JOURNAL OF COMPUTATIONAL PHYSICS **135,** 30–53 (1997)
ARTICLE NO. CP975733

Barotropic-Baroclinic Time Splitting for Ocean Circulation Modeling

Robert L. Higdon* and Roland A. de Szoeke†

*Department of Mathematics and †College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331
E-mail: higdon@math.orst.edu

Received July 22, 1996; revised April 29, 1997

Numerical models of ocean circulation admit motions varying on a wide range of time scales. These motions include fast external gravity waves, which are approximately independent of depth, and slower internal motions which are fully three-dimensional. Explicit time discretizations are impractical for these systems, due to the short timesteps dictated by the fast waves. A commonly used alternative is to confine the fast waves to a two-dimensional system, via vertical averaging, and then to compute the remaining motions explicitly with a long time step. However, this procedure can lead to numerical instability if the latter system admits sufficiently large residual fast motions due to an inexact splitting. In this paper we modify a method developed by R. Bleck and L. T. Smith (*J. Geophys. Res. C* **95,** 3273, 1990) in order to obtain a more precise splitting into fast and slow subsystems. In the vertically averaged momentum equation, we use the exact vertical average of the horizontal pressure gradient in place of the approximation used in *op cit.* We then apply natural time discretizations and show that the modified splitting produces considerable improvements in stability.

1. INTRODUCTION

The general circulation of the ocean is influenced by motions that vary on a wide range of time scales. The major current systems typically exhibit fluid velocities on the order of one meter per second; in other regions these velocities are generally much smaller. Other prominent motions include external gravity waves and internal gravity waves. External waves can propagate with velocities of up to hundreds of meters per second, whereas internal waves have velocities on the order of one meter per second or less. In the case of an external wave, the disturbance is approximately independent of the vertical coordinate, and the motion is manifested by the movement of the free surface at the top of the fluid. On the other hand, internal waves can be regarded as undulations of surfaces of constant density within the fluid.

The vast disparity of time scales poses major challenges when ocean circulation is modeled numerically. If an explicit time discretization is used, then the maximum permissible time step is far smaller than would be the case if the external waves were not present. The restriction on the time step can be removed by using an implicit method [11]; this approach requires that a large system of algebraic equations be solved at each time step. A commonly used alternative is to split the governing equations into subsystems that model the fast and slow motions separately; these will be referred to as the barotropic and baroclinic systems, respectively. Such splittings are used, for example, in [1, 5, 9]. Because the fast motions are approximately independent of z, the barotropic system is two-dimensional and generally resembles the shallow water equations that model the motion of a fluid of constant density. This system can be solved explicitly with short time steps or implicitly with long time steps; in this case, the costs of such methods are not great, due to the restriction to two space dimensions. The baroclinic system is fully three-dimensional and is solved explicitly with a long time step that is appropriate for resolving the slow internal motions. However, if the splitting is inexact, then the baroclinic system can actually admit some fast motions. If the fast motions are present in the baroclinic equations to a sufficiently large degree, then the computational algorithm could be unstable when long time steps are used to solve those equations.

In the present paper we examine the splitting problem in the context of isopycnal coordinates. Here, the vertical coordinate is the specific volume (reciprocal of density) or some other thermodynamic quantity. Surfaces of constant vertical coordinate are approximately material surfaces, so water masses are tracked automatically by the choice of coordinate system. This setting allows for the convenient modeling of the subtle exchanges that occur between fluid layers. Further discussions of the advantages of isopycnal coordinates are given in [1, 2, 8].

Perhaps the first barotropic-baroclinic splitting for isopycnal coordinates was the one developed by Bleck and Smith [1] for the Miami Isopycnic Coordinate Ocean Model (MICOM). Higdon and Bennett [8] recently showed that this splitting yields unstable computational algorithms when applied to a linearized two-layer model with one horizontal dimension. This analysis includes several differ-

0021-9991/97 $25.00

ent timestepping schemes for solving the baroclinic equations. In [8], the instability is attributed to the inexactness in the splitting. Three particular sources of inexactness are the decomposition $\mathbf{u} = \bar{\mathbf{u}} + \mathbf{u}'$ of the velocity field, the decomposition $p = (1 + \eta)p'$ of the pressure field, and the splitting of the term $-\nabla M$ in the momentum equation. The quantities $\bar{\mathbf{u}}$, $\mathbf{u}'$, p', and η are defined in Section 2.1 of the present paper, and ∇M is the horizontal gradient of the Montgomery potential; in an isopycnal setting, $-\nabla M$ provides the pressure forcing in the momentum equation. However, the analysis in [8] does not assess the relative importance of the above sources of inexactness, and it does not suggest an alternate splitting.

In the present paper we find that the primary difficulty lies with the splitting of $-\nabla M$. In a controlled experiment, we retain the same splitting of the pressure and velocity fields and adopt a more precise splitting of the momentum equation. We then find that the stability problem is essentially resolved. The main step is the construction of the barotropic momentum equation, which is based on a vertical average of the three-dimensional momentum equation. In the splitting developed in [1], this equation includes a horizontal pressure gradient that is equivalent to the one used in the constant-density shallow water equations. In effect, the barotropic momentum equation in [1] neglects the variations in density over the depth of the fluid. Our proposed alternative is to incorporate the vertical structure more explicitly by using the exact vertical average of $-\nabla M$. The gradient term used in [1] is an approximation to this vertical average, but we find that the difference between the two has a major effect on the stability of the coupled barotropic-baroclinic algorithm.

In Section 2 we develop the new splitting in the context of the nonlinear primitive equations. In Section 3 we analyze the stability of this splitting when applied to a linearized flow in a two-layer fluid with one horizontal dimension and a flat lower boundary. In Section 4 we extend this analysis to the case of two horizontal dimensions in a rotating reference frame with constant Coriolis parameter. Some numerical tests of the new splitting are described in Section 5. A summary is given in Section 6.

2. DESCRIPTION OF THE SPLITTING

In this section we state the system of governing equations and describe our procedure for splitting these equations into barotropic and baroclinic subsystems. We consider the primitive equations in the form

$$\mathbf{u}_t + (\mathbf{u} \cdot \nabla)\mathbf{u} + f\mathbf{k} \times \mathbf{u} = -\nabla M \tag{2.1a}$$

$$M_\alpha = p \tag{2.1b}$$

$$p_{\alpha t} + \nabla \cdot (\mathbf{u} p_\alpha) = 0 \tag{2.1c}$$

[1, 8]. Here, the vertical coordinate is taken to be the specific volume α, or reciprocal of density; the use of this coordinate requires an assumption that α is an increasing function of height. The quantity $\mathbf{u}(x, y, \alpha, t) = (u(x, y, \alpha, t), v(x, y, \alpha, t))$ is the horizontal velocity field; $p(x, y, \alpha, t)$ is the pressure;

$$M(x, y, \alpha, t) = \alpha p(x, y, \alpha, t) + g z(x, y, \alpha, t) \tag{2.2}$$

is the Montgomery potential; f is the Coriolis parameter; z is the elevation corresponding to specific volume α; g is the acceleration due to gravity; and $\nabla = (\partial/\partial x, \partial/\partial y)$ for fixed α. The notation $\mathbf{k} \times \mathbf{u}$ refers to the first two components of $(0, 0, 1) \times (u, v, 0)$, namely $(-v, u)$.

Equation (2.1b) relies on the assumption that the fluid is in hydrostatic balance; that is, $p_z = -g\alpha^{-1}$, or $p_\alpha = -g\alpha^{-1} z_\alpha$. In the system (2.1) it is assumed that $\varpi \equiv D\alpha/Dt = 0$, so that the density of each fluid particle remains constant with time. This assumption neglects effects such as sources and sinks of heat and the diffusion of temperature and salinity; if the case $\varpi \neq 0$ were to be allowed, then the terms $\varpi \mathbf{u}_\alpha$ and $(\varpi p_\alpha)_\alpha$ would appear on the left sides of (2.1a) and (2.1c), respectively. In a realistic ocean model one might use a different thermodynamic quantity, such as potential density or entropy, for a vertical coordinate; and the governing equations might contain additional terms to account for processes related to salinity and thermodynamics (Davis [4]). However, the system (2.1) is suitable for illustrating the ideas underlying a barotropic-baroclinic modal splitting.

2.1. *Barotropic and Baroclinic Equations*

A major goal of this paper is to develop a new treatment of the barotropic momentum equation, which is essentially a weighted vertical average of the depth-dependent momentum equation (2.1a). For a velocity field in the barotropic system, we use the vertically averaged velocity

$$\bar{\mathbf{u}}(x, y, t) = \frac{1}{p_b(x, y, t)} \int_{\alpha_{\text{bot}}}^{\alpha_{\text{top}}} \mathbf{u}(x, y, \alpha, t)(-p_\alpha)\, d\alpha \tag{2.3}$$

used by Bleck and Smith [1]. Here $p_b(x, y, t)$ denotes the pressure at the bottom of the fluid, and the pressure at the top of the fluid is assumed zero. The averaging defined in (2.3) approximates the projection onto the external mode in the case of a linearized problem for which modal representations of the solution are possible (Higdon and Bennett [8]). However, the linearization of (2.3) does not yield an exact projection in that case.

A vertical averaging of (2.1a) yields the barotropic momentum equation

$$\bar{\mathbf{u}}_t + f\mathbf{k} \times \bar{\mathbf{u}} = -\overline{\nabla M} + \mathbf{G}, \tag{2.4}$$

where $\overline{\nabla M}$ is defined in analogy to $\bar{\mathbf{u}}$, and $\mathbf{G}(x, y, t)$ is a residual term that includes the vertical average of the nonlinear terms. The nonlinear terms are not treated explicitly because particle velocities are far smaller than the velocities of external waves in oceanic flows, so nonlinear effects are unlikely to be significant in the modeling of external motions. The quantity $\mathbf{G}$ also includes a term involving $\mathbf{u}$ and $p_{\alpha t}$, as the time derivative does not commute with the vertical averaging, in general. A similar remark applies to the horizontal gradient ∇, so that $\overline{\nabla M} \neq \nabla(\overline{M})$.

A baroclinic momentum equation is obtained by subtracting (2.4) from (2.1a) to yield

$$\mathbf{u}'_t + (\mathbf{u} \cdot \nabla)\mathbf{u} + f\mathbf{k} \times \mathbf{u}' = -(\nabla M - \overline{\nabla M}) - \mathbf{G}, \quad (2.5a)$$

where

$$\mathbf{u}'(x, y, \alpha, t) = \mathbf{u}(x, y, \alpha, t) - \bar{\mathbf{u}}(x, y, t). \quad (2.5b)$$

The relation (2.5b) implies that the vertical average of $\mathbf{u}'$ is zero. During numerical implementation, the enforcement of this condition will provide values of the residual term $\mathbf{G}$, which will in turn be used to force the barotropic momentum equation (2.4).

In the case of a fluid with constant density, the hydrostatic relation $p_z = -g\alpha^{-1}$ implies that the Montgomery potential $M = \alpha p + gz$ is independent of depth. In that case $M(x, y, t) = M_{\text{top}} = gz_{\text{top}}$, so $-\nabla M$ is exactly the gradient term that appears in the constant-density shallow water equations (Pedlosky [12]). The vertically averaged equation (2.4) thus generalizes the momentum equation in the shallow water system to the case of external wave motions in the variable-density system (2.1).

Bleck and Smith [1] introduced a decomposition of the pressure field given by

$$p(x, y, \alpha, t) = (1 + \eta(x, y, t))p'(x, y, \alpha, t), \quad (2.6)$$

where p' is intended to represent the pressure field due to static effects and slow internal motions, while η is a dimensionless quantity that represents the fast external signal. The quantity p'_b at the bottom of the fluid is assumed to be independent of time. Equation (2.6) is based on the idea that an external signal causes the thickness of each fluid layer to change by approximately the same proportion [1, 8]. Typically $|\eta| \ll 1$. The preceding remarks do not define p' and η uniquely; one can specify these quantities by defining $p'_b(x, y)$ to be the pressure at the bottom of the fluid when the system is in some reference state, such as an equilibrium state, and then letting

$$\eta(x, y, t) = \frac{p_b(x, y, t)}{p'_b(x, y)} - 1, \quad (2.7)$$

so that η represents the relative perturbation in bottom pressure. The quantity $p'(x, y, \alpha, t)$ is then defined by (2.6). In [1] the pressure decomposition (2.6) is used to split the continuity equation (2.1c); the barotropic continuity equation is

$$p'_b \eta_t + \nabla \cdot (p'_b \bar{\mathbf{u}}) = 0, \quad (2.8)$$

and the baroclinic continuity equation is

$$p'_{\alpha t} + \nabla \cdot (\mathbf{u}' p'_\alpha) = -\nabla \cdot (\bar{\mathbf{u}} p'_\alpha) + \frac{p'_\alpha}{p'_b} \nabla \cdot (p'_b \bar{\mathbf{u}}). \quad (2.9)$$

In the derivation of (2.8) and (2.9), the quantity $1 + \eta$ is approximated by 1 whenever $1 + \eta$ appears as a factor. Also see [8].

In the barotropic-baroclinic splitting developed in [1], the barotropic momentum equation has the form (2.4), except that the term $-\overline{\nabla M}$ is approximated by $-\alpha_0 \nabla (p'_b \eta)$, where α_0 is a representative value of α. The quantity $p'_b \eta$ is the perturbation in bottom pressure; the term $-\alpha_0 \nabla (p'_b \eta)$ is thus equivalent to the horizontal pressure gradient that would be used for a hydrostatic fluid having constant density α_0^{-1}. The difference of the quantities $-\alpha_0 \nabla (p'_b \eta)$ and $-\overline{\nabla M}$ includes terms involving the vertical structure of the fluid; see (2.12)–(2.13), (3.6).

An analogy can be made with the barotropic-baroclinic splitting used in the Bryan–Cox class of z-coordinate ocean models (e.g., [3, 5, 9, 13]). There, the decomposition of the pressure field is given by the additive relation $p(x, y, z, t) = \rho_0 g \eta(x, y, t) + p_h(x, y, z, t)$, where ρ_0 is a reference density, η is the perturbation in the elevation in the free surface, and p_h is the pressure relative to the mean free surface. To a close approximation, the term $\rho_0 g \eta(x, y, t)$ represents the effects of external gravity waves, and the term $p_h(x, y, z, t)$ represents the effects of internal motions. These terms are thus regarded to vary on the fast and slow time scales, respectively. In the Bryan–Cox class of models, the horizontal pressure forcing is represented by the Boussinesq approximation $-\rho_0^{-1} \nabla p$. The barotropic momentum equation is based on a vertical average; this equation thus includes a term $-g\nabla\eta$, which is exact for a hydrostatic fluid of constant density, plus some terms that represent the vertical structure of the fluid. The latter terms vary on the slow time scale. However, as seen in the following section, the analogous terms for the isopycnal case can include quantities varying on both the fast and slow time scales.

In the additive pressure splitting $p(x, y, z, t) = \rho_0 g \eta (x, y, t) + p_h(x, y, z, t)$, the effects of the fast external motions are represented in terms of the elevation of the

free surface and are thus regarded as independent of z. This statement is highly accurate. However, the situation is different if the pressure is represented in isopycnal coordinates as $p(x, y, \alpha, t)$. A point of constant (x, y, α) is a material point that moves up and down with the motions of the fluid; for external motions the pressure at that point is approximately due to the vertical distance from the free surface, so variations in this pressure are due to the thickening or thinning of the fluid layer above (x, y, α), not the variation of the free surface elevation relative to its mean. In this case the multiplicative decomposition (2.6) of $p(x, y, \alpha, t)$ is appropriate, as described earlier.

The stability of the barotropic-baroclinic splitting for the Bryan–Cox class of models has apparently not been analyzed, but no stability problems with this splitting have been reported (John Dukowicz, personal communication). In the present paper we analyze the usage of exact vertical averaging in the case of isopycnal coordinates, and we show that this produces major improvements in stability compared to using the gradient term from the constant-density shallow water equations.

2.2 *Vertical Discretization; Representation of* $\overline{\nabla M}$

We next describe a discretization with respect to the vertical coordinate α. In that context, we develop a representation of $\overline{\nabla M}$ that illustrates how $\overline{\nabla M}$ includes quantities varying on the fast time scale and some other quantities varying on the slow time scale. In the analysis given later, the fast and slow terms are handled separately.

In the following description of vertical discretization, the formulas can be regarded either as approximations for a continuously stratified fluid obtained from finite difference approximations with respect to α, or as exact statements about a discretely stratified fluid consisting of a stack of homogeneous fluid layers (Higdon and Bennett [8]). We use terminology motivated by the latter viewpoint.

Suppose that the fluid consists of R layers, and let α_r denote the specific volume of layer r, where the index $r = 1$ refers to the uppermost layer and the index $r = R$ refers to the bottom layer. Let $p_r(x, y, t)$ denote the pressure at the bottom of layer r, and denote the pressure at the free surface by $p_0 = 0$. The quantity $\Delta p_r = p_r - p_{r-1}$ then denotes the pressure increment across layer r and can be regarded as a measure of layer thickness. The baroclinic quantity p_r' is defined via the pressure splitting $p_r(x, y, t) = (1 + \eta(x, y, t))p_r'(x, y, t)$. The increment in α across the rth interface will be denoted by $\Delta\alpha_r = \alpha_r - \alpha_{r+1}$. Let $M_r(x, y, t)$ and $\mathbf{u}_r(x, y, t)$ denote the Montgomery potential and horizontal velocity field, respectively, in layer r. In a layer of constant density, the Montgomery potential is independent of depth; for convenience we assume that this is also the case for the horizontal velocity. At interface r the Montgomery potential satisfies the jump condition

$$M_r = M_{r+1} + p_r \Delta\alpha_r; \tag{2.10}$$

this is a discrete analogue of (2.1b).

In this context, the vertical average of ∇M is given by

$$\overline{\nabla M}(x, y, t) = \sum_{r=1}^{R} \frac{\Delta p_r}{p_R} \nabla M_r, \tag{2.11}$$

which is a discrete analogue of the averaging by integration discussed in Section 2.1. A barotropic velocity $\bar{\mathbf{u}}$ is defined similarly. The weighting coefficients satisfy the condition $\sum_{r=1}^{R} \Delta p_r / p_R = 1$, since $p_0 = 0$. Various explicit formulas for (2.11) are possible; the following is one example.

Recursive application of the interface condition (2.10) yields an expression for M_r in terms of the Montgomery potential M_R in the bottom layer; we then obtain

$$\nabla M_r = \nabla M_R + \Delta\alpha_{R-1} \nabla p_{r-1} + \cdots + \Delta\alpha_r \nabla p_r.$$

When this result is inserted into (2.11), rearrange the sum so that the terms corresponding to a given layer are grouped together. The result is

$$\overline{\nabla M} = \nabla M_R + \sum_{r=1}^{R-1} \left[\Delta\alpha_r (\nabla p_r) \sum_{j=1}^{r} \frac{\Delta p_j}{p_R} \right].$$

But $\sum_{j=1}^{r} \Delta p_j = p_r$, so

$$\overline{\nabla M} = \nabla M_R + \frac{1}{p_R} \sum_{r=1}^{R-1} \Delta\alpha_r \nabla \left(\frac{1}{2} p_r^2 \right).$$

The pressure decomposition $p_r = (1 + \eta)p_r'$ then yields

$$\begin{aligned} \overline{\nabla M} = {} & \nabla(\alpha_R p_R' \eta) + \nabla(\alpha_R p_R' + g z_{\text{bot}}) \\ & + \frac{1}{(1 + \eta) p_R'} \nabla[(p_R')^2 (1 + \eta)^2 Q], \end{aligned} \tag{2.12}$$

where

$$Q(x, y, t) = \frac{1}{2} \sum_{r=1}^{R-1} \left(\frac{p_r'}{p_R'} \right)^2 \Delta\alpha_r. \tag{2.13}$$

The quantity η varies on the fast barotropic time scale, the quantity Q varies on the slow baroclinic time scale, and the remaining quantities are stationary.

2.3. *A Time Discretization*

We next consider discretization with respect to time. In this discussion, a significant issue is the implementation of the residual term $\mathbf{G}$ in (2.4)–(2.5). For the baroclinic equations, we describe a timestepping scheme based on the forward–backward method. When applied to the purely baroclinic equations in the linearized case, this method is stable and nondissipative, provided that the Courant–Friedrichs–Lewy condition is satisfied. The subsequent stability analysis is based mainly on this method. However, the leapfrog scheme can also be used in a similar manner, and a stability analysis using that method is included in the later discussion.

We assume that the solution is known at time t_n, and we wish to compute the solution at time $t_{n+1} = t_n + \Delta t$, where Δt is a long time step that is appropriate for solving the slow baroclinic equations explicitly. Over the time interval $t_n \le t \le t_{n+1}$, the fast barotropic equations could be solved implicitly with time step Δt, or they could be solved explicitly by subcycling through many small substeps of the interval $t_n \le t \le t_{n+1}$.

In the latter case, the term $\overline{\nabla M}$ can be implemented as follows. First suppose that the baroclinic continuity equation (2.9) has already been updated, so that the values of p' at time t_{n+1} are available. Baroclinic quantities appearing in $\overline{\nabla M}$ can then be interpolated (e.g., linearly) with respect to t between times t_n and t_{n+1}. On the other hand, barotropic quantities are updated at each substep. For the sake of efficiency, one does *not* compute a vertical average of the three-dimensional field ∇M at each barotropic substep; instead, an explicit representation of $\overline{\nabla M}$ enables one to work strictly with two-dimensional fields during the subcycling.

We next consider the baroclinic equations. For a vertically discrete medium, the baroclinic continuity equation in layer r has the form

$$\frac{\partial \Delta p'_r}{\partial t} + \nabla \cdot (\mathbf{u}_r \Delta p'_r) = \frac{\Delta p'_r}{p'_R} \nabla \cdot (p'_R \overline{\mathbf{u}}) \tag{2.14}$$

for $1 \le r \le R$, and the baroclinic momentum equation is

$$\frac{\partial \mathbf{u}'_r}{\partial t} + (\mathbf{u}_r \cdot \nabla)\mathbf{u}_r + f\mathbf{k} \times \mathbf{u}'_r = -(\nabla M_r - \overline{\nabla M}) - \mathbf{G}. \tag{2.15}$$

Here $\mathbf{u}'_r(x, y, t) = \mathbf{u}_r - \overline{\mathbf{u}}$, in analogy with (2.5b), and $\Delta p'_r = p'_r - p'_{r-1}$. In the momentum equation, the quantity $\mathbf{G}$ will be regarded as a forcing term that drives the baroclinic velocity field so that it has vertical mean zero, in the sense that

$$\sum_{r=1}^{R} \frac{\Delta p_r}{p_R} \mathbf{u}'_r = \sum_{r=1}^{R} \frac{\Delta p'_r(x, y, t)}{p'_R(x, y)} \mathbf{u}'_r(x, y, t) = 0. \tag{2.16}$$

In the following discussion of time discretization, we do not indicate discretizations with respect to the horizontal space variables, as the emphasis is on the time stepping. Superscripts are used to denote the time index.

First update the continuity equation (2.14) with the forward step

$$\Delta p'^{\,n+1}_r = \Delta p'^{\,n}_r + \Delta t \left(-\nabla \cdot (\mathbf{u}^n_r \Delta p'^{\,n}_r) + \frac{\Delta p'^{\,n}_r}{p'_R} \nabla \cdot (p'_R \overline{\mathbf{u}}^n) \right). \tag{2.17}$$

Then use the values of p' at time t_{n+1} to solve the barotropic equations, as described earlier. The barotropic momentum equation (2.4) includes the term $\mathbf{G}$; when the barotropic equations are solved for $t_n < t \le t_{n+1}$, we use the quantity $\mathbf{G}^n(x, y)$ that was obtained when the baroclinic momentum equation was advanced from time t_{n-1} to time t_n. If the barotropic equations are solved explicitly via subcycling, then the quantity $\mathbf{G}$ is held constant over the barotropic substeps. The computation of $\mathbf{G}$ is described below.

Before updating the momentum equation (2.15), we calculate M_r at time t_{n+1} for $1 \le r \le R$. For the bottom layer, we have $M^{n+1}_R = \alpha_R(1 + \eta^{n+1})p'_R + gz_{\text{bot}}$. Values of M in the other layers can be obtained by repeated application of the jump condition (2.10); this process uses the quantities $p^{n+1}_r = (1 + \eta^{n+1})p'^{n+1}_r$.

When the momentum equation (2.15) is updated, the Coriolis terms and nonlinear terms can be handled by the following predictor–corrector process, which is a generalization of the classical Heun method [10, 14]. In Section 4 we show that this approach provides for a stable treatment of the Coriolis terms in the linearized case. A predicted velocity is defined by

$$\begin{aligned} \mathbf{u}'^{\,\text{pred}}_r = \mathbf{u}'^{\,n}_r + \Delta t \,(&-(\mathbf{u}^n_r \cdot \nabla)\mathbf{u}^n_r - f\mathbf{k} \times \mathbf{u}'^{\,n}_r \\ &- (\nabla M^{n+1}_r - \overline{\nabla M}^{n+1})) - \mathbf{G}^{\text{pred}} \Delta t. \end{aligned} \tag{2.18}$$

Here, we use updated values involving M, but in the nonlinear and Coriolis terms we use old values of velocity. The quantity $\mathbf{G}^{\text{pred}}$ is not yet defined; this quantity is specified by invoking the zero-mean condition (2.16) to obtain

$$\begin{aligned} 0 &= \sum_{r=1}^{R} \frac{\Delta p'^{\,n+1}_r}{p'_R} [\mathbf{u}'^{\,\text{pred}}_r] \\ &= \sum_{r=1}^{R} \frac{\Delta p'^{\,n+1}_r}{p'_R} [\mathbf{u}'^{\,n}_r + \Delta t \,(-(\mathbf{u}^n_r \cdot \nabla)\mathbf{u}^n_r - f\mathbf{k} \times \mathbf{u}'^{\,n}_r \\ &\quad - (\nabla M^{n+1}_r - \overline{\nabla M}^{n+1}))] - \mathbf{G}^{\text{pred}} \Delta t. \end{aligned}$$

The quantity $\mathbf{G}^{\text{pred}}\,\Delta t$ can be inserted back into (2.18) to determine the predicted velocity. An equivalent process is the following: obtain a tentative baroclinic velocity by using Eq. (2.18) without the term $\mathbf{G}^{\text{pred}}\,\Delta t$; compute the vertical mean of the result, using weighting coefficients from time t_{n+1}; and then subtract this mean from the tentative velocity to obtain a field $\mathbf{u}_r'^{\,\text{pred}}$ that satisfies the condition (2.16).

We then obtain corrected velocities by using weighted averages of the nonlinear and Coriolis terms as stated in terms of $\mathbf{u}_r'^{\,n}$ and $\mathbf{u}_r'^{\,\text{pred}}$. The residual term, denoted $\mathbf{G}^{n+1}\,\Delta t$ in the present case, is treated in the same manner as before. The correction step is then given by the relations

$$\begin{aligned}\mathbf{u}_r'^{\,\text{corr}} &= \mathbf{u}_r'^{\,n} + a\Delta t\,(-(\mathbf{u}_r^{\text{pred}}\cdot\nabla)\mathbf{u}_r^{\text{pred}} - f\mathbf{k}\times\mathbf{u}_r'^{\,\text{pred}})\\ &\quad + b\Delta t(-(\mathbf{u}_r^{n}\cdot\nabla)\mathbf{u}_r^{n} - f\mathbf{k}\times\mathbf{u}_r'^{\,n})\\ &\quad - \Delta t(\nabla M_r^{n+1} - \overline{\nabla M}^{n+1}) \quad \text{for } 1\le r\le R\\ \mathbf{G}^{n+1}\,\Delta t &= \sum_{r=1}^{R}\frac{\Delta p_r'^{\,n+1}}{p_R'}\,\mathbf{u}_r'^{\,\text{corr}}\\ \mathbf{u}_r'^{\,n+1} &= \mathbf{u}_r'^{\,\text{corr}} - \mathbf{G}^{n+1}\,\Delta t,\end{aligned} \tag{2.19}$$

where $\mathbf{u}_r^{\text{pred}} = \mathbf{u}_r'^{\,\text{pred}} + \overline{\mathbf{u}}^{n+1}$ and $a + b = 1$. The choice of parameters a and b is discussed in Section 4, along with the effects of these parameters on stability. In the first equation in (2.19), the quantities involving $\mathbf{u}_r^n$, $\mathbf{u}_r'^{\,n}$, and M were already computed for the prediction step and need not be recomputed for the correction step.

3. A LINEARIZED, TWO-LAYER, ONE-DIMENSIONAL MODEL

Here we apply the preceding strategy to a simple one-dimensional setting and analyze the stability of the resulting algorithm. In Section 4 we extend the analysis to the case of two horizontal dimensions with a rotating reference frame.

3.1. *Governing Equations*

Consider a flow in one horizontal dimension with a flat lower boundary and suppose that the flow is linearized about an equilibrium state where the velocity is zero and the pressure depends only on depth. We also suppose that the fluid consists of two layers. Let α_1 and α_2 denote the specific volumes of the upper and lower layers, respectively, with $\Delta\alpha = \alpha_1 - \alpha_2 > 0$. In the present linearized setting, $p_0 = 0$, p_1, and $p_2 = p_b$ denote the perturbations in pressure at the free surface, bottom of upper layer, and bottom of lower layer, respectively. Values of the equilibrium pressures $\tilde p_0 = 0$, $\tilde p_1$, and $\tilde p_2 = \tilde p_b$ are defined analogously. Then $\Delta p_r = p_r - p_{r-1}$ and $\Delta\tilde p_r = \tilde p_r - \tilde p_{r-1}$ denote pressure increments across layer r, for $r = 1$ and $r = 2$. Let u_r denote the horizontal velocity in layer r, and let M_r denote the perturbation in Montgomery potential in that layer. The linearization of the system (2.1), as applied to the present setting, is then

$$\frac{\partial u_r}{\partial t} = -\frac{\partial M_r}{\partial x}, \quad r = 1, 2, \tag{3.1a}$$

$$M_1 = M_2 + p_1\Delta\alpha \tag{3.1b}$$

$$\frac{\partial}{\partial t}(\Delta p_r) + \Delta\tilde p_r\frac{\partial u_r}{\partial x} = 0, \quad r = 1, 2, \tag{3.1c}$$

The boundary condition at the bottom of the fluid is $M_2 = \alpha_2 p_2$.

The linearization of the pressure splitting (2.6) of Bleck and Smith [1] can be written in the form

$$\Delta p_r = \Delta p_r' + \eta\Delta\tilde p_r, \tag{3.2}$$

or equivalently, $p_r = p_r' + \eta\tilde p_r$, with $p_2' = p_b'$ being independent of t (Higdon and Bennett [8]). In (3.2) it is assumed that $\eta = 0$ corresponds to the equilibrium state. These constraints do not determine $\Delta p_1'$, $\Delta p_2'$, and η uniquely for given values of Δp_r and $\Delta\tilde p_r$. However, unique determination is possible if, in addition, one assumes $p_2' = 0$. In that case one has

$$\eta(x, t) = p_2/\tilde p_2, \tag{3.3}$$

and η then represents the relative perturbation in the bottom pressure; this is consistent with the definition given in (2.7) for the nonlinear case.

Define a barotropic velocity $\overline{u}(x, t)$ by the weighted vertical average,

$$\overline{u}(x, t) = \frac{\Delta\tilde p_1}{\tilde p_2}u_1 + \frac{\Delta\tilde p_2}{\tilde p_2}u_2, \tag{3.4}$$

which is a linearization of the kind of averaging defined in (2.11). An average $\overline{M}(x, t)$ is defined in an analogous manner. A vertical averaging of Eq. (3.1a) then yields

$$\frac{\partial\overline{u}}{\partial t} = -\frac{\partial\overline{M}}{\partial x}. \tag{3.5}$$

Equation (3.5) is the linearization of the general barotropic momentum equation (2.4), in the case of one horizontal dimension; in the present simplified case, the residual term G is zero, and the vertical averaging operation commutes with differentiation with respect to x and t. The interface

condition (3.1b) and bottom boundary condition $M_2 = \alpha_2 p_2$ imply

$$\overline{M}(x,t) = \alpha_2 p_2 + \frac{\Delta\tilde{p}_1}{\tilde{p}_2} p_1 \Delta\alpha; \tag{3.6a}$$

the pressure splitting (3.2) and the assumption $p_2' = 0$ then yield

$$\overline{M}(x,t) = \alpha_2\tilde{p}_2\eta + \frac{\Delta\tilde{p}_1}{\tilde{p}_2}(\Delta\alpha)(\Delta p_1' + \eta\Delta\tilde{p}_1). \tag{3.6b}$$

For a barotropic momentum equation, we use Eq. (3.5) with $\overline{M}(x, t)$ given by (3.6b).

The barotropic momentum equation of Bleck and Smith [1] uses the gradient term $-\alpha_0\nabla(p_b'\eta)$ instead of $-\overline{\nabla M}$; if the parameter α_0 is chosen to be α_2, then this term becomes $-\partial/\partial x(\alpha_2\tilde{p}_2\eta)$ in the present linearized case. The term $-\partial/\partial x(\alpha_2\tilde{p}_2\eta)$ is exact for a single-layer fluid having density $1/\alpha_2$; the terms in (3.6b) that involve $\Delta\alpha$ account for the effects of density variations in a two-layer fluid.

Next consider the baroclinic momentum equation. Let $u_r' = u_r - \overline{u}$ for $r = 1, 2$. A comparison with (3.4) shows that the weighted average of u_1' and u_2' is zero, so it suffices to consider u_1' only. Subtracting (3.5) from (3.1a) yields

$$\frac{\partial u_1'}{\partial t} = -\frac{\partial}{\partial x}(M_1 - \overline{M}).$$

Equations (3.1b), (3.2), and (3.6a) then imply

$$\frac{\partial u_1'}{\partial t} = -\frac{\Delta\tilde{p}_1\Delta\tilde{p}_2\Delta\alpha}{\tilde{p}_2}\frac{\partial}{\partial x}\left(\frac{\Delta p_1'}{\Delta\tilde{p}_1} + \eta\right).$$

The linearization of the baroclinic continuity equation (2.14) is

$$\frac{\partial}{\partial t}(\Delta p_r') + \Delta\tilde{p}_r\frac{\partial u_r'}{\partial x} = 0$$

in the case of one horizontal dimension. The sum $\Delta p_1' + \Delta p_2' = p_2'$ is a known quantity, so it suffices to consider the upper layer only. The baroclinic equations can then be written in the form

$$\frac{\partial\delta}{\partial t} + \frac{\partial u_1'}{\partial x} = 0 \tag{3.8a}$$

$$\frac{\partial u_1'}{\partial t} + c_1^2\frac{\partial\delta}{\partial x} = -c_1^2\frac{\partial\eta}{\partial x}, \tag{3.8b}$$

where

$$\delta(x,t) = \frac{\Delta p_1'(x,t)}{\Delta\tilde{p}_1}$$

and

$$c_1 = \left(\frac{\Delta\tilde{p}_1\Delta\tilde{p}_2\Delta\alpha}{\tilde{p}_2}\right)^{1/2}. \tag{3.9}$$

The quantity c_1 is the speed of internal gravity waves in a fluid with two layers.

A comparison of (3.5) and (3.6b) implies that the barotropic momentum equation can be written in the form

$$\frac{\partial\overline{u}}{\partial t} + c_0^2\frac{\partial\eta}{\partial x} = -\gamma c_0^2\frac{\partial\delta}{\partial x}, \tag{3.10}$$

where

$$c_0 = \left(\alpha_2\tilde{p}_2 + \frac{(\Delta\tilde{p}_1)^2\Delta\alpha}{\tilde{p}_2}\right)^{1/2} \tag{3.11}$$

denotes the speed of external gravity waves, and

$$\gamma = \frac{\Delta\tilde{p}_1}{\Delta\tilde{p}_2}\left(\frac{c_1}{c_0}\right)^2 = O(\Delta\alpha/\alpha_2). \tag{3.12}$$

The linearization of the barotropic continuity equation (2.8) is

$$\frac{\partial\eta}{\partial t} + \frac{\partial\overline{u}}{\partial x} = 0. \tag{3.13}$$

3.2. *Computational Algorithm*

We now apply the time discretization described in Section 2.3 to the barotropic-baroclinic system (3.8), (3.10), (3.13). In the present subsection we make a precise statement of the algorithm, and in the following subsections we analyze stability.

In the stability analysis given later, we assume that the system is defined for $-\infty < x < \infty$ and $t > 0$. We then use a Fourier transform with respect to x and examine the evolution of the system through time. In the following discussion, we state the algorithm directly in terms of Fourier transforms. For definiteness, we use centered second-order finite differences on a staggered grid when discretizing with respect to space. Values of η and $\Delta p_1'$ are defined at points with spacing Δx, and grid points for $\overline{u}$ and u_1' are located halfway between the points for η and $\Delta p_1'$. Alter-

nate spatial discretizations can be considered by using different definitions of the symbol K in (3.15).

Let $\hat{\delta}^n(k)$, $\widehat{u_1'}^n(k)$, and $\hat{\eta}^n(k)$ denote the Fourier transforms of δ, u_1', and η with respect to x at time t_n, with k being the dual variable for the Fourier transform. The forward–backward method defined in (2.17)–(2.19), as applied to the linearized baroclinic system (3.8), can then be written in the form

$$\hat{\delta}^{n+1} = \hat{\delta}^n - (iKc_1\Delta t)\frac{\widehat{u_1'}^n}{c_1} \tag{3.14a}$$

$$\frac{\widehat{u_1'}^{n+1}}{c_1} = \frac{\widehat{u_1'}^n}{c_1} - (iKc_1\Delta t)(\hat{\delta}^{n+1} + \hat{\eta}^{n+1}). \tag{3.14b}$$

The action of the spatial differencing is represented by the symbol

$$iK = \frac{(e^{ik\Delta x/2} - e^{-ik\Delta x/2})}{\Delta x} = i\frac{\sin(k\Delta x/2)}{(\Delta x/2)}, \tag{3.15}$$

and the dependent variables are written in nondimensional form for the sake of later analysis. A calculation shows that if $\hat{\eta} = 0$, then the system (3.14) is stable and nondissipative, provided that the Courant–Friedrichs–Lewy condition $c_1\Delta t/\Delta x < 1$ is satisfied.

We assume that the barotropic equations (3.10), (3.13) are discretized with respect to x but are solved exactly with respect to t on the time interval $t_n \le t \le t_{n+1}$. The latter assumption is made partly for simplicity and partly in order to isolate the effects on stability of the modal splitting and baroclinic time stepping.

We consider the transformed barotropic equations in the form

$$\frac{\partial}{\partial t}\left(\frac{\hat{\bar{u}}}{c_0}\right) + (iKc_0)\hat{\eta} = -\gamma(iKc_0)\hat{\delta}(k,t) \tag{3.16a}$$

$$\frac{\partial \hat{\eta}}{\partial t} + (iKc_0)\frac{\hat{\bar{u}}}{c_0} = 0, \tag{3.16b}$$

where $\hat{\delta}(k, t)$ denotes the linear interpolant $\hat{\delta}(k, t) = \hat{\delta}^n(k) + (t - t_n)(\hat{\delta}^{n+1}(k) - \hat{\delta}^n(k))/\Delta t$. This interpolant is an analogue, in the present case, of the interpolation of baroclinic quantities discussed in Section 2.3. Addition and subtraction of (3.16a) and (3.16b) produces scalar equations for the quantities $\hat{\bar{u}}/c_0 + \hat{\eta}$ and $\hat{\bar{u}}/c_0 - \hat{\eta}$, which are solved by standard methods. Impose the initial conditions $\hat{\bar{u}} = \hat{\bar{u}}^n$ and $\hat{\eta} = \hat{\eta}^n$ at time $t = t_n$, and then let $t = t_{n+1}$ to obtain the solution at the new time level. A calculation shows that the solution of the barotropic system (3.16) at time t_{n+1} is

$$\begin{pmatrix} \hat{\bar{u}}^{n+1}/c_0 \\ \hat{\eta}^{n+1} \end{pmatrix} = Y(k\Delta x)\begin{pmatrix} \hat{\bar{u}}^n/c_0 \\ \hat{\eta}^n \end{pmatrix} + V(k\Delta x)\hat{\delta}^n + W(k\Delta x)\hat{\delta}^{n+1} \tag{3.17a}$$

where

$$Y(k\Delta x) = \begin{pmatrix} \cos\theta & -i\sin\theta \\ -i\sin\theta & \cos\theta \end{pmatrix}, \tag{3.17b}$$

V and W are the column vectors

$$\begin{aligned} V(k\Delta x) &= \gamma\begin{pmatrix} -i(\sin\theta - (1-\cos\theta)/\theta) \\ \cos\theta - (\sin\theta)/\theta \end{pmatrix} \\ W(k\Delta x) &= \gamma\begin{pmatrix} -i(1-\cos\theta)/\theta \\ (\sin\theta)/\theta - 1 \end{pmatrix}, \end{aligned} \tag{3.17c}$$

and $\theta = Kc_0\Delta t$. Recall $\gamma = O(\Delta\alpha/\alpha_2)$, from (3.12). For fixed values of the model parameters $\Delta\tilde{p}_1$, $\Delta\tilde{p}_2$, α_1, α_2, and Courant number $c_1\Delta t/\Delta x$, the quantity θ is a function of the dimensionless wavenumber $k\Delta x$ (cf. (3.9), (3.11), (3.15)).

We now combine (3.14) and (3.17) into a single vector equation. Let

$$v^n(k) = (\hat{\delta}^n \quad \widehat{u_1'}^n/c_1 \quad \hat{\bar{u}}^n/c_0 \quad \hat{\eta}^n)^{\mathrm{T}} \tag{3.18}$$

denote a column vector consisting of four unknowns. The coupled barotropic-baroclinic algorithm can then be written in the form

$$E_1 v^{n+1} = E_0 v^n, \tag{3.19a}$$

where

$$E_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ iKc_1\Delta t & 1 & 0 & iKc_1\Delta t \\ -W_1 & 0 & 1 & 0 \\ -W_2 & 0 & 0 & 1 \end{pmatrix} \tag{3.19b}$$

and

$$E_0 = \begin{pmatrix} 1 & -iKc_1\Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ V_1 & 0 & Y_{11} & Y_{12} \\ V_2 & 0 & Y_{21} & Y_{22} \end{pmatrix}. \tag{3.19c}$$

The matrices E_1 and E_0 are functions of $k\Delta x$.

3.3. *Characteristic Polynomial*

In this and the following subsection we analyze the stability of the algorithm defined by (3.19). For such an algorithm to be stable, all solutions of the system (3.19) would have to be bounded as $n \to +\infty$, for each wavenumber k. We will show that this is not the case; however, the unstable behavior is confined to wavenumbers lying in a finite number of extremely narrow intervals. Outside these intervals, the algorithm is stable and nondissipative. This represents a considerable improvement over the splitting of Bleck and Smith [1], as analyzed by Higdon and Bennett [8]; for that splitting, unstable behavior is found for all but finitely many wavenumbers, and the growth rates are larger than those found in the exceptional cases for the present splitting. In some numerical computations described in Section 5, good results are obtained with the splitting developed in the present paper, as applied to a nonlinear ocean model that includes physically appropriate dissipative mechanisms.

We begin by considering solutions of (3.19) having the form $v^n = \lambda^n q$, where λ is a complex scalar and q is a nonzero vector having four components. The superscript on λ is an exponent, and the superscript on v is a time index. The form $v^n = \lambda^n q$ is a nontrivial solution of the difference equation (3.19) if and only if λ and q satisfy the eigenvalue problem

$$\lambda E_1 q = E_0 q, \quad q \neq 0. \tag{3.20}$$

A number λ is an eigenvalue if and only if $\det(\lambda E_1 - E_0) = 0$. We thus consider the characteristic polynomial

$$p(\lambda) = \det(\lambda E_1 - E_0) = \begin{pmatrix} \lambda - 1 & iKc_1\Delta t & 0 & 0 \\ \lambda(iKc_1\Delta t) & \lambda - 1 & 0 & \lambda(iKc_1\Delta t) \\ -\lambda W_1 - V_1 & 0 & \lambda - \cos\theta & i\sin\theta \\ -\lambda W_2 - V_2 & 0 & i\sin\theta & \lambda - \cos\theta \end{pmatrix}, \tag{3.21}$$

where V_1, V_2, W_1, and W_2 denote components of the vectors V and W defined in (3.17c); $\theta = Kc_0\Delta t$, and γ and K are defined in (3.12) and (3.15), respectively.

The determinant of the 2×2 submatrix lying in the upper left corner of $\lambda E_1 - E_0$ is the characteristic polynomial for the forward–backward method (3.14) for the baroclinic equations, when the forcing term $\hat{\eta}^{n+1}$ is removed. Similarly, the determinant of the 2×2 submatrix lying in the lower right corner is associated with the barotropic system (3.16), when the forcing from baroclinic variables is removed. A calculation shows that $p(\lambda)$ consists of the product of these characteristic polynomials, plus some terms involving $\gamma = O(\Delta\alpha/\alpha_2)$ that result from the coupling between the baroclinic and barotropic subsystems. In particular, we obtain

$$p(\lambda) = (\lambda^2 - \tilde{a}\lambda + 1)(\lambda^2 - \tilde{b}\lambda + 1) + \phi(k\Delta x)\lambda(\lambda - 1)^2, \tag{3.22a}$$

where

$$\begin{aligned} \tilde{a}(k\Delta x) &= 2 - (1 - \gamma)(Kc_1\Delta t)^2 \\ &= 2 - 4(1 - \gamma)\nu^2 \sin^2(k\Delta x/2) \\ \tilde{b}(k\Delta x) &= 2\cos\theta = 2\cos(Kc_0\Delta t) \\ &= 2\cos(2\nu(c_0/c_1)\sin(k\Delta x/2)) \\ \phi(k\Delta x) &= \gamma^{3/2} \cdot 2\nu \sin(k\Delta x/2)\sin(Kc_0\Delta t)\left(\frac{\Delta\tilde{p}_2}{\Delta\tilde{p}_1}\right)^{1/2} \\ &= O(\Delta\alpha/\alpha_2)^{3/2} \end{aligned} \tag{3.22b}$$

and $\nu = c_1\Delta t/\Delta x$ denotes the Courant number. The term in $\tilde{a}$ involving γ represents some of the effects of the coupling, and the remainder of the coupling is represented by the term involving $\phi(k\Delta x)$. The polynomial $\lambda^2 - \tilde{a}\lambda + 1$, except for the term involving γ, is the characteristic polynomial for the purely baroclinic equations, and the polynomial $\lambda^2 - \tilde{b}\lambda + 1$ is associated with the barotropic equations. The roots of these quadratic polynomials indicate the time dependence, in the form λ^n, of the baroclinic and barotropic subsystems, respectively. For convenience, we do not denote explicitly the dependence of $p(\lambda)$ on $k\Delta x$. Because of periodicity, we can restrict attention to the case $|k\Delta x| \leq \pi$.

If, for some $k\Delta x$, the polynomial p in (3.22a) has a root λ with $|\lambda| > 1$, then the difference equation (3.19) admits solutions of the form $v^n = \lambda^n q$ that are unbounded as $n \to +\infty$. On the other hand, if $|\lambda| \leq 1$ for each root λ, and if linear independence of eigenvectors implies that all solutions of (3.19) are linear combinations of solutions of the form $\lambda^n q$, then all solutions of (3.19) are bounded as $n \to +\infty$.

Higdon and Bennet [8] analyzed the splitting of Bleck and Smith [1] in the same setting considered here. In that analysis the characteristic polynomial has the form

$$(\lambda^2 - a\lambda + 1)(\lambda^2 - b\lambda + 1) + \psi(k\Delta x)(\lambda + 1)(\lambda - 1)^2, \tag{3.23}$$

where the coefficient a is the same as the coefficient $\tilde{a}$ in (3.22b), except for the term involving γ; $b = \tilde{b}$ exactly; $\psi(k\Delta x) > 0$, except at finitely many values of $k\Delta x$; and

$\psi(k\Delta x) = O(\Delta\alpha/\alpha_2)$, instead of $O(\Delta\alpha/\alpha_2)^{3/2}$. The constant term in (3.23) is $1 + \psi(k\Delta x)$, which is greater than 1 except for finitely many values of $k\Delta x$; the essence of the proof of instability in [8] is that the product of the roots is equal to the constant term, so at least one of the roots must have absolute value greater than 1. In contrast, the constant term in (3.22a) is 1.

3.4. *Analysis of Eigenvalues*

In order to characterize the stability properties of the algorithm defined in (3.19), we analyze the roots of the characteristic polynomial p defined in (3.22). In the following discussion we make the physically realistic assumption $\Delta\alpha/\alpha_2 \ll 1$, so that $\gamma \ll 1$ and $c_1/c_0 \ll 1$ (cf. (3.9), (3.11), (3.12)); in the ocean, the vertical variation of density has an upper bound on the order of one percent and is often much smaller (Gill [6]).

The characteristic polynomial p can be written in the form

$$p(\lambda) = p_a(\lambda)p_b(\lambda) + \phi(k\Delta x)\lambda(\lambda - 1)^2, \quad (3.24a)$$

where

$$\begin{aligned} p_a(\lambda) &= \lambda^2 - \tilde{a}\lambda + 1 \\ p_b(\lambda) &= \lambda^2 - \tilde{b}\lambda + 1. \end{aligned} \quad (3.24b)$$

The polynomials p_a and p_b are associated with the baroclinic and barotropic equations, respectively. We first examine the roots of $p_a(\lambda)p_b(\lambda)$, and we then examine the effects of adding the term $\phi(k\Delta x)\lambda(\lambda - 1)^2$.

LEMMA 3.1. *If the Courant–Friedrichs–Lewy condition $\nu = c_1\Delta t/\Delta x < 1$ is satisfied, then the roots of p_a and p_b have absolute value equal to* 1, *for all values of $k\Delta x$.*

Proof. If $k\Delta x = 0$, then $\tilde{a} = 2$, and $\lambda = 1$ is a double root of p_a. If $0 < |k\Delta x| \le \pi$, then $|\tilde{a}| < 2$, since $\nu < 1$. In that case, the roots of p_a are complex conjugates λ_1 and $\lambda_2 = \overline{\lambda}_1$ with nonzero imaginary part. The product of these roots is the constant term, so $1 = \lambda_1\lambda_2 = \lambda_1\overline{\lambda}_1 = |\lambda_1|^2 = |\lambda_2|^2$. An explicit calculation shows that the roots of p_b are $\exp(i\theta)$ and $\exp(-i\theta)$, with $\theta = Kc_0\Delta t$, for any value of ν. This completes the proof.

As $k\Delta x$ varies from 0 to $\pm\pi$, the coefficient $\tilde{a}$ varies from 2 downward to a minimum that is greater than -2. This coefficient is equal to the sum of the roots of p_a, and thus is equal to twice the real part of each root; the roots therefore move from $\lambda = 1$ around the unit circle, in opposite directions, toward the point $\lambda = -1$ without reaching $\lambda = -1$. These roots are associated with slow baroclinic motions. On the other hand, the coefficient $\tilde{b}$ oscillates between 2 and -2 many times, since $c_0/c_1 \gg 1$. The roots of p_b therefore move around the unit circle many times; these roots are associated with fast barotropic motions. As $k\Delta x$ varies from 0 to $\pm\pi$, the roots of p_b intersect the roots of p_a if and only if

$$\tilde{a}(k\Delta x) = \tilde{b}(k\Delta x). \quad (3.25)$$

In that case, the fourth-degree polynomial $p_a(\lambda)p_b(\lambda)$ has multiple roots on the unit circle. For all other values of $k\Delta x$, the polynomial $p_a(\lambda)p_b(\lambda)$ has four distinct roots on the unit circle.

The characteristic polynomial p in (3.24a) is a small perturbation of the polynomial $p_a(\lambda)p_b(\lambda)$, since $|\phi(k\Delta x)| = O(\Delta\alpha/\alpha_2)^{3/2} \ll 1$. A central question is whether this perturbation causes the roots to leave the unit circle. The following lemma implies some constraints on the manner in which the roots can migrate.

LEMMA 3.2. *Suppose that λ is a root of the polynomial p with $|\lambda| > 1$, for some $k\Delta x$. If λ is not real, then the set of all roots of p is $\{\lambda, 1/\lambda, \overline{\lambda}, 1/\overline{\lambda}\}$. If λ is real, then $1/\lambda$ is also a root, and the other two roots are either complex conjugates on the unit circle or are real roots that are reciprocals of each other.*

Proof. The symmetry of the coefficients in (3.24) implies that the equation $p(\lambda) = 0$ can be written in the form

$$\lambda^4 + A\lambda^3 + B\lambda^2 + A\lambda + 1 = 0$$

for suitable real coefficients A and B that depend on $k\Delta x$. Division by λ^4 reveals that $1/\lambda$ is a root, due to the symmetry. Complex conjugation shows that $\overline{\lambda}$ and $1/\overline{\lambda}$ are also roots. If $|\lambda| = 1$, then $1/\lambda = \overline{\lambda}$, and no new information is obtained here. However, if $|\lambda| > 1$ and λ is complex, then the roots discussed here are distinct, and the entire set of roots is then $\{\lambda, 1/\lambda, \overline{\lambda}, 1/\overline{\lambda}\}$. If $|\lambda| > 1$ with λ real, then the set of roots includes $1/\lambda$. This completes the proof.

Figure 3.1 illustrates the case where $|\lambda| > 1$ and λ is complex. The roots λ and $1/\overline{\lambda}$ lie along a ray through the origin in the complex plane, and the roots $\overline{\lambda}$ and $1/\lambda$ lie along the ray having an angle of opposite sign. We next estimate the maximum magnitudes of perturbations of roots.

LEMMA 3.3. *Suppose that the term $\phi(k\Delta x)\lambda(\lambda - 1)^2$ is added to the polynomial $p_a(\lambda)p_b(\lambda)$ to obtain $p(\lambda)$ in* (3.24a), *for given $k\Delta x$. Simple roots of $p_a(\lambda)p_b(\lambda)$ are perturbed by at most $O(\Delta\alpha/\alpha_2)^{3/2}$, and double roots are perturbed by at most $O(\Delta\alpha/\alpha_2)^{3/4}$.*

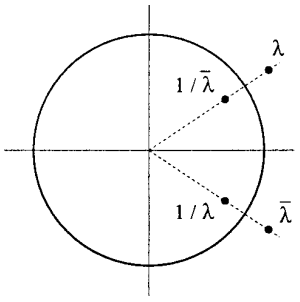

FIG. 3.1. Illustration of roots of the characteristic polynomial (3.24). If the roots leave the unit circle and are not real when the perturbation $\phi(k\Delta x)\lambda(\lambda-1)^2 = O(\Delta\alpha/\alpha_2)^{3/2}$ is added to $p_a(\lambda)p_b(\lambda)$, then they must move into the configuration illustrated here. This can happen only if the unperturbed roots are sufficiently close. This explains why the graph in the upper frame of Fig. 3.2 is exactly flat, except for a few narrow spikes.

Proof. Let λ_1, $\overline{\lambda}_1$, λ_2, and $\overline{\lambda}_2$ denote the roots of $p_a(\lambda)p_b(\lambda) = 0$. The equation $p(\lambda) = 0$ can then be written in the form

$$\begin{aligned}(\lambda-\lambda_1)(\lambda-\overline{\lambda}_1)(\lambda-\lambda_2)(\lambda-\overline{\lambda}_2) \\ = -\phi(k\Delta x)\lambda(\lambda-1)^2 \\ = O(\Delta\alpha/\alpha_2)^{3/2}.\end{aligned} \tag{3.26}$$

Consider, for example, the perturbation in λ_1. If λ_1 is a simple root, then Eq. (3.26) implies $\lambda - \lambda_1 = O(\Delta\alpha/\alpha_2)^{3/2}$, where the denominator in the bounding constant involves the distance between λ_1 and the other roots. If $k\Delta x$ varies so that another root approaches λ_1, then this estimate breaks down; for the case of a double root, a similar estimate for $(\lambda - \lambda_1)^2$ yields $\lambda - \lambda_1 = O(\Delta\alpha/\alpha_2)^{3/4}$. This completes the proof.

We next examine whether the roots leave the unit circle when the term $\phi(k\Delta x)\lambda(\lambda-1)^2$ is added to $p_a(\lambda)p_b(\lambda)$. Centered at each of the roots of $p_a(\lambda)p_b(\lambda)$ is a disk of radius $O(\Delta\alpha/\alpha_2)^{3/4}$ which contains a root of the polynomial p; this is a consequence of Lemma 3.3. Denote these disks by D_1, $\overline{D}_1$, D_2, and $\overline{D}_2$, respectively.

LEMMA 3.4. *Suppose that, for given $k\Delta x$, the disks D_1, $\overline{D}_1$, D_2, and $\overline{D}_2$ are disjoint and none of these disks intersect the real axis. In this case, all of the roots of $p(\lambda) = 0$ must lie exactly on the unit circle.*

Proof. Suppose that one of the roots of p does not lie on the unit circle. Denote this root by μ, and without loss of generality assume that μ lies in disk D_1. It follows from Lemma 3.2 that μ, $1/\mu$, $\overline{\mu}$, and $1/\overline{\mu}$ are distinct roots of p. The root $\overline{\mu}$ lies in the disk $\overline{D}_1$, and the roots $1/\overline{\mu}$ and $1/\mu$ lie along radial lines through μ and $\overline{\mu}$, respectively. However, the disks D_2 and $\overline{D}_2$ also contain roots of p. These roots must be distinct from $\{\mu, 1/\mu, \overline{\mu}, 1/\overline{\mu}\}$, since the disks D_1, $\overline{D}_1$, D_2, and $\overline{D}_2$ are disjoint. The polynomial p would then have at least six distinct roots, which is impossible since p has degree four. Therefore each root of p must lie on the unit circle. This completes the proof.

Next suppose that the disks D_1, $\overline{D}_1$, D_2, and $\overline{D}_2$ are not disjoint, but continue to assume that none of these disks intersect the real axis. The preceding analysis leaves open the possibility that the roots could leave the unit circle in this case; we now analyze whether this actually takes place.

Intersection of disks is possible if the roots of p_a and p_b coincide or if the roots are distinct but sufficiently close. In order to provide a unified treatment of these two possibilities, we consider the polynomial

$$p(\lambda, \tau) = p_a(\lambda)p_b(\lambda) + \tau\phi(k\Delta x)\lambda(\lambda-1)^2$$

for $0 \le \tau \le 1$. The case $\tau = 0$ yields the unperturbed polynomial $p_a(\lambda)p_b(\lambda)$, and the case $\tau = 1$ yields $p(\lambda)$. If the roots of p_a and p_b are distinct, then for sufficiently small τ the roots λ of $p(\lambda, \tau) = 0$ are distinct and lie exactly on the unit circle; this follows from Lemma 3.4, with disks having radii that are scaled by $\sqrt{\tau}$. Now let τ_0 denote the largest number in the interval $(0, 1]$ such that $p(\lambda, \tau)$ has distinct roots λ on the unit circle for all τ satisfying $0 < \tau < \tau_0$. If $\tau_0 = 1$, then the roots of $p(\lambda)$ lie on the unit circle, by continuity. Otherwise, observe that $p(\lambda, \tau_0)$ has double complex roots λ_0 and $\overline{\lambda}_0$ on the unit circle, and

$$\begin{aligned}p(\lambda) &= p(\lambda, \tau_0) + (1-\tau_0)\phi(k\Delta x)\lambda(\lambda-1)^2 \\ &= (\lambda-\lambda_0)^2(\lambda-\overline{\lambda}_0)^2 + (1-\tau_0)\phi(k\Delta x)\lambda(\lambda-1)^2.\end{aligned}$$

Let $\lambda_0 + \varepsilon$ denote a root that is a perturbation of λ_0; we then have $\varepsilon = O(\Delta\alpha/\alpha_2)^{3/4}$. Substitution into the equation $p(\lambda) = 0$, followed by a calculation, shows that the quantity $(\varepsilon/\lambda_0)^2$ alternates in sign between consecutive values of $k\Delta x$ for which the roots of p_a and p_b coincide. In the case where ε/λ_0 is real, the perturbed root leaves the unit circle. In the case where ε/λ_0 is imaginary, the perturbation lies at right angles to the root and is thus tangent to the unit circle; there are two such perturbations ε pointing in opposite directions. An application of Lemma 3.2 then shows that the roots lie exactly on the unit circle.

The roots of p_a and p_b coincide for those values of $k\Delta x$ for which $\tilde{a}(k\Delta x) = \tilde{b}(k\Delta x)$. We have thus shown that in a narrow neighborhood of every other such point, there exist perturbed roots having absolute value greater than 1. At the other points, the perturbed roots remain on the unit circle exactly.

We now estimate the widths of the intervals (in $k\Delta x$) on which the roots of p can have modulus greater than 1. In the following, we sketch the main ideas and omit the details of calculations. Here, we continue to assume that the disks D_1, $\overline{D}_1$, D_2, and $\overline{D}_2$ do not intersect the real axis.

In order for a root of p to leave the unit circle, there must be some overlap among the disks D_1, $\overline{D}_1$, D_2, and $\overline{D}_2$. That is, a root of p_a and a root of p_b must differ by at most $O(\Delta\alpha/\alpha_2)^{3/4}$. If neighborhoods of $\lambda = \pm 1$ are excluded, then this condition is equivalent to $|\tilde{a} - \tilde{b}| = O(\Delta\alpha/\alpha_2)^{3/4}$, since the roots of p_a and p_b have real parts $\tilde{a}/2$ and $\tilde{b}/2$, respectively. We then need to determine the width of an interval on which the quantity $|\tilde{a} - \tilde{b}|$ can vary from zero to $(\Delta\alpha/\alpha_2)^{3/4}$. A calculation of derivatives shows $\tilde{a}'(k\Delta x) = O(\sin(k\Delta x))$ and $\tilde{b}'(k\Delta x) = O((\Delta\alpha/\alpha_2)^{-1/2} \sin\theta \cos(k\Delta x/2))$. If the trigonometric factors are bounded away from zero, then we conclude that the interval has width $O(\Delta\alpha/\alpha_2)^{5/4}$. On the other hand, if the roots of p_a and p_b intersect for $k\Delta x = \pm\pi$, then a separate calculation for this case shows that the interval has width $O(\Delta\alpha/\alpha_2)^{5/8}$. For the intersection points nearest $\lambda = 1$, we have $k\Delta x = O(\Delta\alpha/\alpha_2)^{1/2}$ and $\sin\theta = O(\Delta\alpha/\alpha_2)^{1/2}$, and the estimate of interval width can be improved to $O(\Delta\alpha/\alpha_2)^{7/4}$.

In the preceding discussion, it has been assumed that the disks D_1, $\overline{D}_1$, D_2, and $\overline{D}_2$ do not intersect the real axis. Equivalently, we have assumed that the roots of $p_a(\lambda)p_b(\lambda)$ are located sufficiently far from the real axis that the roots of the perturbed polynomial p cannot be real. We now consider the case where at least one of the disks intersects the real axis. Our analysis shows that perturbed roots cannot leave the unit circle near $+1$, but roots can leave the unit circle near -1. These roots exist for values of $k\Delta x$ lying in extremely narrow intervals lying to one side of points for which $\tilde{b}(k\Delta x) = -2$. Compared to the cases discussed earlier, the intervals are far narrower, and the roots are much closer to the unit circle. Some sample numbers are described later during the discussion of Fig. 3.2. This case is highly unlikely to be significant in practice.

We now assemble the preceding results into a summary of the stability properties of the one-dimensional, barotropic-baroclinic algorithm defined in (3.19).

PROPOSITION 3.5. *If the Courant–Friedrichs–Lewy condition* $\nu = c_1\Delta t/\Delta x < 1$ *is satisfied, then all solutions of the difference equation* (3.19) *are bounded as* $n \to +\infty$, *except for values of* $k\Delta x$ *lying in a finite number of very narrow intervals. Unbounded solutions are found in neighborhoods of every other point for which* $\tilde{a}(k\Delta x) = \tilde{b}(k\Delta x)$. *If* $|k\Delta x| = \pi$ *for such a point, then the neighborhood has width* $O(\Delta\alpha/\alpha_2)^{5/8}$; *if* $|k\Delta x| < \pi$, *then the width is* $O(\Delta\alpha/\alpha_2)^{5/4}$; *if* $k\Delta x$ *is near* 0, *then the estimate can be improved to* $O(\Delta\alpha/\alpha_2)^{7/4}$.

Remainder of proof. The roots of the characteristic polynomial, which have just been discussed, are eigenvalues of the problem (3.20). This eigenvalue problem arose from considering solutions of (3.19) that have the form $v^n = \lambda^n q$, where q is a nonzero vector having four components. If $k\Delta x$ lies in one of the narrow neighborhoods described in Proposition 3.5 then there exist roots λ for which $|\lambda| > 1$, so the difference equation (3.19) admits solutions that are unbounded as $n \to +\infty$. On the other hand, if $k\Delta x$ lies outside all of these neighborhoods, then $|\lambda| = 1$ for all roots λ. Furthermore, the preceding analysis shows that the roots are distinct in this case. Eigenvectors corresponding to distinct eigenvalues are linearly independent, so every solution of (3.29) can be written as a linear combination of special solutions of the form $\lambda^n q$ with $|\lambda| = 1$, and the proposition follows. This completes the proof.

The behavior of the eigenvalues is illustrated in Fig. 3.2. We consider a two-layer model for which $\Delta\tilde{p}_1/\tilde{p}_2 = 0.25$, $\Delta\tilde{p}_2/\tilde{p}_2 = 0.75$, and $\Delta\alpha/\alpha_2 = 0.01$. We also assume that Δt and Δx are chosen so that the Courant number is $\nu = c_1\Delta t/\Delta x = 0.8$. We computed eigenvalues of (3.20) for values of $k\Delta x$ varying from $-\pi$ to π in increments of $\pi/1000$, and also for the values of $k\Delta x$ for which $\tilde{a}(k\Delta x) = \tilde{b}(k\Delta x)$. For each $k\Delta x$, we then determined the maximum of the absolute values of the eigenvalues. These maximum absolute values are plotted versus $k\Delta x$ in the upper frame in Fig. 3.2. The lower frame shows plots of $\tilde{a}(k\Delta x)$ and $\tilde{b}(k\Delta x)$. The small circles on the horizontal axis in the upper frame indicate the horizontal coordinates of the points where $\tilde{a}(k\Delta x) = \tilde{b}(k\Delta x)$. Narrow spikes appear in the upper frame at every other point for which $\tilde{a}(k\Delta x) = \tilde{b}(k\Delta x)$, but elsewhere the graph is exactly flat. This behavior is consistent with the preceding analysis. Close-up views of the spikes show that they are smooth and intersect the horizontal axis at right angles.

The preceding analysis implies the existence of very narrow spikes near points where $\tilde{b} = -2$. Such spikes do not appear in the figure, and they also do not appear when $k\Delta x$ is sampled in increments of $\pi/20000$. Further experimentation shows that these spikes have widths on the order of 10^{-6}, and the largest such spike has height 1.00004, which would not be visible with the vertical scaling used in the figure. These spikes are mentioned here only for the sake of completeness.

Higdon and Bennett [8] analyzed the splitting of Bleck and Smith [1] in the same setting considered here. In that case the plots of eigenvalues also show sharp spikes. However, between the spikes the graphs remain above 1 except for finitely many values of $k\Delta x$, and the spikes are higher; in [8] the maximum magnitude of eigenvalues is $1 + O(\Delta\alpha/\alpha_2)^{1/2}$, but in the present analysis the maximum magnitude of $1 + O(\Delta\alpha/\alpha_2)^{3/4}$.

As noted in (3.25), $\tilde{a}(k\Delta x) = \tilde{b}(k\Delta x)$ if and only if the roots of the polynomials p_a and p_b coincide. The roots of these polynomials indicate the time dependence, in the form λ^n, of solutions of the baroclinic and barotropic subsystems, respectively. Coincidence of roots then occurs when these subsystems have the same time dependence on the coarse time grid that is associated with the baroclinic

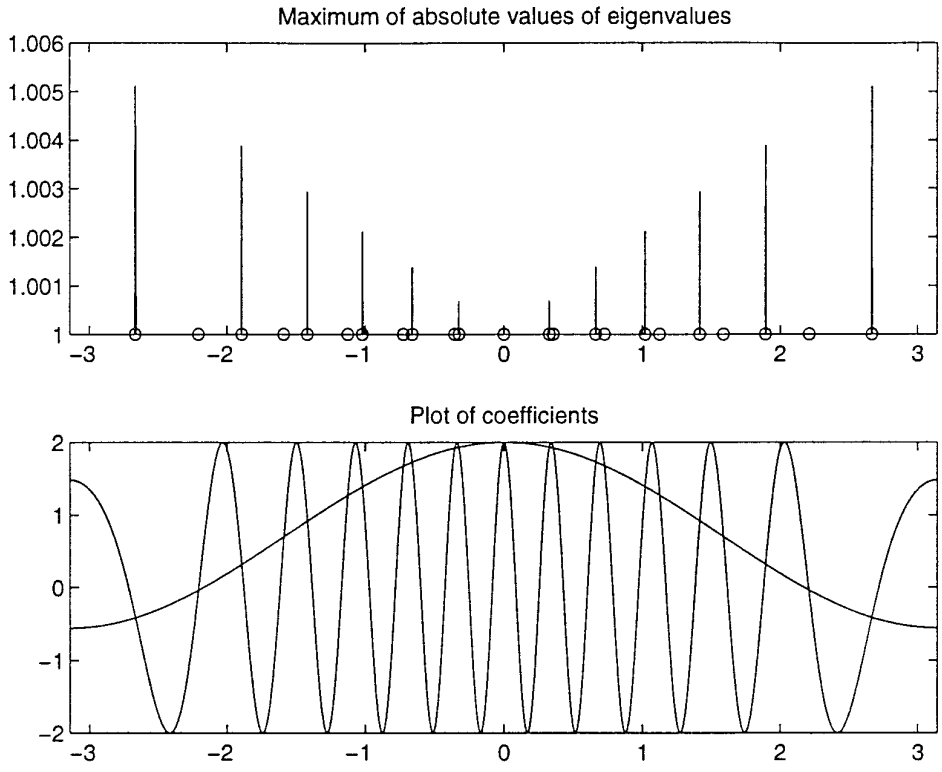

FIG. 3.2. Behavior of eigenvalues. The top frame shows the maximum of the absolute values of the eigenvalues for the algorithm (3.19), as a function of $k\Delta x$ for $|k\Delta x| \le \pi$. Spikes are found in narrow neighborhoods of every other point for which the roots of the polynomials $p_a(\lambda)$ and $p_b(\lambda)$ coincide; these roots coincide if and only if $\tilde{a}(k\Delta x) = \tilde{b}(k\Delta x)$, where $\tilde{a}$ and $\tilde{b}$ are the coefficients in the linear terms in $p_a(\lambda)$ and $p_b(\lambda)$. These coefficients are plotted in the lower frame. The slowly varying function is $\tilde{a}$. In this example, the model parameters are $\Delta\tilde{p}_1/\tilde{p}_2 = 0.25$, $\Delta\tilde{p}_2/\tilde{p}_2 = 0.75$, $\Delta\alpha/\alpha_2 = 0.01$, and $\nu = c_1\Delta t/\Delta x = 0.8$.

equations. Since the barotropic variables typically vary more quickly with respect to t than do the baroclinic variables, the coincidence of roots involves the aliasing of fast barotropic motions to slowly varying motions on a coarse time grid. The spikes appearing in Fig. 3.2 can be interpreted as a kind of resonance that occurs between the baroclinic and barotropic subsystems. This resonance is apparently related to the inexactness of the projections onto the external mode given by the vertical averaging defined in (2.3), (2.11), and (3.4). However, this behavior occurs only at every other point where the baroclinic and barotropic subsystems have the same time dependence, and away from these points the algorithm is stable and nondissipative.

In the preceding analysis, the wavenumber k has been treated as a continuous variable. However, in practice the wavenumber is discrete, since x is discrete. The spikes in Fig. 3.2 can affect a computation only if one of the discrete values of $k\Delta x$ lies within the width of such a spike. Unless $k\Delta x$ lands in the center of that interval, the corresponding value of $|\lambda|$ will be less than the value implied by the preceding discussions.

3.5. *Leapfrog*

Now suppose that the baroclinic equations are discretized with respect to t by using the leapfrog scheme instead of the forward–backward method that has been used up to this point. In the present section we show that the resulting algorithm has stability properties that are very similar to those obtained above.

As in the preceding analysis, we discretize with respect to x by using centered second-order differences on a staggered grid, and we apply a Fourier transform with respect to x. The leapfrog approximation to the linearized baroclinic system (3.8) can then be written in the form

$$\hat{\delta}^{n+1} = \hat{\delta}^{n-1} - 2(iKc_1\Delta t)\frac{\widehat{u_1}^{\,n}}{c_1} \tag{3.29a}$$

$$\frac{\widehat{u_1}^{\,n+1}}{c_1} = \frac{\widehat{u_1}^{\,n-1}}{c_1} - 2(iKc_1\Delta t)(\hat{\delta}^n + \hat{\eta}^n), \tag{3.29b}$$

where K is defined in (3.15). As before, we assume that the barotropic equations are solved exactly with respect

to t for $t_n \le t \le t_{n+1}$, with the solution given in (3.17). The coupled barotropic-baroclinic system can then be written in the form

$$E_1 v^{n+1} = E_0 v^n + E_{-1} v^{n-1}, \tag{3.30a}$$

where v^n is defined in (3.18),

$$E_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -W_1 & 0 & 1 & 0 \\ -W_2 & 0 & 0 & 1 \end{pmatrix}, \tag{3.30b}$$

$$E_0 = \begin{pmatrix} 0 & -2iKc_1\Delta t & 0 & 0 \\ -2iKc_1\Delta t & 0 & 0 & -2iKc_1\Delta t \\ V_1 & 0 & Y_{11} & Y_{12} \\ V_2 & 0 & Y_{21} & Y_{22} \end{pmatrix}, \tag{3.30c}$$

$E_{-1} = \mathrm{diag}(1, 1, 0, 0)$, and $Y(k\Delta x)$, $V(k\Delta x)$, and $W(k\Delta x)$ are defined in (3.17). The matrices E_1, E_0, and E_{-1} are functions of $k\Delta x$.

We now analyze solutions of (3.30) having the form $v^n = \lambda^n q$, where λ is a complex scalar and q is a vector having four components. If $\lambda \neq 0$ for such a solution, then $(\lambda^2 E_1 - \lambda E_0 - E_{-1})q = 0$; for this case, nonzero vectors q can be found if and only if

$$p_8(\lambda) \equiv \det(\lambda^2 E_1 - \lambda E_0 - E_{-1}) = 0.$$

A calculation shows $p_8(\lambda) = \lambda^2 p_6(\lambda)$, where p_6 is a polynomial of degree six. The superfluous factor λ^2 arises from the fact that the barotropic equations involve only two time levels, whereas the matrix formulation (3.30) involves three levels. Nontrivial solutions of the form $\lambda^n q$ thus correspond to roots of the characteristic polynomial p_6, which can be written in the form

$$\begin{aligned} p_6(\lambda) = {} & (\lambda^4 - \tilde{a}\lambda^2 + 1)(\lambda^2 - \tilde{b}\lambda + 1) \\ & - \gamma \cdot 16\nu^2 \sin^2(k\Delta x/2) r(\lambda), \end{aligned} \tag{3.31a}$$

where

$$\tilde{a}(k\Delta x) = 2 - 4(Kc_1\Delta t)^2 = 2 - 16\nu^2 \sin^2(k\Delta x/2)$$

$$\tilde{b}(k\Delta x) = 2\cos\theta$$

$$r(\lambda) = \left[\left(1 - \frac{\sin\theta}{\theta}\right)(\lambda^4 - 2\lambda^3 + \lambda^2) + 2(1 - \cos\theta)\lambda^3 \right]. \tag{3.31b}$$

Here, $\theta = Kc_0\Delta t$ and $\nu = c_1\Delta t/\Delta x$. From (3.12), we have $\gamma = O(\Delta\alpha/\alpha_2)$.

The roots of the factor $\lambda^2 - \tilde{b}\lambda + 1$ in (3.31a) are $\exp(i\theta)$ and $\exp(-i\theta)$. The factor $\lambda^4 - \tilde{a}\lambda^2 + 1$ is the characteristic polynomial for the leapfrog scheme, as applied to the purely baroclinic equations obtained by deleting the term involving ν from (3.14). A calculation shows that this scheme is stable and nondissipative provided $\nu < \frac{1}{2}$; for this scheme, the maximum permissible time step is half that of the forward–backward method. We then conclude that if $\nu < \frac{1}{2}$, then all of the roots of the polynomial

$$(\lambda^4 - \tilde{a}\lambda^2 + 1)(\lambda^2 - \tilde{b}\lambda + 1)$$

lie on the unit circle, for each $k\Delta x$.

The remaining question is whether the term in (3.31a) involving $r(\lambda)$ causes the roots to leave the unit circle. If a complex number λ is a root of $p_6(\lambda) = 0$, then so is $1/\lambda$; this follows from the symmetry of the coefficients of the powers of λ in $p_6(\lambda)$. An analysis similar to that in Section 3.4 shows that the roots must lie exactly on the unit circle, except possibly for values of $k\Delta x$ lying in very narrow neighborhoods of points where roots of $(\lambda^4 - \tilde{a}\lambda^2 + 1)$ and $(\lambda^2 - \tilde{b}\lambda + 1)$ coincide or where the roots of the latter intersect the real axis. Because of the similarity with the previous case, we omit the details. If the roots coincide, then $\exp(i\theta)$ and $\exp(-i\theta)$ are roots of $(\lambda^4 - \tilde{a}\lambda^2 + 1)$; equivalently, $\exp(2i\theta)$ and $\exp(-2i\theta)$ are roots of $(\mu^2 - \tilde{a}\mu + 1)$. The roots therefore coincide if and only if $\tilde{a}(k\Delta x) = 2\cos(2\theta)$.

The behavior of the eigenvalues λ is illustrated in Fig. 3.3. In this example we use the same parameters as in Fig. 3.2, except that the Courant number is chosen to be $\nu = 0.4$. The circles on the horizontal axis denote the values of $k\Delta x$ for which $\tilde{a}(k\Delta x) = 2\cos(2\theta)$. The spikes in the figure are found at every other such point; this is analogous to the behavior found previously for the forward–backward method.

4. THE CASE OF TWO HORIZONTAL DIMENSIONS

In this section we extend the preceding analysis to the case of two horizontal dimensions and a rotating reference frame. An additional issue encountered here is the treatment of the Coriolis terms. We consider one method for treating these terms, and for this example we show that the splitting algorithm has stability properties similar to those of the one-dimensional case.

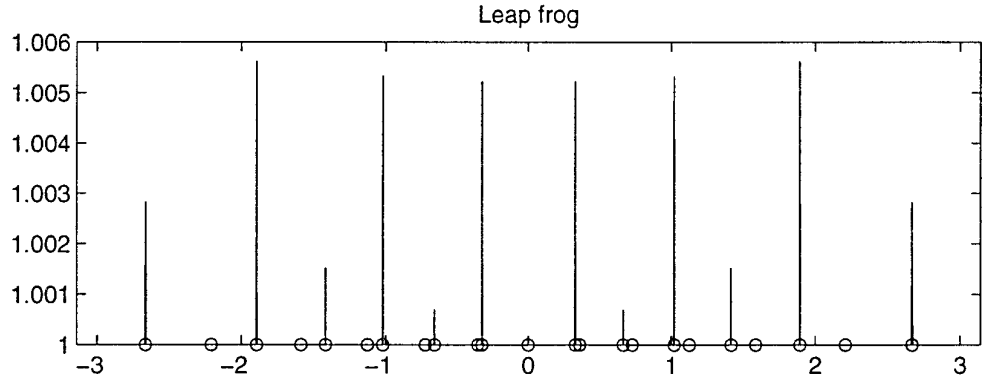

FIG. 3.3. Behavior of eigenvalues when the leap frog method is used to discretize the baroclinic equations with respect to time. The same parameters are used as in Fig. 3.2, except that $\nu = 0.4$.

4.1. *Governing Equations*

As in Section 3, we consider linearized dynamics in a fluid with two layers and a flat lower boundary. The linearization of the system (2.1), as applied to the present setting, is

$$\frac{\partial u_r}{\partial t} - fv_r = -\frac{\partial M_r}{\partial x}, \quad r = 1, 2, \tag{4.1a}$$

$$\frac{\partial v_r}{\partial t} + fu_r = -\frac{\partial M_r}{\partial y}, \quad r = 1, 2, \tag{4.1b}$$

$$M_1 = M_2 + p_1 \Delta\alpha \tag{4.1c}$$

$$\frac{\partial}{\partial t}(\Delta p_r) + \Delta\tilde{p}_r \left(\frac{\partial u_r}{\partial x} + \frac{\partial v_r}{\partial y}\right) = 0, \quad r = 1, 2. \tag{4.1d}$$

Here, $u_r(x, y, t)$ and $v_r(x, y, t)$ denote the x- and y-components of velocity, respectively, in layer r; and the quantities M_r and Δp_r depend on (x, y, t). Otherwise, the notation is the same as in Section 3. The Coriolis parameter f is assumed constant.

The barotropic and baroclinic equations are obtained through a process that is very similar to that used in Section 3.1. In the present case, the barotropic equations are

$$\frac{\partial \overline{u}}{\partial t} - f\overline{v} + c_0^2 \frac{\partial \eta}{\partial x} = -\gamma c_0^2 \frac{\partial \delta}{\partial x} \tag{4.2a}$$

$$\frac{\partial \overline{v}}{\partial t} + f\overline{u} + c_0^2 \frac{\partial \eta}{\partial y} = -\gamma c_0^2 \frac{\partial \delta}{\partial y} \tag{4.2b}$$

$$\frac{\partial \eta}{\partial t} + \frac{\partial \overline{u}}{\partial x} + \frac{\partial \overline{v}}{\partial y} = 0, \tag{4.2c}$$

where c_0 and γ are defined in (3.11)–(3.12). The baroclinic equations are

$$\frac{\partial u_1'}{\partial t} - fv_1' + c_1^2 \frac{\partial \delta}{\partial x} = -c_1^2 \frac{\partial \eta}{\partial x} \tag{4.3a}$$

$$\frac{\partial v_1'}{\partial t} + fu_1' + c_1^2 \frac{\partial \delta}{\partial y} = -c_1^2 \frac{\partial \eta}{\partial y} \tag{4.3b}$$

$$\frac{\partial \delta}{\partial t} + \frac{\partial u_1'}{\partial x} + \frac{\partial v_1'}{\partial y} = 0, \tag{4.3c}$$

where c_1 is defined in (3.9), and

$$\delta(x, y, t) = \frac{\Delta p_1'(x, y, t)}{\Delta\tilde{p}_1}.$$

In the following analysis of numerical discretization, we use potential vorticity and the divergence of velocity as dependent variables instead of the velocity components listed above. This is due to the following conservation principle. Let $\overline{\zeta}(x, y, t) = \partial\overline{v}/\partial x - \partial\overline{u}/\partial y$ denote the barotropic vorticity; a manipulation of the system (4.2) then yields

$$\frac{\partial}{\partial t}(\overline{\zeta} - f\eta) = 0. \tag{4.4}$$

Up to a multiplicative constant, the quantity $\overline{\zeta} - f\eta$ is a linearization of the potential vorticity that is conserved in solutions of the constant-density shallow water equations via vortex stretching (Gill [6], Pedlosky [12]). A similar analysis shows that the quantity $v_x' - u_y' - f\delta$ is conserved in solutions of the baroclinic system (4.3). Discrete ana-

logues of these principles will lead to considerable simplifications of the analysis given in Sections 4.2 and 4.3.

4.2. *Computational Algorithm*

We now apply the time discretization described in Section 2.3 to the barotropic-baroclinic system (4.2)–(4.3).

The algorithm is stated directly in terms of Fourier transforms with respect to x and y. For definiteness, we use a spatial discretization based on the C-grid of Arakawa [7, 10]. To construct this grid, partition the spatial domain into rectangular grid cells; here we assume $\Delta x = \Delta y = h$. For each cell, values of η and δ are associated with the center of the cell, values of $\overline{u}$ and u' are taken at the midpoints of the sides corresponding to fixed x, and values of $\overline{v}$ and v' are taken at the midpoints of the sides corresponding to fixed y. Alternate spatial discretizations can be handled by using different definitions of the symbols K and L in (4.6).

As in Section 3, we assume that the barotropic equations are discretized with respect to space but are solved exactly with respect to t on the time interval $t_n \le t \le t_{n+1}$. Let $\hat{\overline{u}}(k,l,t)$, $\hat{\overline{v}}(k,l,t)$, $\hat{\eta}(k,l,t)$, and $\hat{\delta}(k,l,t)$ denote the Fourier transforms of $\overline{u}$, $\overline{v}$, η, and δ with respect to (x, y) for fixed t. The transformed barotropic system, for $t_n \le t \le t_{n+1}$, then has the form

$$\frac{\partial \hat{\overline{u}}}{\partial t} - f_0 \hat{\overline{v}} + iKc_0^2 \hat{\eta} = -iK\gamma c_0^2 \hat{\delta}(k,l,t) \tag{4.5a}$$

$$\frac{\partial \hat{\overline{v}}}{\partial t} + f_0 \hat{\overline{u}} + iLc_0^2 \hat{\eta} = -iL\gamma c_0^2 \hat{\delta}(k,l,t) \tag{4.5b}$$

$$\frac{\partial \hat{\eta}}{\partial t} + iK\hat{\overline{u}} + iL\hat{\overline{v}} = 0, \tag{4.5c}$$

where

$$iK = i\frac{\sin(kh/2)}{(h/2)}, \quad iL = i\frac{\sin(lh/2)}{(h/2)} \tag{4.6}$$

in analogy to (3.15),

$$f_0 = f\cos(kh/2)\cos(lh/2), \tag{4.7}$$

and $\hat{\delta}(k, l, t)$ denotes the linear interpolant

$$\hat{\delta}(k,l,t) = \hat{\delta}^n(k,l) + (t - t_n)(\hat{\delta}^{n+1}(k,l) - \hat{\delta}^n(k,l))/\Delta t.$$

The quantity $\cos(kh/2)\cos(lh/2)$ in (4.7) arises from the fact that the values of $\overline{u}$ and $\overline{v}$ are not associated with the same points on the staggered C-grid used here; in the momentum equations involving $\partial\overline{u}/\partial t$ and $\partial\overline{v}/\partial t$, the Coriolis terms $-f\overline{v}$ and $f\overline{u}$ are approximated by four-point averages.

We next express the barotropic system (4.5) in terms of divergence and potential vorticity. Let

$$\hat{D}(k,l,t) = iK\hat{\overline{u}} + iL\hat{\overline{v}}$$

denote a discretization of the horizontal velocity divergence $\partial\overline{u}/\partial x + \partial\overline{v}/\partial y$, and let

$$\hat{P}(k,l,t) = iK\hat{\overline{v}} - iL\hat{\overline{u}} - f_0\hat{\eta}$$

denote a discretization of the quantity $\overline{\zeta} - f\eta$ in (4.4) that is a constant multiple of the linearized potential vorticity. Some manipulations of the system (4.5) then yield

$$\frac{\partial \hat{D}}{\partial t} - f_0\hat{P} - s_0^2\hat{\eta} = \gamma\omega_0^2\hat{\delta} \tag{4.8a}$$

$$\frac{\partial \hat{P}}{\partial t} = 0 \tag{4.8b}$$

$$\frac{\partial \hat{\eta}}{\partial t} + \hat{D} = 0, \tag{4.8c}$$

where

$$\begin{aligned} \omega_0^2 &= c_0^2(K^2 + L^2) \\ s_0^2 &= c_0^2(K^2 + L^2) + f_0^2. \end{aligned} \tag{4.9}$$

The quantity s_0 is the frequency of oscillation in the system (4.8), and ω_0 is the frequency that would be obtained if rotation were absent. Equation (4.8b) expresses the conservation of discrete potential vorticity $\hat{P}$, in analogy to (4.4).

Differentiation of Eqs. (4.8a) and (4.8c) yields second-order forced wave equations for $\hat{D}$ and $\hat{\eta}$, which are solved by standard methods. Next impose the initial conditions $\hat{D} = \hat{D}^n$, $\hat{\eta} = \hat{\eta}^n$, and $\hat{P} = \hat{P}^n$ at time $t = t_n$, and then let $t = t_{n+1}$ to obtain the solution at the new time level. A calculation shows that the solution of (4.8) at time t_{n+1} is

$$\begin{aligned} \begin{pmatrix} \hat{\eta}^{n+1} \\ \hat{D}^{n+1}/s_0 \\ \hat{P}^{n+1}/s_0 \end{pmatrix} &= \begin{pmatrix} \cos\theta & -\sin\theta & -(f_0/s_0)(1-\cos\theta) \\ \sin\theta & \cos\theta & (f_0/s_0)\sin\theta \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \hat{\eta}^n \\ \hat{D}^n/s_0 \\ \hat{P}^n/s_0 \end{pmatrix} \\ &\quad + \gamma(\omega_0/s_0)^2 \begin{pmatrix} \cos\theta - (\sin\theta)/\theta \\ \sin\theta - (1-\cos\theta)/\theta \\ 0 \end{pmatrix} \hat{\delta}^n \\ &\quad + \gamma(\omega_0/s_0)^2 \begin{pmatrix} (\sin\theta)/\theta - 1 \\ (1-\cos\theta)/\theta \\ 0 \end{pmatrix} \hat{\delta}^{n+1}, \end{aligned} \tag{4.10}$$

where $\theta = s_0\Delta t$. The quantities θ, f_0/s_0, and $(\omega_0/s_0)^2$ in (4.10) can be written in terms of trigonometric functions of the quantities kh and lh and of the dimensionless parameters c_0/fh and $f\Delta t$. Due to periodicity, we can assume $|kh| \le \pi$ and $|lh| \le \pi$. The sign of s_0 is not determined in (4.9); however, (4.10) is invariant under a change of sign of s_0.

Next consider the baroclinic equations. In the algorithm outlined in Section 2.3, the baroclinic continuity equation is advanced explicitly, the barotropic equations are then solved, and finally the baroclinic momentum equations are advanced with a variation on a backward method that includes a predictor–corrector step for the nonlinear and Coriolis terms. When this algorithm is applied to the linearized baroclinic system (4.3), the discrete continuity equation is

$$\hat{\delta}^{n+1} = \hat{\delta}^n - \Delta t(iK\widehat{u_1'}^n + iL\widehat{v_1'}^n), \tag{4.11}$$

the prediction step for the momentum equations is

$$\begin{aligned}\widehat{u_1'}^{\,\mathrm{pred}} &= \widehat{u_1'}^n + \Delta t(f_0\widehat{v_1'}^n - iKc_1^2(\hat{\delta}^{n+1} + \hat{\eta}^{n+1}))\\ \widehat{v_1'}^{\,\mathrm{pred}} &= \widehat{v_1'}^n + \Delta t(-f_0\widehat{u_1'}^n - iLc_1^2(\hat{\delta}^{n+1} + \hat{\eta}^{n+1})),\end{aligned} \tag{4.12}$$

and the correction step is

$$\begin{aligned}\widehat{u_1'}^{n+1} &= \widehat{u_1'}^n + \Delta t\,[f_0(a\widehat{v_1'}^{\,\mathrm{pred}} + b\widehat{v_1'}^n)\\ &\quad - iKc_1^2(\hat{\delta}^{n+1} + \hat{\eta}^{n+1})]\\ \widehat{v_1'}^{n+1} &= \widehat{v_1'}^n + \Delta t\,[-f_0(a\widehat{u_1'}^{\,\mathrm{pred}} + b\widehat{u_1'}^n)\\ &\quad - iLc_1^2(\hat{\delta}^{n+1} + \hat{\eta}^{n+1})],\end{aligned} \tag{4.13}$$

where a and b are constants with $a + b = 1$.

We now express the system (4.11)–(4.13) in terms of divergence and potential vorticity. Let

$$\begin{aligned}\hat{D}'^n(k,l) &= iK\widehat{u_1'}^n + iL\widehat{v_1'}^n\\ \hat{\zeta}'^n(k,l) &= iK\widehat{v_1'}^n - iL\widehat{u_1'}^n\end{aligned} \tag{4.14}$$

denote horizontal discretizations of divergence and vorticity, respectively. In (4.13), replace $\widehat{u_1'}^{\,\mathrm{pred}}$ and $\widehat{v_1'}^{\,\mathrm{pred}}$ by the values given in (4.12). Some manipulations of the resulting system then yield

$$\begin{aligned}\hat{D}'^{n+1} &= (1 - a\phi^2)\hat{D}'^n + \phi\hat{\zeta}'^n\\ &\quad + (K^2 + L^2)c_1^2\Delta t(\hat{\delta}^{n+1} + \hat{\eta}^{n+1})\end{aligned} \tag{4.15a}$$

$$\begin{aligned}\hat{\zeta}'^{n+1} &= (1 - a\phi^2)\hat{\zeta}'^n - \phi\hat{D}'^n\\ &\quad - a\phi(K^2 + L^2)c_1^2\Delta t(\hat{\delta}^{n+1} + \hat{\eta}^{n+1}),\end{aligned} \tag{4.15b}$$

where $\phi = f_0\Delta t$. In (4.15b), eliminate the terms involving $\hat{\delta}^{n+1}$ and $\hat{\eta}^{n+1}$ by using (4.15a). Further manipulation yields $\hat{P}'^{n+1} = \hat{P}'^n$, where

$$\hat{P}'^n = \hat{\zeta}'^n - f_0(1 + a^2\phi^2)\hat{\delta}^n + a\phi\hat{D}'^n. \tag{4.16}$$

In this calculation, the terms of the form $f_0(1 + a^2\phi^2)\hat{\delta}$ arise from a substitution involving the continuity equation (4.11) when written in the form $\hat{\delta}^{n+1} = \hat{\delta}^n - \Delta t\hat{D}'^n$.

The quantity $\hat{P}'^n$ in (4.16) can be written in the form $\hat{P}'^n = \hat{\zeta}'^n - f_0\hat{\delta}^n + O(\Delta t)$. This quantity is conserved exactly by the present computational algorithm, and it is a consistent approximation to the linearized potential vorticity $\zeta' - f\delta = v_x' - u_y' - f\delta$ that is conserved in the continuous baroclinic system (4.3).

After the divergence equation (4.15a) is written in terms of $\hat{P}'^n$, the discrete baroclinic equations can be summarized as

$$\begin{aligned}&\begin{pmatrix}1 & 0 & 0\\ -\rho\theta/r & 1 & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\hat{\delta}^{n+1}\\ \hat{D}'^{n+1}/s_0\\ \hat{P}'^{n+1}/s_0\end{pmatrix}\\ &\quad = \begin{pmatrix}1 & -\theta & 0\\ (f_0/s_0)\phi(1 + a^2\phi^2) & 1 - 2a\phi^2 & \phi\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\hat{\delta}^n\\ \hat{D}'^n/s_0\\ \hat{P}'^n/s_0\end{pmatrix}\\ &\qquad + \begin{pmatrix}0\\ \rho\theta/r\\ 0\end{pmatrix}\hat{\eta}^{n+1},\end{aligned} \tag{4.17a}$$

where

$$\rho(kh, lh) = \gamma(\omega_0/s_0)^2 = \left(\frac{c_1}{c_0}\right)^2\frac{\Delta\tilde{p}_1}{\Delta\tilde{p}_2}\left(\frac{\omega_0}{s_0}\right)^2 = O(\Delta\alpha/\alpha_2) \ll 1 \tag{4.17b}$$

cf. (3.12)), $r = \Delta\tilde{p}_1/\Delta\tilde{p}_2$, and $\theta = s_0\Delta t$.

We now combine the barotropic system (4.10) and baroclinic system (4.17) into a single vector equation. Let

$$\psi^n(k,l) = \left(\hat{\eta}^n \quad \frac{\hat{D}^n}{s_0} \quad \frac{\hat{P}^n}{s_0} \quad \hat{\delta}^n \quad \frac{\hat{D}'^n}{s_0} \quad \frac{\hat{P}'^n}{s_0}\right)^{\mathrm{T}}$$

denote a column vector consisting of six unknowns. In the system (4.10), use the continuity equation $\hat{\delta}^{n+1} = \hat{\delta}^n -$

$\Delta t\hat{D}'^{n}$ to eliminate $\hat{\delta}^{n+1}$. The coupled barotropic-baroclinic system can then be written in the form

$$E_1\psi^{n+1} = E_0\psi^n, \tag{4.18a}$$

where

$$E_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ -\rho\theta/r & 0 & 0 & -\rho\theta/r & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{4.18b}$$

and

$$E_0 = \begin{pmatrix} \cos\theta & -\sin\theta & -\phi(1-\cos\theta)/\theta & -\rho(1-\cos\theta) & \rho(\theta-\sin\theta) & 0 \\ \sin\theta & \cos\theta & \phi(\sin\theta)/\theta & \rho\sin\theta & -\rho(1-\cos\theta) & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -\theta & 0 \\ 0 & 0 & 0 & \phi^2(1+a^2\phi^2)/\theta & 1-2a\phi^2 & \phi \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{4.18c}$$

The matrices E_1 and E_0 are functions of (kh, lh). In (4.18c), the quantity θ appears as a denominator; this does not cause a singularity, as $|\phi/\theta| = |f_0/s_0| \le 1$.

4.3. *Stability Analysis*

We now examine the stability of the preceding algorithm. In this discussion we use the same general framework as in Sections 3.3 and 3.4.

We begin with the purely baroclinic system that is obtained from (4.17a) by deleting the forcing term involving $\hat{\eta}$. A calculation shows that the characteristic polynomial for this system is

$$(\lambda - 1)\left[\lambda^2 - \left(2 - 2a\phi^2 - \frac{\rho\theta^2}{r}\right)\lambda + (1 - a\phi^2)^2 + \phi^2\right]. \tag{4.19}$$

A comparison with (4.6), (4.9), and (4.17b) reveals

$$\frac{\rho\theta^2}{r} = 4\nu^2[(\sin kh/2)^2 + (\sin lh/2)^2],$$

where $\nu = c_1\Delta t/h$ denotes the baroclinic Courant number. The root $\lambda_1 = 1$ of (4.19) corresponds to solutions that are constant with respect to time; it follows from the structure of the matrices in (4.17a) that these solutions are a consequence of the conservation of discrete potential vorticity (4.16).

Let λ_2 and λ_3 denote the roots of the quadratic factor in (4.19). If these roots are complex conjugates, then

$$|\lambda_2|^2 = |\lambda_3|^2 = \lambda_2\overline{\lambda}_2 = \lambda_2\lambda_3 = (1 - a\phi^2)^2 + \phi^2,$$

and analyzing the stability of the purely baroclinic system reduces to studying the magnitude of the quantity $(1 - a\phi^2)^2 + \phi^2$. A calculation shows that λ_2 and λ_3 are complex conjugates if

$$\begin{aligned} &2\nu^2[(\sin kh/2)^2 + (\sin lh/2)^2] \\ &\quad < 1 - a\phi^2 + \sqrt{(1 - a\phi^2)^2 + \phi^2}. \end{aligned} \tag{4.20a}$$

If the inequality (4.20a) is satisfied for all (k, l), and if in addition

$$\sqrt{(1 - a\phi^2)^2 + \phi^2} \le 1 \tag{4.20b}$$

for all (k, l), then the purely baroclinic system is stable. In the case where $f = 0$ (i.e., no rotation), (4.20a) is satisfied for all (k, l) if and only if $\nu < 1/\sqrt{2}$.

The magnitude of the eigenvalues is illustrated in Fig. 4.1. It is assumed that (4.20a) holds, so that $|\lambda_2| = |\lambda_3| = \sqrt{(1 - a\phi^2)^2 + \phi^2}$. The figure shows a contour plot of $|\lambda_2|$ as a function of (a, ϕ) for $0 \le a \le 1$ and $0 \le \phi \le 1$. For fixed (k, l), the set of all (a, ϕ) for which $|\lambda_2| \le 1$ is the set of all (a, ϕ) for which the purely baroclinic algorithm is stable; this set will be regarded as the stability region for that algorithm. Most of the right half of Fig. 4.1 is contained in this stability region. The boundary of the region is characterized by $a = 1/(1 + \sqrt{1 - \phi^2})$; for example, if $\phi = 0.1$, then the corresponding value of a is approximately 0.5013.

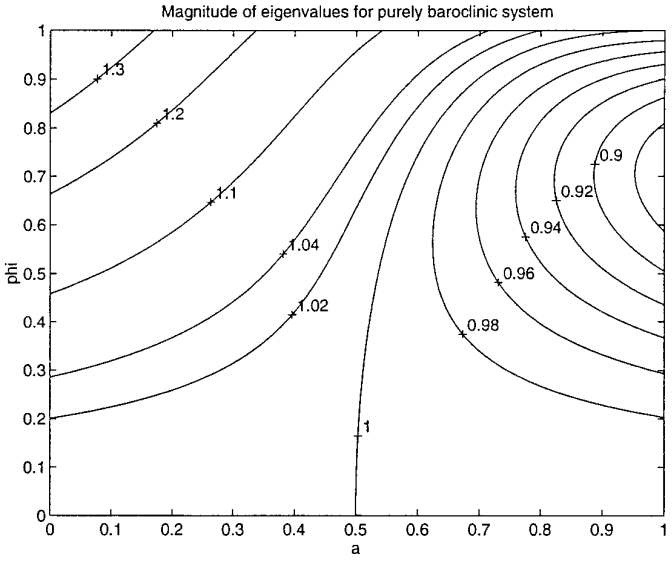

FIG. 4.1. Contour plot of absolute values of eigenvalues for the purely baroclinic system. This plot excludes the root $\lambda = 1$ which results from the conservation of potential vorticity. The two remaining roots are complex conjugates, and their modulus is a function of (a, ϕ). Here, a is the weighting coefficient for the predicted values in (4.12)–(4.13), and $\phi = f_0\Delta t = f\Delta t \cos(kh/2) \cos(lh/2)$.

The quantity ϕ is defined by $\phi = f_0\Delta t = f\Delta t \cos(kh/2) \cos(lh/2)$ (cf. 4.7)). Due to the convexity of the stability region, if (a, ϕ) is in the region when $k = l = 0$, then (a, ϕ) is in the region for all (k, l). If $k = l = 0$, then the discrete baroclinic system (4.11)–(4.13) reduces to a predictor–corrector method for the system of ordinary differential equations $\partial u_1'/\partial t = fv_1'$, $\partial v_1'/\partial t = -fu_1'$, which describes inertial oscillations. The preceding discussion thus shows that the purely baroclinic algorithm is stable if the condition (4.20a) is satisfied for all (k, l) and if the parameter a and the time step Δt are chosen so that the predictor–corrector method is stable when applied to inertial oscillations.

This predictor–corrector scheme was also discussed by Wang and Ikeda [14] during an analysis of time-stepping schemes for inertial oscillations in linearized shallow water (single-layer) models. In that analysis it was assumed that the solutions are constant in space, so that $k = l = 0$. If $a = 0.5$, then this method reduces to the classical Heun scheme [10], which is weakly unstable. An explicit treatment of the Coriolis terms is obtained when $a = 0$, as indicated by (4.13); this method is also unstable. The preceding analysis shows that the predictor–corrector method provides a stable treatment of the Coriolis terms if the parameter a is chosen suitably.

Next consider the coupled barotropic-baroclinic system (4.18). The characteristic polynomial for that system is $p(\lambda) = \det(\lambda E_1 - E_0)$. Because the system is written in a form that expresses the conservation of discrete potential vorticity (cf. (4.8b), (4.16)), the calculation of this determinant reduces immediately to the 4×4 case. We obtain

$$p(\lambda) = (\lambda - 1)^2[p_a(\lambda)p_b(\lambda) + \varepsilon\lambda^2], \qquad (4.21a)$$

where

$$\begin{aligned} p_a(\lambda) &= \lambda^2 - [2 - 2a\phi^2 - \rho\theta^2/r + \rho^2\theta(\theta - \sin\theta)/r]\lambda \\ &\quad + (1 - a\phi^2)^2 + \phi^2 \\ p_b(\lambda) &= \lambda^2 - (2\cos\theta)\lambda + 1 \\ \varepsilon(kh, lh) &= -2(\rho^2/r)\theta\sin\theta(1 - \cos\theta) = O(\Delta\alpha/\alpha_2)^2. \end{aligned} \qquad (4.21b)$$

The coefficients in the polynomial p are functions of (kh, lh) for $|kh| \le \pi$ and $|lh| \le \pi$.

The factor $(\lambda - 1)^2$ corresponds to the equations in (4.18) that express the conservation of discrete barotropic and baroclinic potential vorticity. The quadratic polynomial $p_b(\lambda)$ is associated with the barotropic subsystem. The factor $p_a(\lambda)$ is the same as the quadratic factor in the characteristic polynomial (4.19) for the purely baroclinic system, except for the term involving $\rho^2 = O(\Delta\alpha/\alpha_2)^2$. This term represents some of the coupling between the barotropic and baroclinic subsystems, and the remainder of the coupling is represented by the term $\varepsilon\lambda^2$. For later reference, we also write the characteristic polynomial in the form

$$p(\lambda) = (\lambda - 1)^2 q(\lambda), \qquad (4.22a)$$

where

$$q(\lambda) = p_a(\lambda)p_b(\lambda) + \varepsilon\lambda^2. \qquad (4.22b)$$

In the following discussion, we examine the roots of the polynomial q.

The roots of p_b are $\exp(\pm i\theta)$ and thus lie on the unit circle. The roots of p_a can be analyzed in the same manner as for the purely baroclinic system. In the present case the timestep restriction (4.20a) is modified by a term of order $O(\rho^2) \ll 1$. Subject to that modified restriction, the roots of p_a are complex conjugates with modulus $\sqrt{(1 - a\phi^2)^2 + \phi^2}$. We assume that a and $f\Delta t$ are chosen so that this modulus is at most 1, for all (k, l).

The term $\varepsilon\lambda^2$ represents a small perturbation of the polynomial $p_a(\lambda)p_b(\lambda)$, and we wish to examine the effects of this perturbation. If the roots of p_a and p_b are well separated, then an argument similar to that in Lemma 3.3 shows that the roots of the polynomial $p_a(\lambda)p_b(\lambda)$ are perturbed by a magnitude of at most $O(\varepsilon) = O(\Delta\alpha/\alpha_2)^2$. If two roots coincide or nearly coincide, then this estimate

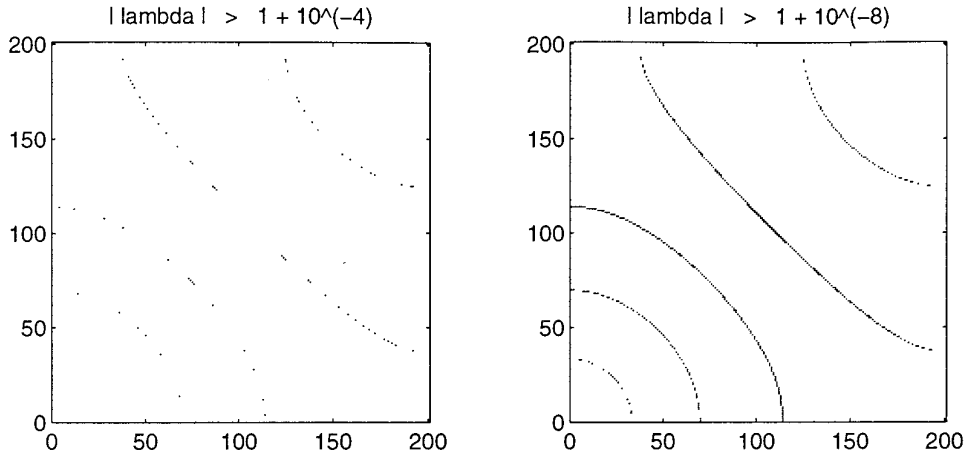

FIG. 4.2. Eigenvalues for the coupled barotropic-baroclinic system (4.18). These plots show locations of values of (kh, lh) for which $|\lambda| > 1$. Here, kh and lh are sampled in increments of $\pi/200$. The horizontal and vertical axes correspond to kh and lh, respectively, for $0 < kh \leq \pi$ and $0 < lh \leq \pi$; the labels on the axes are array indices. The left frame shows the points for which $\max|\lambda| > 1 + 10^{-4}$, and the right frame shows the points for which $\max|\lambda| > 1 + 10^{-8}$. Model parameters are $\nu = c_1\Delta t/h = 0.5$, $f\Delta t = 0.1$, $a = 0.51$, $\Delta\alpha/\alpha_2 = 0.01$, $\Delta\tilde{p}_1/\tilde{p}_2 = 0.25$, and $\Delta\tilde{p}_2/\tilde{p}_2 = 0.75$.

breaks down, and the roots are perturbed by at most $O(\sqrt{\varepsilon}) = O(\Delta\alpha/\alpha_2)$.

The direction of the perturbation determines whether the eigenvalues leave the unit disk. However, the argument used in Section 3.4 cannot be used here, as the constant term in p is not exactly equal to 1 in the present case. Instead, we compute numerical values of eigenvalues for an example.

We consider a two-layer model for which $\Delta\tilde{p}_1/\tilde{p}_2 = 0.25$, $\Delta\tilde{p}_2/\tilde{p}_2 = 0.75$, and $\Delta\alpha/\alpha_2 = 0.01$. We also assume that the time and space steps are chosen so that $f\Delta t = 0.1$ and the baroclinic Courant number is $\nu = c_1\Delta t/h = 0.5$. The Courant number can be written in the form $\nu = (c_1/fh)(f\Delta t)$; with the present choice of parameters, we then have $c_1/fh = 5$. The quantity c_1/f is the baroclinic Rossby radius, so the present choice of parameters corresponds to five grid intervals per baroclinic Rossby radius. In the weighted average that appears in the predictor–corrector method for the Coriolis terms, we use $a = 0.51$. According to the earlier discussion, the purely baroclinic algorithm is then stable; in particular, the magnitude $\sqrt{(1 - a\phi^2)^2 + \phi^2}$ of the roots of p_a varies between 1 and approximately 0.999913 as kh and lh vary.

With the above choices of parameters, we compute eigenvalues of the problem $\lambda E_1 w = E_0 w$, where E_0 and E_1 are given in (4.18). The matrices E_0 and E_1 are functions of (kh, lh) for $|kh| \leq \pi$ and $|lh| \leq \pi$. Due to symmetries, we can restrict attention to the domain $0 \leq kh \leq \pi$, $0 \leq$

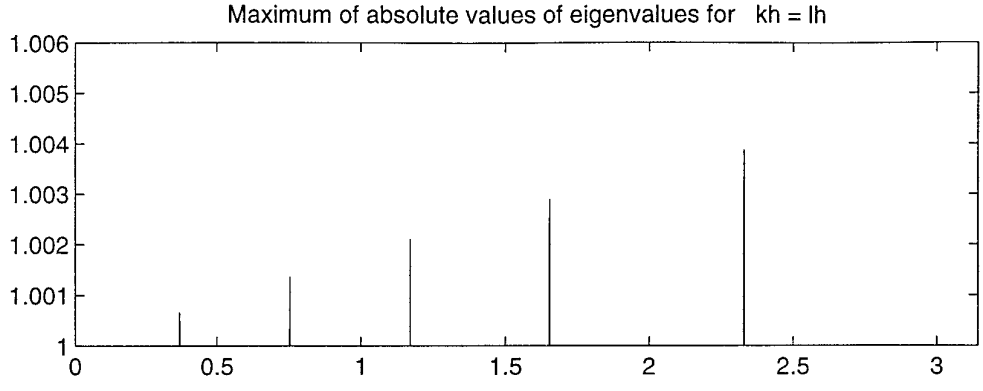

FIG. 4.3. Cross section of Fig. 4.2 corresponding to $kh = lh$. The vertical coordinate is the maximum of the absolute values of eigenvalues. The horizontal coordinate is kh, with kh sampled in increments of $\pi/18000$. The model parameters are the same as in Fig. 4.2. Five spikes are shown in the plot; the short vertical lines located at the axis labels are tick marks. This figure illustrates how the curved patterns of dots in Fig. 4.2 represent the locations of extremely thin walls. These walls are the higher-dimensional analogues of the spikes seen in Figs. 3.2 and 3.3.

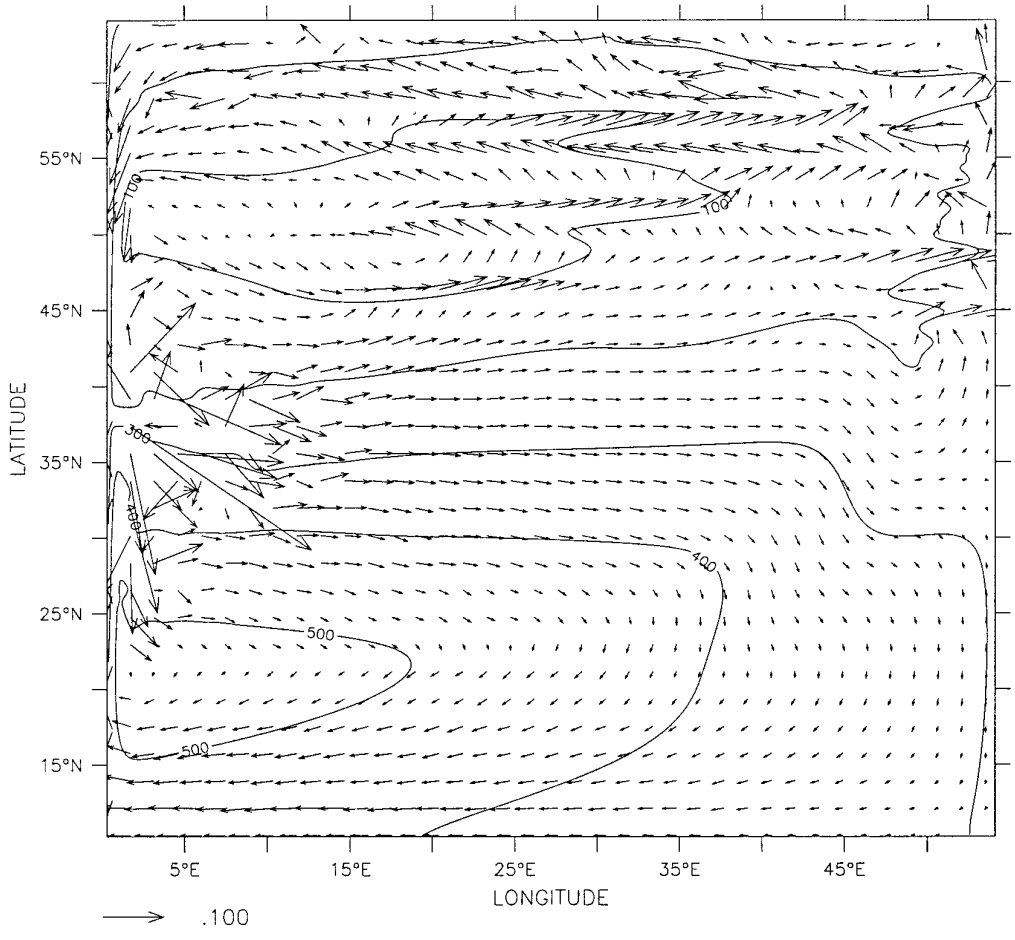

FIG. 5.1. Large-scale flow in the top layer. The arrows represent the time-averaged total velocity $\mathbf{u} = \bar{\mathbf{u}} + \mathbf{u}'$ in the top layer; the velocity vectors are shown only for every tenth grid point, and the reference vector has units of meters per second. The curves are contours of time-averaged total pressure $p = (1 + \eta)p'$ at the bottom of the top layer. The contours are in units of 10^4 Pascals, which corresponds roughly to one meter of depth.

$lh \leq \pi$. For a finite set of (kh, lh), we compute the maximum of the absolute values of the eigenvalues.

For the graphs shown in Fig. 4.2, we use values of kh and lh varying independently from $\pi/200$ to π in increments of $\pi/200$. In these graphs the horizontal and vertical axes correspond to kh and lh, respectively, although the labels on the axes are array indices. The dots in the left frame in Fig. 4.2 indicate the locations of those values of (kh, lh) for which $\max |\lambda| > 1 + 10^{-4}$; the right frame indicates the locations for which $\max |\lambda| > 1 + 10^{-8}$. These points lie along five arcs in the (kh, lh) plane.

The coefficients in the characteristic polynomial (4.21) are smoothly varying functions of (kh, lh). If the values of (kh, lh) are restricted to rays through the origin in the (kh, lh) plane, then the behavior of the roots must vary smoothly with the directions of the rays. In order to illustrate further the properties of the roots, we can therefore examine the behavior along one such ray. Figure 4.3 shows a high-resolution plot along the ray for which $kh = lh$. In this plot the vertical coordinate is the maximum of the absolute values of the eigenvalues, and the horizontal coordinate is kh, with kh sampled in increments of $\pi/18000$.

The five arcs in Fig. 4.2 thus represent the locations of thin curving walls rising upward from the (kh, lh) plane; the spikes in Fig. 4.3 are cross sections of those walls. These walls are higher dimensional analogues of the spikes encountered in the analysis of the one-dimensional case given in Section 3.

5. NUMERICAL COMPUTATIONS

Here we outline the results of a numerical simulation involving a nonlinear ocean model that employs the barotropic-baroclinic splitting developed in the present paper.

In this computation, the spatial domain occupies a spherical sector, with the latitude varying from 10.18° N to 64.54°

N and the longitude varying over a range of 54.36°. Spherical coordinates are used, with a horizontal grid spacing of 0.18° in both dimensions. The bottom of the fluid domain is assumed level. The fluid consists of three layers having specific volumes of 0.975×10^{-3}, 0.974×10^{-3}, and 0.973×10^{-3} m^3/kg and initial thickness of 300, 700, and 3000 m, respectively. No mass exchange between the layers is permitted. The system is driven by a steady zonal (east–west) wind stress of the form $\tau_\phi = \tau_0 \cos(2\pi(\theta - 37.36)/54.36)$, with $\tau_0 = 0.1$ N m^{-2}; the wind stress is applied to the upper layer as a body force. The baroclinic time step is 2000 s, and the barotropic time step is 40 s. In the predictor–corrector scheme analyzed in Section 4, the coefficient a is chosen to be 0.51.

Physical dissipation is represented by a horizontal viscosity term having coefficient 100 m^2 s^{-1}. This value is chosen in order to resolve adequately the Munk boundary layer along the western boundary. In addition, some numerical dissipation is generated by an upwind method that is used to solve the baroclinic continuity equation. However, no time filtering is used; geophysical fluid calculations often use time smoothing at each time step in order to suppress numerical noise [2, 7], but this procedure is not employed in the present computation.

The computation was run for 3750 days of simulated time. Snapshots of flow fields at various fixed times showed stable behavior. In particular, the computations showed the development of an anticyclonic subtropical gyre, a cyclonic subpolar gyre, and an intense western boundary current. Superimposed on these large-scale, slowly varying features were various shorter-period disturbances.

Figures 5.1 and 5.2 show the large scale patterns in the flow. In order to show these features more clearly, we suppress the transient disturbances by computing time averages of the final computed fields over the last 375 days of the simulation; this averaging is very different from the process of reducing numerical noise by time-smoothing at each time step during the course of the computation. In

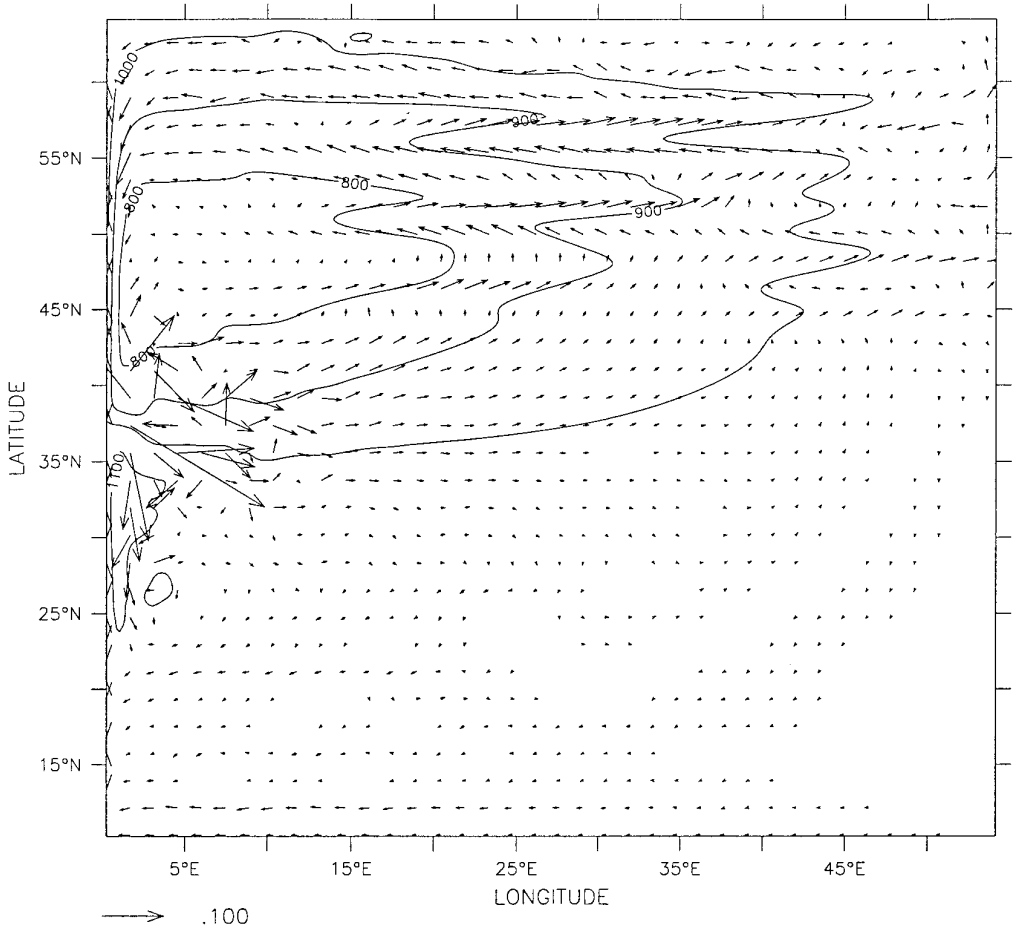

FIG. 5.2. Large-scale flow in the second layer. The arrows represent the time-averaged velocity in the second layer, and the curves are contours of time-averaged total pressure at the bottom of the second layer.

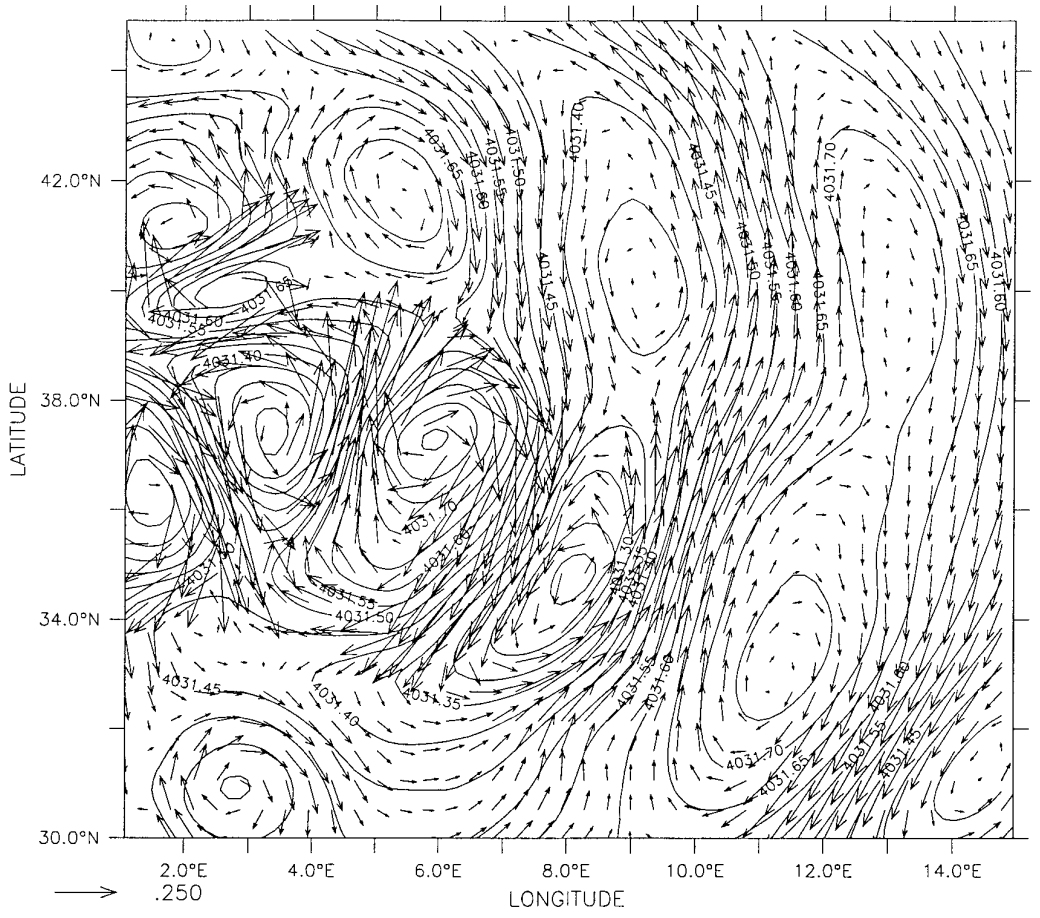

FIG. 5.3. Close-up view of the flow near the western boundary at day 3750. The arrows represent the instantaneous vertically-averaged velocity $\bar{\mathbf{u}}$; velocity vectors are shown at every other grid point. The curves are contours of total pressure at the bottom of the fluid. The units of pressure are the same as in Figs. 5.1 and 5.2.

Fig. 5.1, the arrows show the time-averaged total velocity field $\mathbf{u} = \bar{\mathbf{u}} + \mathbf{u}'$ in the top layer, and the solid curves are contours of the time-averaged total pressure $p = (1 + \eta)p'$ at the bottom of the top layer. Figure 5.2 shows the total velocity field in the second layer and the total pressure at the bottom of the second layer. In the third layer, which is not known here, the flow is relatively quiescent, except near the western boundary. Over much of the region, the velocity vectors align closely with the contours of pressure, which indicates a near-geostrophic balance.

Figure 5.3 is a snapshot of the flow near the western boundary at day 3750. The arrows indicate the vertically averaged velocity $\bar{\mathbf{u}}$, and the curves are contours of the total pressure at the bottom of the fluid. The figure shows numerous eddies that are generated due to shear instabilities in the western boundary current.

The general features seen in this simulation are to be expected from the wind forcing used here. These numerical results suggest the practical stability of the proposed barotropic-baroclinic time splitting.

6. SUMMARY AND CONCLUSIONS

The goal of a barotropic-baroclinic time splitting is to separate the fast and slow motions into separate subsystems. However, if errors in the splitting cause the baroclinic system to admit significant energy in fast motions, then the algorithm could be unstable when the baroclinic equations are solved explicitly with long time steps.

In the present paper we modify the splitting developed in [1]. The main modification is in the barotropic momentum equation, which describes the time evolution of the vertically averaged velocity $\bar{\mathbf{u}}$. In [1] this equation is forced by a pressure gradient that is equivalent to the one that appears in the constant-density shallow water equations. We replace this term by $\overline{\nabla M}$, the vertical average of the Montgomery potential. The gradient term in [1] is an approximation to $\overline{\nabla M}$; the present formulation takes into account the density variations over the depth of the fluid.

We then show that this modification leads to major improvements in stability. The present approach yields algo-

rithms that are stable in the linearized case, except for wavenumbers that are confined to a finite number of extremely small neighborhoods. This splitting yielded good results in a numerical simulation involving nonlinearity, level bottom topography, physically reasonable dissipation, and no time filtering.

ACKNOWLEDGMENTS

We thank Andrew Bennett and Scott Springer for useful conversations regarding the matters discussed in this paper. We also thank Springer for performing the numerical computations described in Section 5. Higdon's work was supported by National Science Foundation Grant DMS-9407509. De Szoeke's work was supported by National Science Foundation Grants OCE-9319892 and OCE-9402891.

REFERENCES

1. R. Bleck and L. T. Smith, A wind-driven isopycnic coordinate model of the north and equatorial Atlantic Ocean 1. Model development and supporting experiments, *J. Geophys. Res. C* **95,** 3273 (1990).
2. R. Bleck, C. Rooth, D. Hu, and L. T. Smith, Salinity-driven thermocline transients in a wind- and thermohaline-forced isopycnic coordinate model of the North Atlantic, *J. Phys. Oceanogr.* **22,** 1486 (1992).
3. K. Bryan, A numerical method for the study of the circulation of the world ocean, *J. Comput. Phys.* **4,** 347 (1969).
4. R. E. Davis, Diapycnal mixing in the ocean: equations for large-scale budgets, *J. Phys. Oceanogr.* **24,** 777 (1994).
5. J. K. Dukowicz and R. D. Smith, Implicit free-surface method for the Bryan–Cox–Semtner ocean model, *J. Geophys. Res. C* **99,** 7991 (1994).
6. A. E. Gill, *Atmosphere-Ocean Dynamics* (Academic Press, San Diego, 1982).
7. G. J. Haltiner and R. T. Williams, *Numerical Prediction and Dynamic Meteorology* (Wiley, New York, 1980).
8. R. L. Higdon and A. F. Bennett, Stability analysis of operator splitting for large-scale ocean modeling, *J. Comput. Phys.* **123,** 311 (1996).
9. P. D. Killworth, D. Stainforth, D. J. Webb, and S. M. Paterson, The development of a free-surface Bryan–Cox–Semtner ocean model, *J. Phys. Oceanogr.* **21,** 1333 (1991).
10. F. Mesinger and A. Arakawa, *Numerical Methods Used in Atmospheric Models,* GARP Publications Series No. 17, Vol. 1 (WMO-ICSU Joint Organizing Committee, Geneva, 1976).
11. J. M. Oberhuber, Simulation of the Atlantic circulation with a coupled sea ice—mixed layer—isopycnal general circulation model. Part I. Model description, *J. Phys. Oceanogr.* **23,** 808 (1993).
12. J. Pedlosky, *Geophysical Fluid Dynamics,* 2nd ed. (Springer-Verlag, New York, 1987).
13. A. J. Semtner, Finite-difference formulation of a world ocean model, in *Advanced Physical Oceanographic Numerical Modelling,* edited by J. J. O'Brien (Reidel, Norwell, MA, 1986), p. 187.
14. J. Wang and M. Ikeda, On inertial stability and phase error of time integration schemes in ocean general circulation models, *Mon. Weather Rev.,* to appear.