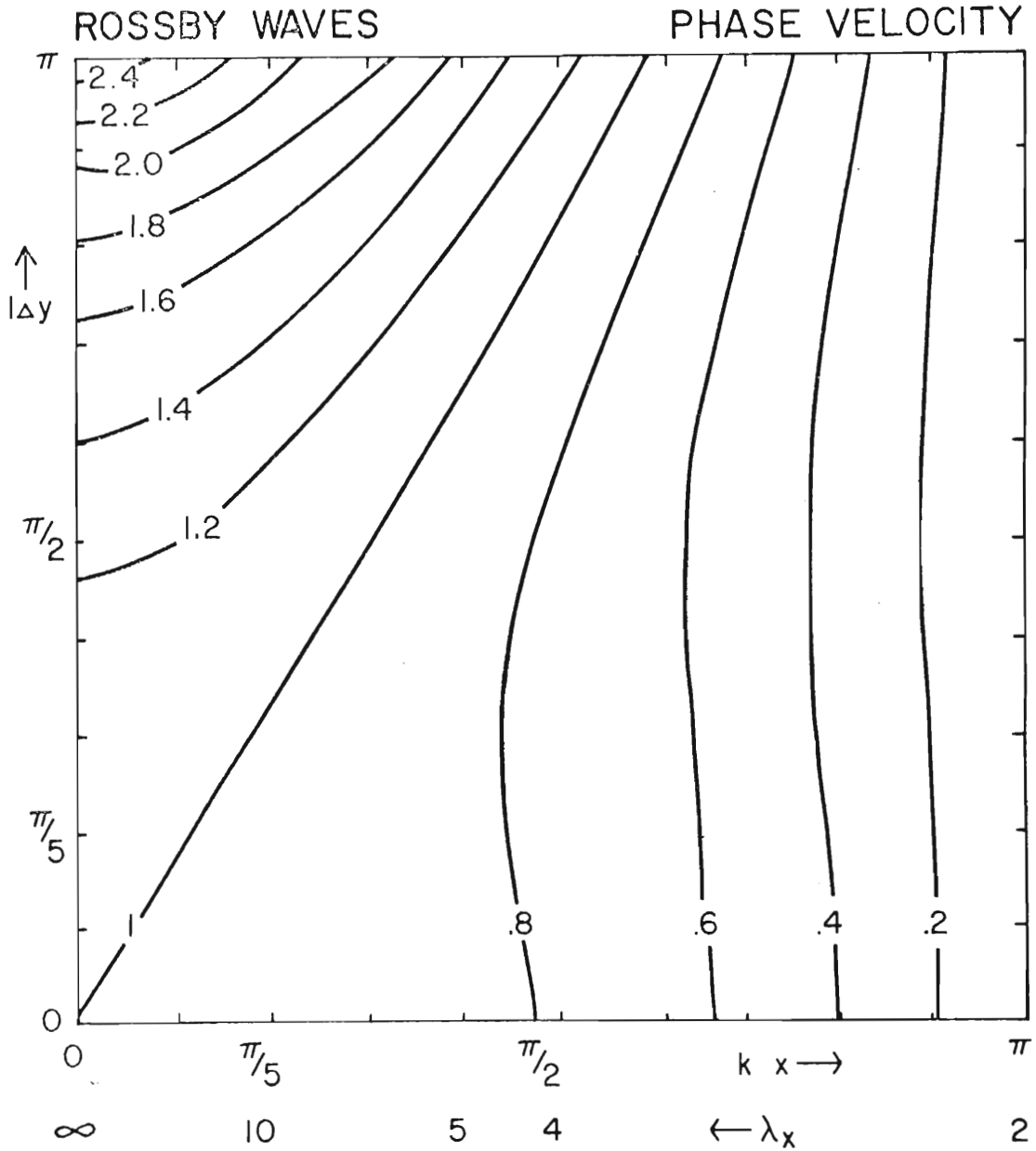
Fig. 14. Phase velocity for the numerical treatment of the nondivergent barotropic vorticity equation normalized by the correct velocity of the Rossby waves.

are illustrated in Figs. 15 and 16, respectively.

The final case examined will be the primitive equations on an equatorial β-plane. Our governing equations are a special case of Laplace's tidal equations (LTE):

$$u_t + Uu_x = \beta y v - \phi_x$$

$$v_t + Uv_y = -\beta y u - \phi_y \qquad (35)$$

$$\phi_t + U\phi_x + \Phi(u_x + v_y) = 0$$

This and similar versions of LTE have been extensively studied analytically. Miles (1974) presents an excellent review of the classical tidal problem. The solutions of (35) are contained in Matsuno (1966); they are two waves which behave like gravity waves and a third which is like a Rossby wave. Lindzen (1970) referred to the latter solution as a Rossby-Haurwitz wave. He allowed for vertical propagation in his treatment and also derived an approximate solution for a mid-latitude β-plane.

To solve (35) substitute a solution of the form

$$F(y) \ e^{i(kx-\omega t)}$$

for u, v and φ, where the y dependence, denoted by some function F, is left unspecified. The set of equations is then manipulated into an ordinary differential equation (ODE) in y for v. (The details of this derivation are presented in Appendix 2.)
(ODE) in y for v. (The details of this derivation are presented in Appendix 2.)

$$\frac{d^2v}{dy^2} + \left(\frac{A^2}{\Phi} - k^2 - \frac{\beta k}{A} - \frac{\beta^2 y^2}{\Phi} \right) v = 0 \qquad (36)$$
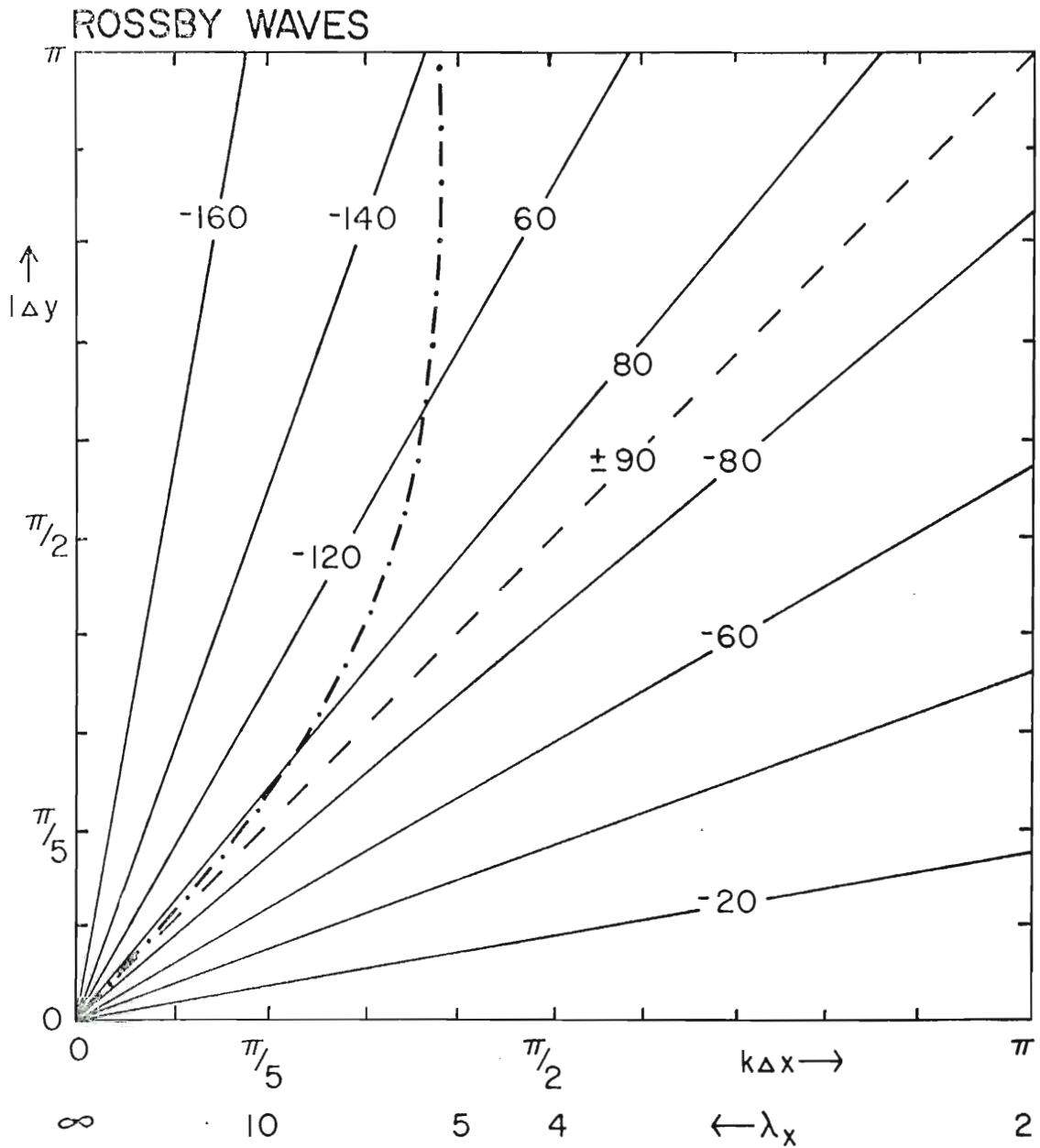
Fig. 15. Contour plot of the directional error in the vector group velocity for the numerical formulations of the vorticity equation. The x components of the correct and numerical group velocities are zero along the dashed and dot-dashed lines, respectively.
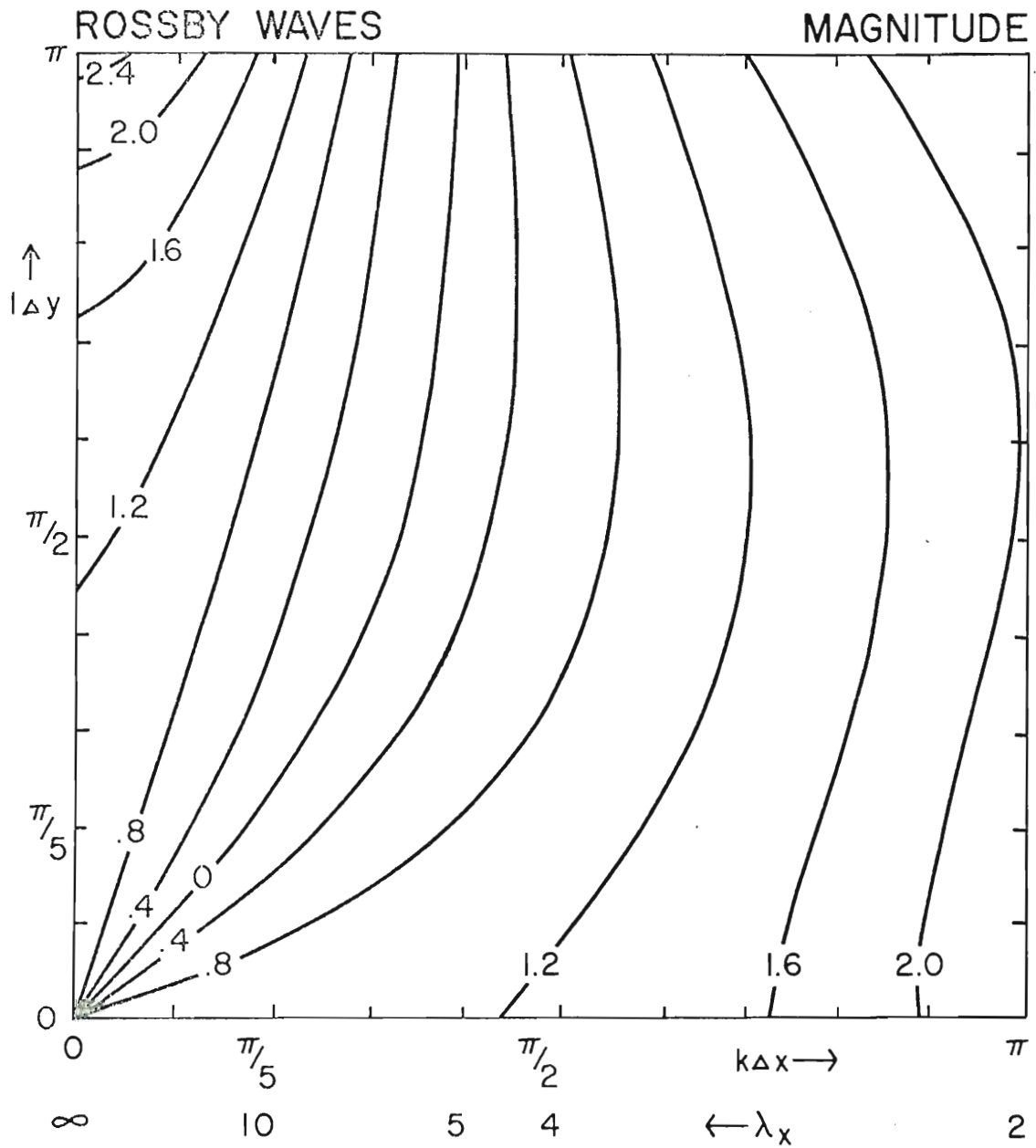
Fig. 16. Plot of the magnitude of the group velocity for the numerical schemes normalized by the correct value.

where

$$A = \omega - Uk$$

The boundary conditions for (36) are $v \to 0$ as $y \to \pm \infty$.

Assume that the y dependent part of the solution for v has the form

$$v(y) = G(y) e^{-\frac{1}{2} \frac{\beta}{\Phi} y^2} \tag{37}$$

This will leave the following ODE for G:

$$G'' - \frac{2\beta}{\sqrt{\Phi}} y \, G' + \left(\frac{A^2}{\Phi} - k^2 - \frac{\beta k}{A} - \frac{\beta^2 y^2}{\Phi} \right) G = 0$$

where primes denote differentiation with respect to y. If we assume a polynomial solution for G:

$$G = \sum_{\eta=0}^{\infty} C_\eta \, y^\eta$$

then one obtains the recursion formula

$$C_{\eta+2} = -C_\eta \frac{\left(\frac{A^2}{\Phi} - k^2 - \frac{\beta k}{A} - \frac{\beta}{\sqrt{\Phi}} - \frac{2\beta\eta}{\sqrt{\Phi}} \right)}{(\eta + 2)(\eta + 1)}$$

In order for v to satisfy the boundary conditions, G must be truncated at some value of η. Thus there are an infinite number of modes in the y direction — one for each value of η. They satisfy the dispersion relation

$$A^3 + aA + b = 0 \tag{38}$$

where

$$a = -(k^2\Phi + (2\eta + 1) \beta \sqrt{\Phi})$$

where

$$a = -(k^2\Phi + (2\eta + 1) \beta \sqrt{\Phi})$$

$$b = -k \Phi \beta$$

The roots of (38) are found via the closed form solution to a cubic. Matsuno discusses several of the lowest modes. Only the $\eta = 0$ mode will be considered in the following evaluation.

An explicit formulation of (35) is

$$u^{n+1} = u^{n-1} - U \frac{\Delta t}{\Delta x} \left[u^n_{j+1} - u^n_{j-1} \right] + 2\beta m \Delta y \Delta t \quad v^n$$

$$- \frac{\Delta t}{\Delta x} \left[\phi^n_{j+1} - \phi^n_{j-1} \right]$$

$$v^{n+1} = v^{n-1} - U \frac{\Delta t}{\Delta x} \left[v^n_{j+1} - v^n_{j-1} \right] - 2\beta m \Delta y \Delta t \quad u^n$$

$$- \frac{\Delta t}{\Delta y} \left[\phi^n_{m+1} - \phi^n_{m-1} \right] \tag{39}$$

$$\phi^{n+1} = \phi^{n-1} - U \frac{\Delta t}{\Delta x} \left[\phi^n_{j+1} - \phi^n_{j-1} \right] - \Phi \frac{\Delta t}{\Delta x} \left[u^n_{j+1} - u^n_{j-1} \right]$$

$$- \Phi \frac{\Delta t}{\Delta y} \left[v^n_{m+1} - v^n_{m-1} \right]$$

The analytic phase frequencies have no y dependence and we anticipate that the frequencies of the numerical scheme do not either. As was pointed out in Section 2, in this problem it is not strictly correct to replace y by $m\Delta y$ in the form of the solution because the functional form of the derivatives will be inconsistent. This may cause the frequency derived for the finite differenced set of equations to depend on y. This difficulty in obtaining a consistent solution form illustrates a limitation of this analysis method. However, we will use the form consistent solution form illustrates a limitation of this analysis method. However, we will use the form

$$\exp \left\{ \frac{-\beta m^2 \Delta y^2}{2\sqrt{\Phi}} + i \; (kj\Delta x - \omega n \Delta t) \right\} \qquad (40)$$

which corresponds to the first, $\eta = 0$, mode of the analytical solution. We will consider this approximation reasonable, if we obtain phase frequencies that vary only by a minute amount with y. This would indicate that (40) is very close to the correct solution. If the y dependence turned out to be strong, then an eigenvector analysis (see Appendix 2) would be required.

After substitution of the form (40), and considerable algebraic manipulation, one obtains the cubic equation

$$D^3 + aD + b = 0$$

where

$$D = \sin \omega \Delta t - U \frac{\Delta t}{\Delta x} \sin k\Delta x$$

$$c = -\frac{\beta \Delta y^2}{2\sqrt{\Phi}}$$

$$a = -\left[\Phi \frac{\Delta t^2}{\Delta x^2} \sin^2 k\Delta x + \beta^2 \Delta t^2 m^2 \Delta y^2 \right.$$
$$\left. + (1 - e^{4c} \cosh (4cm)) \; \Phi \frac{\Delta t^2}{2\Delta y^2} \right]$$

$$b = \beta \Delta t \Delta y \; \Phi \frac{\Delta t}{\Delta x} \sin k\Delta x \; e^c \cosh (2cm)$$

because β is so small, for a terrestrial problem the dependence of D on y (and therefore of ω_E) is weak. The difference in ω between m = 0 and m = 10 for $\Delta y = 10^5$ m is in the 4th or

in ω between m = 0 and m = 10 for $\Delta y = 10^5$ m is in the 4th or larger significant digit, thus our use of (40) appears reasonable.

The phase frequencies for the eastward moving gravity wave (E), westward moving gravity wave (W) and Rossby-Haurwitz wave (R) are presented in Figs. 17 and 18. The details of the curves for $0 \le k\Delta x \le \pi/20$ are shown in the insert. The rather peculiar behavior of the Rossby-like waves may be due to the use of the approximate solution (40). The analytical roots are complex for values of $\sim 0.01 < k\Delta x < 0.05$ (for $\Phi = 10^4 \ m^2 \ s^{-2}$, $\beta = 10^{-11} \ m^{-1} \ s^{-1}$ and $\Delta x = 10^5 \ m$) whereas the numerical frequencies are real; no comparison was made in that range. The Rossby-Haurwitz waves move at about a tenth the speed of the gravity waves, so the linear stability limit, $\Lambda = 1$, is determined by the gravity waves. For small Λ, the explicit and implicit versions are again nearly identical. Therefore, Fig. 17, where $\Lambda \sim 0.5$, is representative of either scheme. The curves have the same basic shape as in previous, simpler, problems. Figure 18 compares the implicit and analytical phase velocities for $\Lambda \sim 5$. The gravity waves are slowed down appreciably, but the much slower Rossby-Haurwitz wave is not appreciably changed by the tenfold increase in Λ.

Since these results are analogous to those from simpler systems it seems reasonable to conclude that they are general properties of the finite difference scheme. That is, one would expect the truncation error of this scheme to have these general properties for similar types That is, one would expect the truncation error of this scheme to have these general properties for similar types of wave motions.
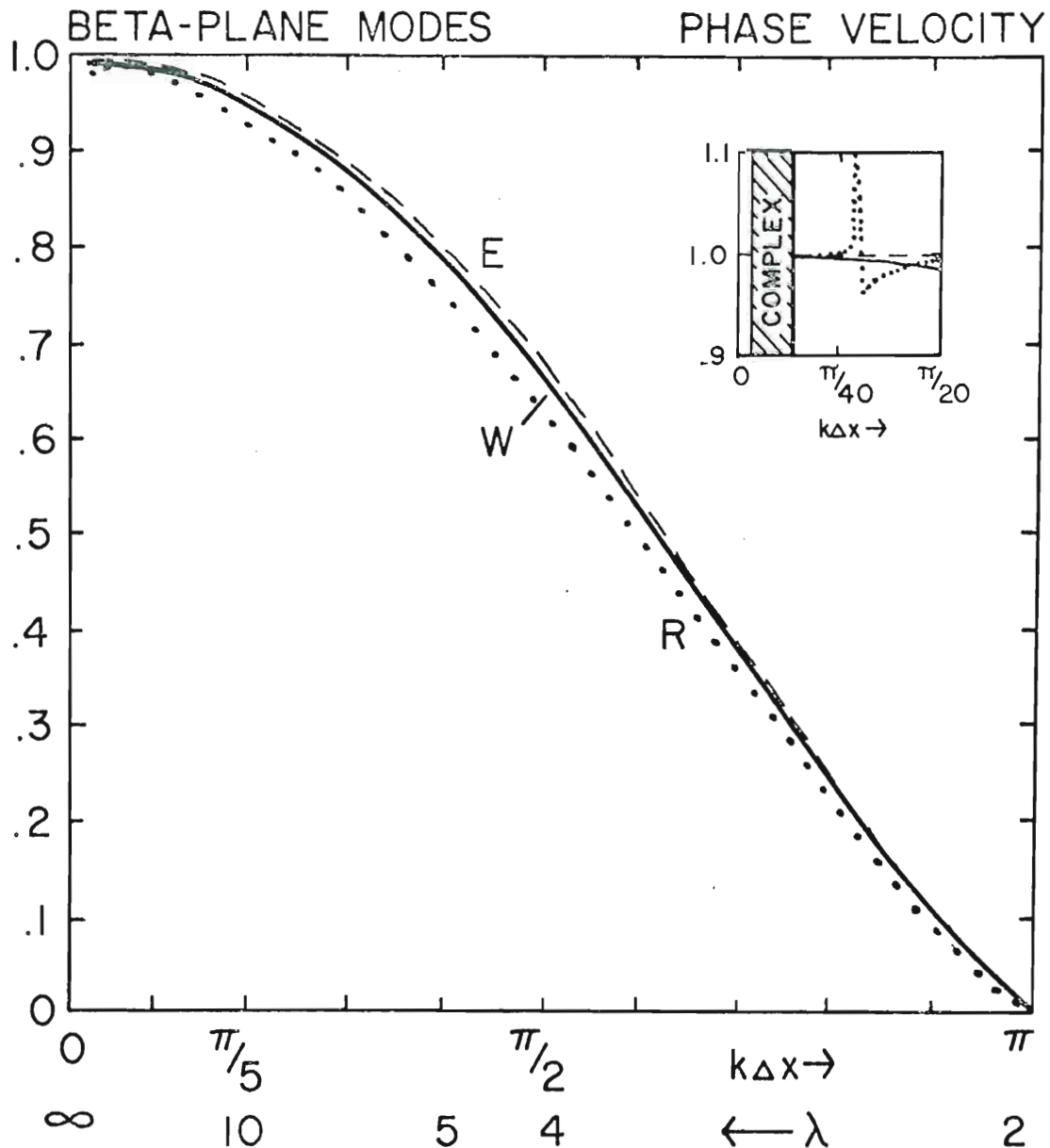
Fig. 17. Phase velocity for the explicit formulation
(Λ = 0.5) of the equatorial β-plane system normalized by the
correct speeds. (E) is the eastward propagating gravity
wave, (W) the westward propagating gravity wave and (R) the
Rossby-Haurwitz wave. The insert shows the details of the
error at small values of k. The analytical solution is complex
for ~ 1×10^{-7} < k < 5×10^{-7} m^{-1} for β = 10^{-11} m^{-1} s^{-1} and Φ =
10^4 m^2 s^{-2}, here Δx = 10^5 m.
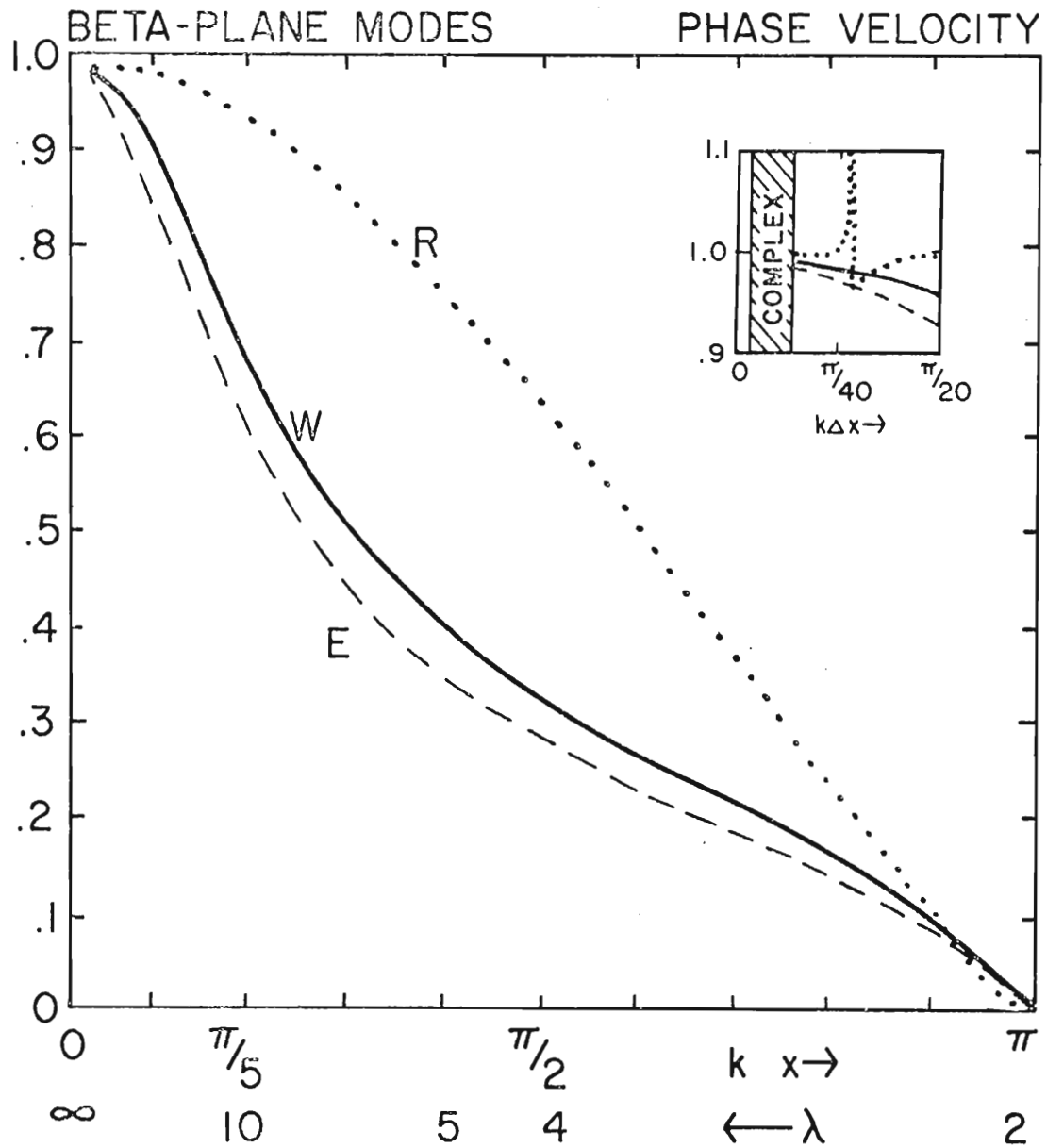
Fig. 18. Same as Fig. 17 but for the implicit scheme at $\Lambda \simeq 5$. Note that the tenfold increase in Λ had little effect on the Rossby-Haurwitz waves. These results are consistent with previous results for simpler systems.

It would be possible, though tedious, to compute the group velocity for this problem. However, since an approximate solution was used, it may not be appropriate to do so. That is, the misrepresentation caused by using (40) may be magnified by taking the derivative of the numerical frequency solutions.

5. CONCLUSIONS

This study was undertaken to attempt to quantitatively determine the error introduced by finite differencing. Since the field is so broad, testing many schemes, and various boundary and staggered grid formulations was not feasible. Therefore the study was also designed to outline a straight-forward method of determining the error which can be applied in practice. The determination is imperative since statements concerning the order of a Taylor series are clearly inadequate to represent the truncation error.

This study confirms several previous results. Namely, that explicit formulations tend to overestimate oscillations whereas implicit formulations underestimate them. Space differencing produces dispersive waves even if the true waves are not dispersive. The phase velocity errors increase as the wavenumber increases, or equivalently, as the number of grid points resolving the wave decreases. Implicit schemes retain stability by slowing down the waves, which is the price extracted for the economy of large time steps. Generally, only the longest waves are handled reasonably well. Several new aspects of the truncation error have been revealed, particularly through comparison of the group velo-Several new aspects of the truncation error have been revealed, particularly through comparison of the group velocities. The dispersive nature introduced by the finite

51

differencing causes significant angular, as well as magnitude, errors in the group velocity. Since the group velocity is an indication of the propagation of the energy, this could cause grave dynamical effects in a numerical model, particularly if the short waves carry a significant fraction of the energy. It is also apparent that some of the characteristics of the error evident in simple systems reappear when the same scheme is used in more complex models with similar solutions.

This work examined several linear systems. The intention being, to study the error in the scheme without the additional complications of nonlinear effects. This method could be extended to other types of problems. Further work is currently being considered on formulating boundary and initial conditions based on the interrelationships of variables in the discrete system.

Outline of the method applied to a simple equation

$$\phi_t + U\phi_x + V\phi_y = 0 \qquad (A1.1)$$

insert

$$\phi = A e^{i(kx+\ell y-\omega t)}$$

into (A1.1) and

$$- i\omega + ikU + i\ell V = 0$$

$$\omega = Uk + V\ell$$

An explicit formulation of (A1.1) is

$$\phi^{n+1} = \phi^{n-1} - U \frac{\Delta t}{\Delta x}\left(\phi^n_{j+1} - \phi^n_{j-1}\right) - V \frac{\Delta t}{\Delta y}\left(\phi^n_{m+1} - \phi^n_{m-1}\right)$$

$$\qquad (A1.2)$$

use a solution

$$\phi = A e^{i(kj\Delta x + \ell m\Delta y - \omega n\Delta t)} \qquad (A1.3)$$

which, when inserted into (A1.2), reduces to

$$e^{-i\omega\Delta t} = e^{i\omega\Delta t} - U \frac{\Delta t}{\Delta x}\left(e^{ik\Delta x} - e^{-ik\Delta x}\right) - V \frac{\Delta t}{\Delta y}\left(e^{i\ell\Delta y} - e^{-i\ell\Delta y}\right)$$

using Euler's relations:

$$\sin \omega\Delta t = U \frac{\Delta t}{\Delta x} \sin k\Delta x + V \frac{\Delta t}{\Delta y} \sin \ell\Delta y$$

thus, for the explicit scheme (A1.2)

thus, for the explicit scheme (A1.2)

$$\omega_E = \frac{1}{\Delta t} \arcsin\left[U \frac{\Delta t}{\Delta x} \sin k\Delta x + V \frac{\Delta t}{\Delta y} \sin \Delta y\right]$$

An implicit scheme for (A1.1) is

$$\phi^{n+1} = \phi^n - U \frac{\Delta t}{4\Delta x} \left(\phi^n_{j+1} + \phi^{n+1}_{j+1} - \phi^n_{j-1} - \phi^{n+1}_{j-1} \right)$$

$$- V \frac{\Delta t}{4\Delta y} \left(\phi^n_{m+1} + \phi^{n+1}_{m+1} - \phi^n_{m-1} - \phi^{n+1}_{m-1} \right)$$

in a similar fashion, substitution of (A1.3) yields

$$\sin \frac{\omega\Delta t}{2} = U \frac{\Delta t}{4\Delta x} \left[2 \cos \frac{\omega\Delta t}{2} \sin k\Delta x \right] + V \frac{\Delta t}{4\Delta y} \left[2 \cos \frac{\omega\Delta t}{2} \sin \ell\Delta y \right]$$

and

$$\omega_I = \frac{2}{\Delta t} \arctan \left[U \frac{\Delta t}{2\Delta x} \sin k\Delta x + V \frac{\Delta t}{2\Delta y} \sin \ell\Delta y \right]$$

APPENDIX 2

Analytic solutions for the beta-plane primitive equations are derived. The more vigorous eigenvector analysis is sketched.

Analytic Beta-Plane Solutions

Given:

$$u_t + Uu_x = fv - \phi_x$$

$$v_t + Uv_x = - fu - \phi_y$$

$$\phi_t + U\phi_x + \Phi (u_x + v_y) = 0$$

substitute solutions for u, v and of the form

$$F(y)\ e^{i(kx-\omega t)}$$

where the y variation is left unspecified. Then

$$- iAu = \beta yv - ik\phi \qquad (A2.1)$$

$$- iAv = - \beta yu - \frac{\partial \phi}{\partial y} \qquad (A2.2)$$

$$- iA\phi + i\Phi ku + \Phi \frac{\partial v}{\partial y} = 0 \qquad (A2.3)$$

where

$$A = \omega - Uk$$

Take $\partial/\partial y$ (A2.3), substitute that into (A2.2), and multiply by iA;

$$iA\ (- iAv + \beta yu) + ik\Phi \frac{\partial u}{\partial y} + \Phi \frac{\partial^2 v}{\partial y^2} = 0 \qquad (A2.4)$$

by iA,

$$iA\ (- iAv + \beta yu) + ik\Phi \frac{\partial u}{\partial y} + \Phi \frac{\partial^2 v}{\partial y^2} = 0 \qquad (A2.4)$$

55

Eliminate ϕ between (A2.1) and (A2.3): multiply (A2.1) by A and (A2.3) by k, add;

$$- iA^2 u - \beta y A v + i\Phi k^2 u + \Phi k \frac{\partial v}{\partial y} = 0 \qquad (A2.5)$$

Eliminate ϕ between (A2.1) and (A2.2) to get;

$$iA \frac{\partial u}{\partial y} + i\beta y k u + A k v + \beta y \frac{\partial v}{\partial y} + \beta v = 0 \qquad (A2.6)$$

Eliminate $\partial v / \partial y$ term between (A2.5) and (A2.6):

$$- iA^2 \beta y u - \beta^2 y^2 A v - ik A\Phi \frac{\partial u}{\partial y} - A\Phi k^2 v - \beta \Phi k v = o \qquad (A2.7)$$

Use (A2.7) as a relation for

$$iA\beta y u - ik\Phi \frac{\partial u}{\partial y}$$

Substitute that into (A2.4), and obtain

$$\frac{\partial^2 v}{\partial y^2} + \left(\frac{A^2}{\Phi} - k^2 - \frac{\beta k}{A} - \frac{\beta^2 y^2}{\Phi} \right) v = 0 \qquad (A2.8)$$

let

$$v(y) = G(y) \, e^{-\frac{1}{2} \frac{\beta}{\sqrt{\Phi}} y^2}$$

substitute this into (A2.8):

$$G'' - \frac{2\beta y}{\sqrt{\Phi}} G' + \left(\frac{A^2}{\Phi} - k^2 - \frac{\beta k}{A} - \frac{\beta^2 y^2}{\Phi} \right) G = 0 \qquad (A2.9)$$

Assume G is polynomial:

$$G = \sum_{\eta=0}^{\infty} C_\eta y^\eta$$

Insert this into (A2.9) to obtain the recursion formula:

$$\underline{} \quad C_\eta \left[\frac{A^2}{\Phi} - k^2 - \frac{\beta k}{A} - \frac{\beta}{\sqrt{\Phi}} - 2 \frac{\beta}{\sqrt{\Phi}} \eta \right] \qquad (A2.10)$$

$$C_{\eta+2} = - C_\eta \frac{\left[\frac{A^2}{\Phi} - k^2 - \frac{\beta k}{A} - \frac{\beta}{\sqrt{\Phi}} - 2 \frac{\beta}{\sqrt{\Phi}} \eta \right]}{(\eta + 2)(\eta + 1)} \qquad (A2.10)$$

For large η,

$$\frac{C_{\eta+2}}{C_\eta} \sim \frac{2\beta}{\eta\sqrt{\Phi}}$$

thus

$$G \sim \sum_{\eta=0}^{\infty} \frac{2\beta y^\eta}{\sqrt{\Phi}\ \eta!}$$

Thus for finite v at large y, G must be truncated. For each value of η there will be a dispersion relation, obtained from the recursion formula (A2.10):

$$2\eta + 1 = \frac{A^2}{\sqrt{\Phi}\beta} - \frac{k^2\sqrt{\Phi}}{\beta} - \frac{k\sqrt{\Phi}}{A} \tag{A2.11}$$

or

$$A^3 - A\left(k^2\Phi + (2\eta+1)\beta\sqrt{\Phi}\right) - k\Phi\beta = 0 \tag{A2.12}$$

Note that if (A2.12) is written

$$A^3 - aA + b = 0$$

then $\frac{a^3}{27} + \frac{b^2}{4} < 0$ and the roots are real, so

$$\omega_i = Uk + A_i$$

(except for a small range of k: $1.3\times10^{-7}\ \mathrm{m}^{-1} \lesssim k \lesssim 5\times10^{-7}\mathrm{m}^{-1}$ where the roots are complex, for $\Phi = 10^4\ \mathrm{m}^2\ \mathrm{s}^{-2}$ and $\beta = 10^{-11}$ $\mathrm{m}^{-1}\ \mathrm{s}^{-1}$.)

Numerical Beta Plane Solutions

An explicit formulation is

$$u^{n+1} = u^{n-1} - U\frac{\Delta t}{\Delta x}\left(u_{j+1}^n - u_{j-1}^n\right) + 2\beta m\Delta y\Delta t\ v^n$$

$$u^{n+1} = u^{n-1} - U\frac{\Delta t}{\Delta x}\left(u_{j+1}^n - u_{j-1}^n\right) + 2\beta m\Delta y\Delta t\ v^n$$

$$- \frac{\Delta t}{\Delta x}\left(\phi_{j+1}^n - \phi_{j-1}^n\right)$$

$$v^{n+1} = v^{n-1} - U \frac{\Delta t}{\Delta x} \left(v^n_{n+1} - v^n_{n-1} \right) - 2\beta m \Delta y \Delta t \ u^n$$

$$- \frac{\Delta t}{\Delta y} \left(\phi^n_{m+1} - \phi^n_{m-1} \right)$$

$$\phi^{n+1} = \phi^{n-1} - U \frac{\Delta t}{\Delta x} \left(\phi^n_{j+1} - \phi^n_{j-1} \right) - \Phi \frac{\Delta t}{\Delta x} \left(u^n_{j+1} - u^n_{j-1} \right)$$

$$- \Phi \frac{\Delta t}{\Delta y} \left(v^n_{m+1} - v^n_{m-1} \right)$$

where $y = m\Delta y$.

Though the y varying part of the solution is not strictly correct, we will use a solution of the form

$$v = \exp \left\{ \frac{-\beta (m\Delta y)^2}{2\sqrt{\Phi}} + i \ (kj\Delta x - \omega n\Delta t) \right\}$$

$$u = Av \qquad \text{and} \qquad \phi = Bv$$

where the arbitrary coefficients, A and B will be eliminated. The correct solution form can be numerically found through an eigenvector analysis, this will be sketched later. Insert the solution forms into the explicit formulation to obtain:

$$- 2iA \sin \omega\Delta t = - 2iA \ U \frac{\Delta t}{\Delta x} \sin k\Delta x + 2\beta m \Delta y \Delta t$$

$$- \frac{\Delta t}{\Delta x} 2i \ B \sin k\Delta x \qquad (A2.13)$$

$$-2i \sin \omega\Delta t = - 2i \ U \frac{\Delta t}{\Delta x} \sin k\Delta x - 2\beta m \Delta y \Delta t \ A$$

$$- \frac{\Delta t}{\Delta y} \left(B_{m+1} \exp \ (C(m+1)^2) \right.$$

$$- \frac{\Delta t}{\Delta y} \left(B_{m+1} \exp \ (C(m+1)^2) \right.$$

$$\left. - B_{m-1} \exp \left[C(m-1)^2 \right] \right] e^{-Cm^2} \qquad (A2.14)$$

$$- 2i \ B \ \sin \omega \Delta t \ = \ - \ 2i \ U \ \frac{\Delta t}{\Delta x} \ B \ \sin k\Delta x \ - \ 2i \ \Phi \ \frac{\Delta t}{\Delta x} \ A \ \sin k\Delta x$$

$$- \ \Phi \ \frac{\Delta t}{\Delta y} \ \left[\exp \left(C(m+1)^2 \right) \ - \ \exp \left(C(m-1)^2 \right) \right] \ e^{-Cm^2}$$

$$(A2.15)$$

where

$$C \ \equiv \ \frac{- \ \beta \Delta y^2}{2 \ \sqrt{\Phi}}$$

If we define:

$$E \ \equiv \ \frac{\Delta t}{\Delta x} \ \sin k\Delta x$$

$$D \ \equiv \ \sin \omega \Delta t \ - \ UE$$

$$F' \ \equiv \ \beta \Delta y \Delta t, \qquad\qquad F \ = \ F'm$$

$$G \ \equiv \ 2e^C \ \sinh (2 \ cm)$$

then (A2.13) and (A2.15) can be written

$$- \ 2i \ AD \ = \ 2F \ - \ 2iBE$$

$$- \ 2i \ BD \ = \ - \ 2i \ \Phi \ EA \ - \ \Phi \ \frac{\Delta t}{\Delta y} \ G$$

Solving for A and B, one obtains

$$A \ = \ \frac{i \ \left[FD \ - \ E\Phi \ \frac{\Delta t}{2\Delta y} \ G \right]}{D^2 \ - \ \Phi E^2}$$

$$B \ = \ \frac{i \ \left[\Phi EF \ - \ \Phi D \ \frac{\Delta t}{2\Delta y} \ G \right]}{D^2 \ - \ \Phi E^2}$$

Insert these relations into (A2.14), and after some manipu-
lation derive the cubic equation in D:

$$D^3 + \left[- \Phi E^2 - F^2 - (1 - e^{4C} \cosh (4 \ cm)) \ \Phi \ \frac{\Delta t^2}{2\Delta y^2} \right] D$$

$$- \ F' \ \Phi E \ \frac{\Delta t}{\Delta y} \ e^C \ \cosh (2 \ cm) \ = \ 0 \qquad\qquad (A2.16)$$

which is of the form

$$D^3 + aD + b = 0$$

If the quantity

$$\frac{a^3}{27} + \frac{b^2}{4} < 0$$

then the roots are given by

$$r_1 = 2 (- a/3)^{1/2} \cos (\phi/3)$$

$$r_2 = 2 (- a/3)^{1/2} \cos (\phi/3 + 2\pi/3) \qquad \text{(A2.17)}$$

$$r_3 = 2 (- a/3)^{1/2} \cos (\phi/3 + 4\pi/3)$$

where

$$\phi \equiv \text{arc cos} (- b/(2(- a^3/27)^{1/2}))$$

The relations (A2.17) are used to determine the phase frequencies, through the relation

$$\omega_i = \frac{1}{\Delta t} \text{ arc sin} \left\{ UE + r_i \right\} \qquad \text{(A2.18)}$$

Matrix Formulation

As an alternative, and more rigorous, method, the frequencies can be found from the eigenvector. One proceeds in an analogous fashion as for the continuous system of equations by first keeping the y variation arbitrary. Thus the explicit formulation is written:

$$- 2i \; Du_m = 2FV_m - 2iE\phi_m$$

$$- 2i \; Du_m = 2FV_m - 2iE\phi_m$$

$$- 2i \; D\phi_m = - 2i \; \Phi E \; u_m - \Phi \frac{\Delta t}{\Delta y} (v_{m+1} - v_{m-1})$$

for the x momentum and continuity equations.

Eliminate for u_m and ϕ_m:

$$\phi_m = \frac{2F\phi EV_m - D\Phi \frac{\Delta t}{\Delta y} (v_{m+1} - v_{m-1})}{2i (D^2 - \Phi E^2)}$$

$$u_m = \frac{2FDV_m - \Phi E \frac{\Delta t}{\Delta y} (v_{m+1} - v_{m-1})}{2i (D^2 - \Phi E^2)}$$

Insert these into the y momentum equation

$$2i\ Dv_m = 2\ Fu_m + \frac{\Delta t}{\Delta y} (\phi_{m+1} - \phi_{m-1})$$

to obtain:

$$-4v_m D (D^2 - \Phi E^2) = 2F \left[2FDv_m - \Phi E \frac{\Delta t}{\Delta y} (v_{m+1} - v_{m-1})\right]$$

$$+ \frac{\Delta t}{\Delta y} \left[2F\Phi E (v_{m+1} - v_{m-1}) - \Phi D \frac{\Delta t}{\Delta y} (v_{m+2} + v_{m-2} - 2v_m)\right]$$

which is a pentdiagonal system of the form:

$$Av_{m+2} + Bv_m + Av_{m-2} = 0$$

where

$$A = \Phi D \frac{\Delta t^2}{\Delta y^2}$$

$$B = -4D (D^2 - \Phi E^2 - F^2) + 2\Phi D \frac{\Delta t^2}{\Delta y^2}$$

away from the boundaries in y.

From this matrix the eigenvalues are determined and the frequencies derived.

To clarify one might define a matrix N which con-

the frequencies derived.

To clarify, one might define a matrix, N, which con-tains no D's as

$$M + P(D) = N$$

where P is a function of D, for example a cubic. N will
contain known quantities (ω, contained only in D, is the
sole unknown). Then P(D) = eigenvalues of N, and the
frequencies can be found. Note that if the eigenvalues
must be found numerically, it is not necessarily clear
which frequency of the numerical system corresponds to a
frequency of the continuous system.

REFERENCES

Baer, F., and T. J. Simmons, 1970: Computational stability
 and time truncation of coupled nonlinear equations
 with exact solutions. Mon. Wea. Rev., 98, 665-679.

Blumen, W., 1972: Geostrophic adjustemt. Rev. Geophys.,
 10, 485-528.

Charney, J. G., R. Fjörtoft, and J. von Neumann, 1950:
 Numerical integration of the barotropic vorticity
 equation. Tellus, 2, 237-254.

Courant, R., K. O. Friedrichs, and H. Lewy, 1928: Uber de
 Partiellen Differenzen gleichurgen der Mathemati-
 schen Physik. Math. Annal., 100, 32-74. Translated,
 1967, IBM Jour., 215-234.

Elvius, T., and A. Sundström, 1973: Computationally effici-
 ent schemes and boundary conditions for a fine-mesh
 barotropic model based on the shallow-water equations.
 Tellus, 25, 132-156.

Gates, W. L., 1959: On the truncation error, stability, and
 convergence of difference solutions of the barotropic
 vorticity equation. J. Meteor., 16, 556-568.

_____, and C. A. Riegel, 1962: A study of numerical
 errors in the integration of barotropic flow on a
 sperical grid. J. Geophys. Res., 67, 773-784.

Grammeltvedt, A., 1969: A survey of finite difference
 schemes for the primitive equations for a baro-
 tropic fluid. Mon. Wea. Rev., 97, 384-404.

Kurihara, Y., 1965: On the use of implicit and iterative
 methods for the time integration of the wave equation.
 Mon. Wea. Rev., 93, 33-46.

Kwizak, M., 1970: Semi-implicit integration of a grid point
 model of the primitive equations. Pub. in Meteorol-
 ogy No. 98, Arctic Met. Res. Group, McGill Univ.

 model of the primitive equations. Pub. in Meteorol-
 ogy No. 98, Arctic Met. Res. Group, McGill Univ.

Kreiss, H., and J. Oliger, 1972: Comparison of accurate
    methods for the integration of hyperbolic equations.
    Tellus, 24, 199-215.

Lindzen, R. D., 1967: Planetary waves on beta-planes.
    Mon. Wea. Rev., 95, 441-451.

Matsuno, T., 1966: Quasi-geostrophic motions in the
    equatorial area. J. Meteor. Soc. Japan, 44, 25-43.

Miles, J. W., 1974: On Laplace's tidal equations. J. Fluid
    Mech., 66, 241-260.

Orzag, S. A., 1971: Numerical simulation of incompressible
    flows within simple boundaries: Accuracy. J. Fluid
    Mech., 49, 75-112.

Whitham, G. B., 1974: Linear and nonlinear waves. John
    Wiley and Sons, 636 pp.

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

VITA

Richard Grotjahn was born September 1, 1951 in Seattle, Washington. In June 1973 he graduated from the University of Washington with a B.S. degree in Atmospheric Sciences. He began his graduate study in the Department of Meteorology at the Florida State University in September 1973. He was a participant in the 1974 Summer Fellowship Program in Scientific Computing at the National Center for Atmospheric Research.