general integral constraints that these eddies must satisfy. There has as yet been no complete analysis on the detailed internal structure and interaction characteristics of these eddies with the surrounding fluid. The principal purpose of this paper is to present a unified asymptotic theory describing the propagation and structure of coherent cold-core domes on a sloping bottom and their (dynamical and thermodynamical) interaction with the surrounding fluid.

One important conclusion of the analysis presented by the Nof (1983) study was that, in the absence of any interaction of the eddy with the surrounding environment, the along-shelf propagation speed was given simply by $g's^*/f_0$ (hereafter called the *Nof speed*), where g', s^* and f_0 are the reduced gravity, the constant topographic slope and the Coriolis parameter, respectively, and that the transverse or cross-shelf velocity was identically zero. Note that the Nof speed is independent of the detailed spatial structure of the eddy and depends only on the density difference between the eddy and the surrounding fluid, the bottom slope and the constant Coriolis parameter. The order of magnitude of the along-shelf speed predicted by the Nof formula is in qualitative agreement with the observations presented by Houghton *et al.* (1982). However, the data taken from the rotating-tank experiments described by Mory *et al.* (1987) did not completely agree with the Nof theory. The experiments clearly indicated that a non-negligible cross-shelf drift was present. Another interesting observation was that the azimuthal velocity field above the eddy in the surrounding water was appreciable and was probably larger in magnitude than the swirl velocity in the eddy interior. This last observation strongly suggests that it is important to include the dynamical response of the surrounding fluid in determining the evolution of an initially isolated cold-core eddy.

Mory *et al.* (1987) suggested various mechanisms which may account for the above discrepancies. The first proposal centred on the action of an induced wave drag on the dome due to the excitation of topographic vorticity waves generated by the baroclinic vortex-tube compression associated with the passage of the cold dome through the surrounding fluid. Earlier work by Flierl (1984*a*) had shown how this drag could result in a mean meridional drift for surface trapped warm eddies on a β-plane.

A second mechanism suggested by Mory *et al.* (1987) was that frictional forces between the eddy and the bottom (which in the experimental configuration may have been particularly important) could result in cross-shelf motion. Mory *et al.* (1987) were unable, however, to distinguish which of these two mechanisms was dynamically dominant in their experiment.

A third effect that may have dynamical importance, especially in the oceanographic context, is the process of ventilation between the cold eddy and the surrounding relatively warmer slope water. Ou & Houghton (1982) argued that this effect was of particular importance during the evolution of the cold dome observed on the Middle Atlantic Bight during 1979. Although the heating rate for the Middle Atlantic Bight cold eddy was a function of the distance the eddy travelled, Houghton *et al.* (1982) estimated that on average it was approximately between 0.5 °C and 2.0 °C per month. Based on a representative eddy temperature of about 4 °C a Newtonian heating timescale for the cold pool will be on the order of 10^7 s or about 3–4 months. It is possible to show (see §3.2) that this (eddy to slope water) mass conversion timescale is comparable with the timescale of the topographic Rossby waves induced by the vortex-tube compression in the slope water associated with the propagating eddy. We may expect, therefore, that the heating processes documented by Houghton *et al.* (1982) and Ou & Houghton (1982) will lead to a non-steady

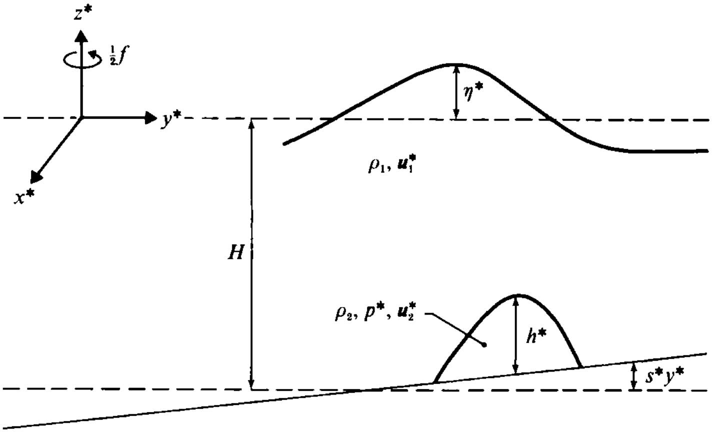


FIGURE 1. Geometry of the two-layer model used in this paper.

dynamical interaction between the cold eddy and the relatively warmer slope water involving diabatic heating, vortex-tube compression and the radiation of topographic Rossby waves.

In this paper a theory is developed to describe the propagation of a coherent cold-core eddy on a linearly sloping bottom. The theory will incorporate both dynamical interactions of the eddy with the surrounding fluid and ventilation processes. Based on 'synoptic' parameter values suggested by the oceanographic and experimental data, the new model equations that we derive to study the above processes correspond to strongly interacting 'hybrid' quasi-geostrophic, intermediate-lengthscale geostrophic dynamics (see Charney & Flierl 1981). Specifically, we argue that the dynamics of the surrounding slope water is essentially quasi-geostrophic but that the eddy dynamics while geostrophic is *not* quasi-geostrophic because eddy thickness changes are not small in comparison with the eddy-scale height itself. The above model is derived in a formal asymptotic expansion based on two-layer shallow-water theory assuming a small (appropriately scaled) shelf slope parameter.

The parameterization we adopt to describe the ventilation is the CI or cross-interfacial mass flux model of Dewar (1987, 1988*a*, *b*). Physically, this parameterization models the ventilation process as a continuous conversion of cold eddy water into relatively warmer slope water and is convenient for the two-layer model considered here. There have been other ventilation parameterizations suggested. For example, Chapman & Nof (1988) have proposed a parameterization for continuously stratified eddy models based on potential vorticity conservation principles.

The nonlinear model equations (see (3.14) and (3.15)) contain two non-dimensional parameters: an 'interaction' parameter (denoted μ) which physically measures the ratio of eddy-induced vortex-tube compression in the surrounding slope water to the background vorticity gradient associated with the shelf slope; and a diabatic heating parameter (denoted β) which measures the ratio of the timescale associated with the Nof translation speed to the ventilation timescale. In the absence of ventilation processes these equations admit an exact radially symmetric isolated-eddy slope-water solution which travels at the Nof speed and satisfies the Stern integral constraint (Mory 1985). However, the presence of the diabatic ventilation terms leads to temporal evolution in the eddy height which in turn, it is shown, leads to a weak topographic wave field in the slope water and a wave-drag-induced upslope drift in the propagating eddy.

The plan of the paper is as follows. In §2 the non-dimensional problem is formulated based on shallow-water theory and some preliminary remarks are made. In §3 the small-topographic-slope approximation is introduced and the 'quasi-geostrophic, intermediate lengthscale geostrophic' model is derived. In §4, the multiple-scale asymptotic theory is developed to describe the propagation and evolution of an initially isolated eddy due to the presence of ventilation processes and the subsequent emergence of the topographic wave field. In §5, the various salient features of our theory are illustrated with a simple example which assumes that the eddy has a parabolic configuration. The paper is summarized and some concluding remarks are made in §6.

2. Problem formulation

The basic model we assume is an f-plane two-layer system (both layers are assumed hydrostatic, homogeneous and incompressible) with a linearly varying bottom slope (see figure 1). The *dimensional* equations of motion for the upper layer or slope water (layer 1) can be written in the form

$$[\partial_{t^*}+\boldsymbol{u}_1^*\cdot\nabla^*]\boldsymbol{u}_1^*+f\hat{\boldsymbol{e}}_3\times\boldsymbol{u}_1^*+g\nabla^*\eta^*=\boldsymbol{0}, \tag{2.1a}$$

$$(\eta^*-h^*)_{t^*}+\nabla^*\cdot[\boldsymbol{u}_1^*(H+\eta^*-h^*-s^*y^*)]=-\mathscr{H}^*(x^*,y^*,t^*), \tag{2.1b}$$

where $\boldsymbol{u}_1^*=(u_1^*,v_1^*)$ is the horizontal velocity field, and f, g, η^*, h^*, H, and s^* are the constant Coriolis parameter, gravitational acceleration, reduced layer-1 pressure, cold dome thickness, 'mean' layer-1 thickness and bottom slope parameter, respectively. The dimensional coordinates are (x^*,y^*) and t^* is dimensional time. Subscripts x^*, y^* and t^* indicate partial differentiation, and $\nabla^*=(\partial_{x^*},\partial_{y^*})$. The term $-\mathscr{H}^*(x^*,y^*,t^*)$ in (2.1b) models the slope-water mass gain associated with the conversion of cold eddy water into relatively warmer slope water as a result of the diabatic processes. We will comment more completely on this term below.

The *dimensional* equations of motion describing the evolution of the cold-core eddy can be written in the form

$$[\partial_{t^*}+\boldsymbol{u}_2^*\cdot\nabla^*]\boldsymbol{u}_2^*+f\hat{\boldsymbol{e}}_3\times\boldsymbol{u}_2^*+\frac{1}{\rho_2}\nabla^*p^*=0, \tag{2.2a}$$

$$h_t^*+\nabla^*\cdot[\boldsymbol{u}_2^*h^*]=\mathscr{H}^*(x^*,y^*,t^*), \tag{2.2b}$$

where the notation convention is similar to that used in (2.1), and where p^* is the dynamic pressure field in the eddy.

The diabatic heating of the cold pool is determined by the function $\mathscr{H}^*(x^*,y^*,t^*)$ in the eddy continuity equation (2.2b). Our parameterization for this term is based on the cross-interfacial (CI) model proposed by Dewar (1987, 1988a, b) for dynamically active warm rings given by

$$\mathscr{H}^*(x^*,y^*,t^*)=-\beta^*h^*. \tag{2.2c}$$

In this formulation (2.2c) is a simple Newtonian cooling/heating law which models the conversion of relatively colder eddy water to warmer slope water on a timescale determined by $1/\beta^*$. The reader is referred to Dewar (1987, 1988a, b) for a complete account of the physical principles for the model (2.2c). However, we point out here that this model implicitly assumes that the fluid must exist in either the cold-eddy density state or the relatively warmer slope-water density state. Accordingly, if the cold eddy is heated by the surrounding slope water an appropriate volume of eddy water is converted into slope water and, as a result, the effect of the warming will be to induce a mass loss in the cold eddy which is modelled by (2.2b, c). It is important to add, however, that other eddy diabatic heating/cooling parameterizations have been suggested (e.g. Chapman & Nof 1988).

The system of equations (2.1) and (2.2) is closed with the requirement that the dynamic pressure be continuous across the eddy–slope-water interface, i.e.

$$p^* = \rho_1 g\eta^* + \rho_2 g'(h^* + s^* y^*), \tag{2.3}$$

where ρ_1, ρ_2 and g' are the layer 1 and 2 densities, and the reduced gravity $g' = (\rho_2 - \rho_1)g/\rho_2 > 0$ (stable stratification), respectively.

Further analysis is facilitated by reformulating the governing equations into non-dimensional form. The non-dimensionalization scheme we adopt is motivated by (but is somewhat different from) the scalings used in Flierl (1984*b*). The non-dimensional (unasterisked) variables are given by

$$\left.\begin{aligned} (x^*, y^*) = L(x,y), \quad t^* = (fL/g's^*)\,t, \quad h^* = h_0 h, \\ \mathbf{u}_1^* = \delta f L \mathbf{u}_1, \quad \eta^* = \delta(fL)^2 g^{-1}\eta, \quad \mathbf{u}_2^* = g's^* f^{-1}\mathbf{u}_2, \\ p^* = \rho_2 L g' s^* p, \end{aligned}\right\} \tag{2.4}$$

where $L = (g'H)^{\frac{1}{2}}/f$, $h_0 = h^*(x^* = y^* = t^* = 0)$ and $\delta = h_0/H$. The horizontal length-scale is the internal deformation radius and time has been scaled advectively based on the theoretical Nof translation speed. The scalings for $\mathbf{u}_1^*$ and η^* are based on the assumption that changes in upper-layer relative vorticity are due to vortex compression with $\mathbf{u}_1^*$ and η^* in geostrophic balance. As we will see, the above scalings will imply that the upper-layer dynamics will be quasi-geostrophic under a small-slope-parameter assumption. The lower-layer dynamics will be primarily geostrophic. The scaling for p^* is geostrophic and the scaling for $\mathbf{u}_2^*$ is the Nof translation speed. However, changes in the eddy thickness $h(x,y,t)$ will be as important as the horizontal divergence term $\nabla\cdot\mathbf{u}_2$ in the continuity equation. Thus while momentum advection can be ignored for the eddy, the evolution of $h(x,y,t)$ is strongly coupled to $\mathbf{u}_2$. These scalings are similar to those presented in Whitehead *et al.* (1990).

Substitution of (2.4) into (2.1), (2.2) and (2.3) gives the non-dimensional problem

$$s\mathbf{u}_{1_t} + \delta(\mathbf{u}_1\cdot\nabla)\,\mathbf{u}_1 + \hat{\mathbf{e}}_3\times\mathbf{u}_1 + \nabla\eta = 0, \tag{2.5a}$$

$$s(h - g'\eta/g)_t + \nabla\cdot[\mathbf{u}_1(\delta h + sy - 1 - g'\eta/g)] = -s\beta h, \tag{2.5b}$$

$$s\mathbf{u}_{2_t} + s(\mathbf{u}_2\cdot\nabla)\,\mathbf{u}_2 + \hat{\mathbf{e}}_3\times\mathbf{u}_2 + \nabla p = 0, \tag{2.5c}$$

$$h_t + \nabla\cdot(h\mathbf{u}_2) = -\beta h, \tag{2.5d}$$

$$(1 - g'/g)\,\delta\eta + \delta h + s(y-p) = 0, \tag{2.5e}$$

where $\beta = \beta^*(fs)^{-1}$, and $s = s^*L/H$. We will estimate the magnitudes of δ, s and β in §3.2.

The following boundary conditions are imposed on the model. Suppose the projection of the intersection of the eddy with the sloping bottom on the plane $z = 0$ is given by $\phi(x,y,t) = 0$. The kinematic condition, which physically expresses the fact that a fluid parcel on the deforming eddy boundary remains on the boundary, is given by

$$\phi_t + \mathbf{u}_2\cdot\nabla\phi = 0 \quad \text{on} \quad \phi(x,y,t) = 0. \tag{2.6}$$

The eddy thickness must satisfy

$$h(x,y,t) = 0 \quad \text{on} \quad \phi(x,y,t) = 0. \tag{2.7}$$

For the upper layer we will impose the additional constraint that $r^{\frac{1}{2}}\eta \rightarrow 0$ as $r\rightarrow\infty$ 'ahead' of the propagating eddy. As pointed out by Miles (1968), this is the correct 'no upstream waves' condition (see also McKee 1971; McCartney 1975; Flierl 1984*a,b*).

3. Analysis for small topographic slope

3.1. *Modified Nof formulae*

It is possible to see how the theoretical expressions that Nof (1983) derived for the along-shelf and cross-shelf velocities are modified by the presence of a dynamically active upper layer as follows. Let us suppose that a steadily travelling ansatz can be made in the form $\boldsymbol{u}_2 = \boldsymbol{u}_2(\xi,\zeta)$, $h = h(\xi,\zeta)$, $\eta = \eta(\xi,\zeta)$ and $\phi = \phi(\xi,\zeta)$ with $\xi = x - c_x t$ and $\zeta = y - c_y t$ where $\boldsymbol{c} = (c_x, c_y)$ is the translation velocity vector. In the absence of any diabatic processes, the eddy momentum and continuity equations can be written in the form

$$s^2[(\boldsymbol{u}_2-\boldsymbol{c})\cdot\boldsymbol{\nabla}](\boldsymbol{u}_2-\boldsymbol{c})+s\hat{\boldsymbol{e}}_3\times\boldsymbol{u}_2+\boldsymbol{\nabla}(\delta\eta+\delta h+s\zeta)=\boldsymbol{0}, \tag{3.1}$$

$$\boldsymbol{\nabla}\cdot[h(\boldsymbol{u}_2-\boldsymbol{c})]=0, \tag{3.2}$$

where $\boldsymbol{\nabla} = (\partial_\xi, \partial_\zeta)$, respectively. The eddy boundary conditions (2.6) and (2.7) can be written in the respective forms,

$$\left.\begin{aligned}(\boldsymbol{u}_2-\boldsymbol{c})\cdot\boldsymbol{\nabla}\phi&=0,\\ h&=0,\end{aligned}\right\}\quad\text{on}\quad \phi(\xi,\zeta)=0. \qquad \begin{aligned}&(3.3a)\\&(3.3b)\end{aligned}$$

If the product (3.1)$\cdot h$ is integrated over the eddy region $R = \{(\xi,\zeta)\,|\,\partial R = \phi\}$ it follows that

$$\iint_R \hat{\boldsymbol{e}}_3\times(h\boldsymbol{u}_2)+\hat{\boldsymbol{e}}_\zeta\iint_R h=-\frac{\delta}{s}\iint_R h\boldsymbol{\nabla}\eta. \tag{3.4}$$

In deriving (3.4) we have already used the fact that the integral of $h(\xi,\zeta)$ times the nonlinear terms in (3.1) is identically zero since if we integrate by parts these terms it follows that

$$\iint_R h(\boldsymbol{u}_2-\boldsymbol{c})\cdot\boldsymbol{\nabla}(\boldsymbol{u}_2-\boldsymbol{c})=\oint_{\phi=0} h(\boldsymbol{u}_2-\boldsymbol{c})\cdot\boldsymbol{n}(\boldsymbol{u}_2-\boldsymbol{c})-\iint_R(\boldsymbol{u}_2-\boldsymbol{c})\boldsymbol{\nabla}\cdot[h(\boldsymbol{u}_2-\boldsymbol{c})],$$

and the first term on the right-hand side is zero because of the boundary condition (3.3*b*) and the second term on the right-hand side is zero because of (3.2); see also Nof (1983).

From the continuity equation (3.2) we can infer the existence of a co-moving mass transport stream function $\psi = \psi(\xi,\zeta)$ satisfying

$$h(\boldsymbol{u}_2-\boldsymbol{c})=\hat{\boldsymbol{e}}_3\times\boldsymbol{\nabla}\psi. \tag{3.5}$$

Note that it follows from (3.3*a*) that ψ is constant on $\phi = 0$, i.e. the eddy boundary forms a streamline in the co-moving frame of reference. Substituting (3.5) into (3.4) yields the relation

$$(\hat{\boldsymbol{e}}_\zeta+\hat{\boldsymbol{e}}_3\times\boldsymbol{c})=-\frac{\delta}{s}\iint_R h\boldsymbol{\nabla}\eta\bigg/\iint_R h. \tag{3.6}$$

It follows, immediately, that the components of the translation velocity vector will be given by

$$c_x=-1-\frac{\delta}{s}\iint_R h\eta_\zeta\bigg/\iint_R h, \tag{3.7}$$

$$c_y=\frac{\delta}{s}\iint_R h\eta_\xi\bigg/\iint_R h. \tag{3.8}$$

In the no interaction limit ($\delta \rightarrow 0$), these expressions reduce exactly to the Nof formulae.

3.2. *Approximation of small topographic slope*

The modified Nof formulae (3.7) and (3.8) state that the interaction between the eddy and the surrounding fluid is $O(\delta/s)$ in comparison with the leading-order (non-interaction) terms. In this section we shall examine the small-slope asymptotic limit defined by

$$\delta = \mu s, \tag{3.9}$$

where $0 < s \ll 1$ and $\mu = O(1)$. This limit will correspond to a strongly interacting eddy–slope water configuration in the sense that although $0 < s \approx \delta \ll 1$ the fact that $\mu = O(1)$ in (2.5*e*) will mean that the leading-order dynamics for the eddy height cannot be decoupled from the dynamics for the geostrophic pressure in the slope water (see (3.14)). Physically, this limit has a simple interpretation. The scaled topographic slope parameter $s = s^*L/H$ can be rearranged into the form $s = (s^*g'/f)/(g'H)^{\frac{1}{2}}$. Thus the parameter s can be interpreted as the ratio of the Nof speed to the phase or group speed of the long baroclinic gravity wave solutions in the model. The limit $s \rightarrow 0^+$ can be viewed as a low-bandpass filter which will effectively remove the long baroclinic gravity waves in the slope water and focus attention on the vorticity wave processes.

The observations reported in Armi & D'Asaro (1980) and Houghton *et al.* (1982) suggest that the small (non-dimensional)-slope approximation and the scaling (3.9) are physically relevant. For example, the observations of the cold pool reported by Houghton *et al.* (1982) correspond to approximate parameter values of $s^* \approx 1.2$ m/km, $L \approx 15$ km, $h_0 \approx 30$–40 m and $H \approx 200$–300 m suggesting $s \approx 4 \times 10^{-2}$, $\delta \approx 2 \times 10^{-2}$ and consequently that $\mu = 0(1)$. The oceanographic observations of Armi & D'Asaro and the rotating-tank data of Mory *et al.* scale similarly. The magnitude of the non-dimensional heating parameter β can be estimated as follows. Based on a heating rate of about 2 °C/month and a scale eddy temperature of 4 °C, it follows that $\beta^* \approx 2 \times 10^{-7}\ \mathrm{s}^{-1}$. Based on $f \approx 10^{-4}\ \mathrm{s}^{-1}$ and $s \approx 4 \times 10^{-2}$ it therefore follows that $\beta = \beta^*(fs)^{-1} \approx 10^{-1}$. Consequently, as a rough first approximation, we see that $s \approx \delta$ and that heating processes, while relatively weak, are an order of magnitude larger than the ageostrophic terms in the eddy momentum balance.

Substitution of (3.9) into (2.5) yields the following set of equations for the eddy and surrounding slope-water problem:

$$\hat{\boldsymbol{e}}_3 \times \boldsymbol{u}_1 + \nabla \eta = -s\boldsymbol{u}_{1_t} - s\mu(\boldsymbol{u}_1 \cdot \nabla)\boldsymbol{u}_1, \tag{3.10a}$$

$$\nabla \cdot \boldsymbol{u}_1 = sh_t + s\nabla \cdot (y\boldsymbol{u}_1) + \mu s \nabla \cdot (h\boldsymbol{u}_1) + s\beta h, \tag{3.10b}$$

$$\hat{\boldsymbol{e}}_3 \times \boldsymbol{u}_2 + \hat{\boldsymbol{e}}_2 + \mu \nabla (h + \eta) = -s\boldsymbol{u}_{2_t} - s(\boldsymbol{u}_2 \cdot \nabla)\boldsymbol{u}_2, \tag{3.11a}$$

$$h_t + \nabla \cdot (h\boldsymbol{u}_2) = -\beta h, \tag{3.11b}$$

subject to the eddy boundary conditions

$$\left.\begin{aligned} \phi_t + (\boldsymbol{u}_2 \cdot \nabla)\phi &= 0, \\ h(x,y,t) &= 0, \end{aligned}\right\} \quad \text{on} \quad \phi(x,y,t) = 0, \qquad \begin{aligned} &(3.12a) \\ &(3.12b) \end{aligned}$$

and the appropriate radiation condition. In (3.10*b*) we have neglected those terms in (2.5*b*) which are $O(\eta g'/g)$. The elimination of these terms will filter the long surface gravity waves out of the model.

The location of the slope parameter s in (3.10) occurs in such a way that to $O(s)$ the dynamics of the shelf water will be quasi-geostrophic. The location of the slope parameter s in (3.11*a*, *b*) implies that the interior eddy dynamics is essentially geostrophic but not quasi-geostrophic since changes in h are comparable with h itself. This is analogous to the 'intermediate lengthscale dynamics' identified in Charney & Flierl (1981).

We can exploit the fact that $0 < s \ll 1$ by constructing a straightforward asymptotic expansion of the form

$$(\eta, p, \boldsymbol{u}_1, \boldsymbol{u}_2, h, \phi) \sim (\eta_0, p_0, \boldsymbol{u}_{10}, \boldsymbol{u}_{20}, h_0, \phi_0) + s(\eta_1, p_1, \boldsymbol{u}_{11}, \boldsymbol{u}_{21}, h_1, \phi_1) + \ldots . \tag{3.13}$$

Substitution of this expansion into (3.10), (3.11) and (3.12) yields the $O(1)$ problem in the form

$$(\nabla^2 \partial_t + \partial_x)\eta_0 + h_{0_x} + \mu J(\eta_0, \nabla^2 \eta_0) = 0, \tag{3.14a}$$

$$h_{0_t} - h_{0_x} + \mu J(\eta_0, h_0) = -\beta h_0. \tag{3.14b}$$

The quasi-geostrophic potential vorticity equation (3.14*a*) is obtained in the usual way (see e.g. Pedlosky 1987, §3.12), by forming the vorticity equation for the $O(s)$ equations associated with (3.10*a*), eliminating the $O(s)$ divergence term using (3.10*b*), and then finally simplifying the resulting expression using the $O(1)$ equations obtained from (3.11).

The leading-order velocities and dynamic pressure in the eddy will be obtained from the auxiliary relations

$$\boldsymbol{u}_{10} = \hat{\boldsymbol{e}}_3 \times \nabla \eta_0, \tag{3.15a}$$

$$\boldsymbol{u}_{20} = -\hat{\boldsymbol{e}}_1 + \mu \hat{\boldsymbol{e}}_3 \times \nabla(\eta_0 + h_0), \tag{3.15b}$$

$$p_0 = y + \mu(\eta_0 + h_0), \tag{3.15c}$$

and the eddy boundary conditions can be written in the form

$$\left.\begin{aligned} \phi_{0_t} - \phi_{0_x} + \mu J(\eta_0 + h_0, \phi_0) &= 0, \\ h_0 &= 0, \end{aligned}\right\} \quad \text{on} \quad \phi_0 = 0, \qquad \begin{aligned} &(3.16a) \\ &(3.16b) \end{aligned}$$

where the Jacobian is given by $J(A, B) = A_x B_y - A_y B_x$.

The problem posed by the set of equations (3.14) retains all of the physics associated with the Nof and Mory *et al.* theories. If we were to introduce (as we eventually will in a suitable form) co-moving coordinates $\xi = x - c_x t$ and $\zeta = y - c_y t$ into (3.14) (with $\beta = 0$) and multiply (3.14*b*) by ξ and integrate over the eddy region R, the result would be the generalized Nof formula (3.7). Similarly, multiplying (3.14*b*) by ζ and integrating over R would yield (3.8).

The model of Ou & Houghton (1982) for the along-shelf heating of the Middle Atlantic Cold Pool can be viewed as corresponding to (3.14*b*) with η_0 as a prescribed geostrophic pressure associated with a steady along-shelf current and the additional approximation $h_{0_y} \ll h_{0_x}$. In Ou & Houghton's theory the slope water does not dynamically evolve and thus there is no analogue of (3.14*a*) in their model. The equations (3.14*a*, *b*) without the heating terms have also been derived in Whitehead *et al.* (1989). Their analysis develops an integral constraint for isolated cold eddies and a lower bound for the eddy radius. Although no explicit analytical solutions are presented, a series of numerical experiments is described.

Earlier in this section we argued that, based on observed order of magnitude estimates for our scalings, the interaction parameter $\mu = O(1)$. Consequently, (3.14)–(3.16) constitute a fully interacting coupled eddy–slope water model. We have not been able to find a *general* solution for these equations for arbitrary initial conditions. However, exact nonlinear solutions can be obtained for some special cases. For example, in the absence of any diabatic processes (i.e. $\beta = 0$) we can find exact radially symmetric non-radiating steadily travelling solutions (described in the next section).

4. Weak diabatic heating theory

4.1. *Problem formulation*

Although we have not been able to find a *general* solution to (3.14) and (3.16), it is possible to make further progress under a weak diabatic heating limit. In our discussion in §3.2 it was argued that the interaction parameter $\mu = O(1)$ but that $\beta = O(10^{-1})$. In this Section we shall develop a theory describing the dynamical evolution of an isolated eddy and slope-water configuration, which is initially balanced so that there is no fully developed $O(1)$ wave field in the slope water for $t \approx 0^+$, under the asymptotic limit $0 \ll s \ll \beta \ll 1$. We shall show that the vortex-tube dynamics associated with the diabatic mass conversion will result in the generation of topographic Rossby waves in the slope water and an induced upslope motion in the propagating eddy.

Assuming the cold-core eddy is to leading order radially symmetric, it is natural to reformulate the governing equations in the co-moving polar coordinates $r = (\xi^2+\zeta^2)^{\frac{1}{2}}$ and $\theta = \tan^{-1}(\zeta/\xi)$, and the slow time $T = \beta t$, where ξ and ζ are the co-moving phase variables

$$\xi = x - \beta^{-1}\int_0^{\beta t} c_x(t')\,\mathrm{d}t', \qquad (4.1a)$$

$$\zeta = y - \beta^{-1}\int_0^{\beta t} c_y(t')\,\mathrm{d}t'. \qquad (4.1b)$$

Note that the along-shelf and cross-shelf velocities given by $-\xi_t = c_x(T)$ and $-\zeta_t = c_y(T)$ respectively have both been formally scaled $O(1)$ on account of the assumption that $\mu = O(1)$ in (3.7) and (3.8). As it turns out, however, the assumption that the leading-order eddy and slope-water configuration contains no wave field will imply that $c_y = O(\beta)$. We have also allowed the translation speeds to be time dependent on the slow timescale $O(\beta^{-1})$ because of the presence of the diabatic heating terms.

It follows from (4.1) that derivatives in the governing equations will map according to

$$\partial_x \rightarrow \cos(\theta)\,\partial_r - r^{-1}\sin(\theta)\,\partial_\theta, \qquad (4.2a)$$

$$\partial_y \rightarrow \sin(\theta)\,\partial_r + r^{-1}\cos(\theta)\,\partial_\theta, \qquad (4.2b)$$

$$\partial_t \rightarrow -c_x(T)\,[\cos(\theta)\,\partial_r - r^{-1}\sin(\theta)\,\partial_\theta] - c_y(T)\,[\sin(\theta)\,\partial_r + r^{-1}\cos(\theta)\,\partial_\theta] + \beta\,\partial_T. \qquad (4.2c)$$

As well, the boundary of the eddy will be written in the form

$$\phi(x,y,t) = r - a(\theta,T) = 0.$$

Substitution of (4.2) and the assumed form for the eddy boundary into (3.14), (3.15) and (3.16) yields

$$[\cos(\theta)\,\partial_r - r^{-1}\sin(\theta)\,\partial_\theta][-c_x\nabla^2\eta + \eta + h] + \mu J(\eta, \nabla^2\eta)$$
$$- c_y\,[\sin(\theta)\,\partial_r + r^{-1}\cos(\theta)\,\partial_\theta]\,\nabla^2\eta = -\beta\nabla^2\eta_T, \qquad (4.3)$$

$$(c_x+1)\,[\cos(\theta)\,\partial_r - r^{-1}\sin(\theta)\,\partial_\theta]\,h - \mu J(\eta, h)$$
$$+ c_y[\sin(\theta)\,\partial_r + r^{-1}\cos(\theta)\,\partial_\theta]\,h = \beta h_T + \beta h, \qquad (4.4)$$

subject to the eddy boundary conditions

$$h = 0, \qquad (4.5a)$$

$$(c_x+1)\,[a\sin(\theta)]_\theta - c_y[a\cos(\theta)]_\theta = -\mu a_\theta(\eta+h)_r - \mu(\eta+h)_\theta - \beta a a_T, \qquad (4.5b)$$

evaluated on $r = a(\theta, T)$. In polar coordinates, the Jacobian takes the form

$J(A,B) = r^{-1}(A_rB_\theta - A_\theta B_r)$ and $\nabla^2 = \partial_{rr} + r^{-1}\partial_r + r^{-2}\partial_{\theta\theta}$. Note that we have deleted the zero subscript from the dependent variables for subsequent notational convenience.

The radiation condition on the slope-water geostrophic pressure can be written in the form

$$\lim_{r\to\infty} r^{\frac{1}{2}}\eta(r,\theta;T) = 0, \tag{4.6}$$

in the sector $|\theta-\theta_*| < \frac{1}{2}\pi$ where $\theta_* = \tan^{-1}(c_y/c_x)$. This condition is obtained as follows. The slowly modulated eddy will propagate along a path whose tangent vector is given by $(c_x(T), c_y(T))$ on the (x, y)-coordinate plane at each instant of time. The velocity vector will form an angle $\theta_* = \tan^{-1}(c_y/c_x)$ with the positive x-axis. Defining 'ahead' of the propagating eddy to be those coordinates with angle θ satisfying $-\frac{1}{2}\pi < \theta-\theta_* < \frac{1}{2}\pi$, then the 'no waves' constraint that $r^{\frac{1}{2}}\eta \to 0$ as $r\to\infty$ ahead of the propagating eddy (see Miles 1968 or Flierl 1984*b*) can be written as above.

4.2. *The leading-order solution*

We shall develop a weakly ventilated theory for an initially radially symmetric isolated baroclinic eddy. The (formal) asymptotic expansion is given by

$$h(r,\theta;T) = h^{(0)}(r;T) + \beta h^{(1)}(r,\theta;T) + \dots, \tag{4.7a}$$

$$\eta(r,\theta;T) = \eta^{(0)}(r;T) + \beta\eta^{(1)}(r,\theta;T) + \dots, \tag{4.7b}$$

$$a(\theta;T) = a^{(0)}(T) + \beta a^{(1)}(\theta;T) + \dots, \tag{4.7c}$$

$$c_x(T) = c_x^{(0)}(T) + O(\beta^2), \tag{4.7d}$$

$$c_y(T) = c_y^{(0)}(T) + \beta c_y^{(1)}(T) + O(\beta^2). \tag{4.7e}$$

As well, the eddy boundary conditions (4.5*a*, *b*) will have to be Taylor expanded about $r = a^{(0)}(T)$. Neglecting terms of $O(\beta^3)$ and higher, the approximate boundary conditions associated with (4.5*a*, *b*) can be written in the respective forms

$$h + (a-a^{(0)})h_r + \tfrac{1}{2}(a-a^{(0)})^2 h_{rr} + O(\beta^3) = 0, \tag{4.8a}$$

$$\begin{aligned}(c_x+1)[a\sin(\theta)]_\theta - c_y[a\cos(\theta)]_\theta = {} & -\mu a_\theta(\eta+h)_r - \mu(\eta+h)_\theta - \beta a a_T \\ & -\mu a_\theta(a-a^{(0)})(\eta+h)_{rr} - \mu(a-a^{(0)})(\eta+h)_{\theta r} + O(\beta^3),\end{aligned} \tag{4.8b}$$

evaluated on $r = a^{(0)}(T)$.

In the expansion for the along-shelf translation speed $c_x(T)$ given in (4.7*d*) we have implicitly assumed $c_x^{(1)}(T) = 0$. It turns out that if this term is retained in the asymptotic expansion it is possible to formally determine completely the $O(\beta)$ solutions (in particular the exterior wave field) *and* satisfy all known solvability conditions associated with the $O(\beta)$ equations with this parameter left unspecified. We have therefore chosen to set $c_x^{(1)}(T) = 0$. We hasten to add that it may be that $c_x^{(1)}(T)$ is determined from solvability conditions associated with the $O(\beta^2)$ or higher-order problems. However, our inability to solve the $O(\beta)$ equations for the eddy in closed form for even the simplest $h^{(0)}$ and $\eta^{(0)}$ has not made this determination tractable. This is unfortunate because the experiments carried out by Mory *et al.* (1987) showed that the observed along-shelf translation speed of the eddy was much smaller than the Nof speed and we have not been able to reproduce this result for an eddy subjected to diabatic heating. (Recent informal discussions with D. Nof and M. Stern suggest that errors in the measuring of the eddy density may have led to overestimating the appropriate Nof speed for the Mory *et al.* experiment and as a result the discrepancy may not be nearly as large as originally thought.)

Substitution of the expansion (4.7) into (4.3), (4.4) and (4.8) yields the $O(1)$ set of equations

$$\cos(\theta)\left[-c_x^{(0)}\nabla^2\eta^{(0)}+\eta^{(0)}+h^{(0)}\right]_r - c_y^{(0)}\sin(\theta)\nabla^2\eta_r^{(0)} = 0, \tag{4.9a}$$

$$\left[(c_x^{(0)}+1)\cos(\theta)+c_y^{(0)}\sin(\theta)\right]h_r^{(0)} = 0, \tag{4.9b}$$

subject to the $O(1)$ boundary conditions

$$h^{(0)}[a^{(0)}(T);T] = 0, \tag{4.10a}$$

$$\left[\cos(\theta)(c_x^{(0)}+1)+c_y^{(0)}\sin(\theta)\right]a^{(0)} = 0. \tag{4.10b}$$

Assuming a non-trivial solution for $h^{(0)}$ and $a^{(0)}$, it follows directly from (4.9*b*) or (4.10*b*) and the orthogonality of the trigonometric functions over $0<\theta<2\pi$ that

$$c_x^{(0)} = -1, \quad c_y^{(0)} = 0, \tag{4.11a, b}$$

and consequently from (4.9*a*) that

$$\nabla^2\eta^{(0)}+\eta^{(0)} = -h^{(0)}, \tag{4.12}$$

where the constant of integration is zero because the eddy is isolated.

Note that $h^{(0)}(r;T)$ is undetermined to this order except for the boundary condition (4.10*a*). The fact that $h^{(0)}(r,T)$ is undetermined at this order is a difficulty which occurs in many models of isolated eddies (e.g. Killworth 1983; Nof 1983; Flierl 1984*b*, among others). Flierl (1984*b*) chose to close the problem by specifying the potential vorticity distribution in the eddy interior which in turn would uniquely determine $h^{(0)}(r;T)$. Nof (1983) preferred to specify the interior eddy *swirl* velocity (i.e. the eddy azimuthal velocity in a frame of reference moving with the along-shelf Nof speed) and thereby determine the appropriate corresponding eddy height. Here, it will be more convenient to specify $h^{(0)}(r;0)$ and simply compute the other corresponding fields. However, because of our scaling, we can interpret the specification of $h^{(0)}(r;0)$ as equivalent to a potential vorticity specification in the following sense. The potential vorticity associated with (2.5*c*) and (2.5*d*) may be written in the form $P=[s\hat{\boldsymbol{e}}_3\cdot(\boldsymbol{\nabla}\times\boldsymbol{u}_2)+1]/h$. Thus in the limit as $s\rightarrow 0$ we have $P\approx 1/h_0+O(s)$. Consequently in the small-slope limit adopted here, the specification of h_0 is equivalent to a specification of the leading-order potential vorticity.

Given the leading-order eddy height $h^{(0)}(r;T)$, the *isolated* solution for the geostrophic pressure in the surrounding slope water can be written in the form (see Flierl 1984*b*)

$$\eta^{(0)}(r;T) = -\tfrac{1}{2}\pi Y_0(r)\int_0^r r'J_0(r')\,h^{(0)}(r';T)\,\mathrm{d}r' - \tfrac{1}{2}\pi J_0(r)\int_r^{a^{(0)}} r'Y_0(r')\,h^{(0)}(r';T)\,\mathrm{d}r', \tag{4.13a}$$

in $r\leqslant a^{(0)}$, and

$$\eta^{(0)}(r;T) = 0, \tag{4.13b}$$

in $r>a^{(0)}$, with the 'isolation' constraint

$$\int_0^{a^{(0)}} rJ_0(r)\,h^{(0)}(r;T)\,\mathrm{d}r = 0. \tag{4.13c}$$

The constraint (4.13*c*) is the necessary and sufficient condition for the annihilation of the external topographic wave field (Flierl 1984*a*) and can be physically interpreted as a zero-wave drag condition. It also follows from (4.13*c*) that the relative circulation in the upper layer is zero and that $\eta_r^{(0)}(a^{(0)};T)=0$.

The condition (4.13*c*) will imply that the allowed leading-order eddy radius can only take on a discrete set of values determined by the particular form for $h^{(0)}(r;T)$.

The no radiation condition (4.13*c*) raises an interesting issue concerning the characterization or determination of the class of functions $h^{(0)}(r;T)$ that will allow (4.13*c*) to have non-trivial solutions ($a^{(0)}=0$ is always a solution). For example, if we take $h^{(0)}(r;0)=J_0(r)$ it immediately follows that *only* $a^{(0)}=0$ is allowed. We have been unable to determine general necessary and sufficient conditions on $h^{(0)}(r;T)$ which will guarantee that non-trivial solutions for $a^{(0)}$ exist and this aspect of our theory remains problematic. The constraint (4.13*c*) may be viewed as resulting from the fact that the leading-order time evolution is modelled solely as the result of quasi-steady advection (see (4.2*c*)). This assumption will remove the possibility of generating an $O(1)$ wave field during the early adjustment period that would occur in a true initial-value problem. Accordingly, it is important to emphasize again that the theory developed here is only valid for an appropriately adjusted or dynamically balanced initial eddy–slope water configuration which can satisfy the $O(1)$ zero-wave-drag condition. We would also like to comment here that the eddy–slope-water solution configuration just obtained can be interpreted as corresponding to a steadily travelling baroclinic monopole on a continental shelf.

The leading-order isolated eddy solution just constructed will satisfy the *Stern integral constraint* for isolated eddies (Mory 1983, 1985), which for our theory can be expressed in the form

$$\int_0^{a^{(0)}} r[\eta^{(0)}(r;T)+h^{(0)}(r;T)]\,\mathrm{d}r=0. \tag{4.14}$$

This result follows immediately from (4.12) and the fact that $\eta_r^{(0)}(a^{(0)};T)=0$ as a consequence of (4.13).

Stern's integral constraint has an interesting consequence for the eddy swirl velocity (i.e. the eddy azimuthal velocity relative to the steady along-shelf Nof translation speed) distribution. It follows from (3.15*c*) that the leading-order eddy swirl velocity, denoted by $v_s(r;T)$, is given by

$$v_s(r;T)=\mu(\eta^{(0)}+h^{(0)})_r. \tag{4.15}$$

However, from (4.14) and (4.15) it is easy to show that

$$\int_0^{a^{(0)}} r^2 v_s(r;T)\,\mathrm{d}r=0. \tag{4.16}$$

Clearly the constraint (4.16) will imply that within the eddy there must be regions of cyclonic *and* anticyclonic circulation if there is to be any swirl velocity at all. At first this result may seem counter-intuitive because it is natural to think that the swirl velocity in a cold-core eddy would be strictly anticyclonic. But this conclusion rests on the assumption that the dynamic pressure in the eddy is relatively unaffected by the dynamical response of the surrounding slope water. However, the theory presented here suggests (see (3.15*c*)) that the contribution to the eddy dynamic pressure from the slope-water geostrophic pressure and the eddy thickness are comparable. Since $\eta_r^{(0)}<0$ where $h_r^{(0)}>0$ (i.e. flow above the eddy is cyclonic), it can easily follow from (3.15*c*) that the eddy swirl velocity can take on positive as well as negative values. We shall illustrate these points with an example calculation in §5.

4.3. *The first-order perturbation equations*

There are several dynamical characteristics of the leading order solution that have yet to be determined and for which we will have to examine the $O(\beta)$ equations of motion. We shall be able to determine, based on relatively direct solvability conditions on the $O(\beta)$ equations, the slow-time ventilation of the cold eddy as well as the leading-order cross-shelf translation velocity. Also, the $O(\beta)$ exterior topographic wave field in the slope water can be determined.

The $O(\beta)$ problem in the eddy region $r < a^{(0)}$ can be written in the form

$$[\cos(\theta)\,\partial_r - r^{-1}\sin(\theta)\,\partial_\theta][\nabla^2\eta^{(1)} + \eta^{(1)} + h^{(1)}] - \mu\eta_\theta^{(1)}\,\nabla^2\eta_r^{(0)}/r + \mu\eta_r^{(0)}\,\nabla^2\eta_\theta^{(1)}/r$$

$$= c_y^{(1)}\sin(\theta)\,\nabla^2\eta_r^{(0)} - \nabla^2\eta_T^{(0)}, \tag{4.17a}$$

$$c_y^{(1)}\sin(\theta)\,h_r^{(0)} - \mu r^{-1}[\eta_r^{(0)}\,h_\theta^{(1)} - \eta_\theta^{(1)}\,h_r^{(0)}] = h_T^{(0)} + h^{(0)}, \tag{4.17b}$$

subject to the $O(\beta)$ eddy boundary conditions

$$h^{(1)}(a^{(0)},\theta;T) + h_r^{(0)}(a^{(0)};T)\,a^{(1)}(\theta;T) = 0, \tag{4.18a}$$

$$\mu\eta_\theta^{(1)} = -c_y^{(1)}\,a^{(0)}\sin(\theta) - a^{(0)}\,a_T^{(0)}, \tag{4.18b}$$

evaluated on $r = a^{(0)}$, where (4.18*a*) and $\eta_r^{(0)}(a^{(0)},T) = 0$ has been used to obtain (4.18*b*).

In the region *exterior* to the eddy (i.e. $r > a^{(0)}$) the upper-layer geostrophic pressure satisfies

$$\nabla^2\eta^{(1)} + \eta^{(1)} = 0, \tag{4.19}$$

where, again, the constant of integration has been set to zero because it is assumed that $\eta^{(1)} \rightarrow 0$ as $r \rightarrow \infty$.

The leading-order response of the eddy to the diabatic heating is determined by (4.17*b*). Because of the orthogonality of the trigonometric functions and the periodicity that $\eta^{(1)}$ and $h^{(1)}$ must have over the interval $0 < \theta < 2\pi$, it follows from (4.17*b*) that

$$h_T^{(0)} + h^{(0)} = 0. \tag{4.20}$$

Similarly it follows from (4.18*b*) that $a_T^{(0)} = 0$. The solution to (4.20) is given by simply

$$h^{(0)}(r;T) = \tilde{h}(r)\exp(-T), \tag{4.21}$$

where $\tilde{h}(r)$ corresponds to the initial radially symmetric eddy height profile.

4.3.1. *Determination of the cross-shelf translation speed*

Even for a very simple $O(1)$ eddy height configuration (e.g. $h^{(0)}(r;0)$ given as a parabola; see §5), the $O(\beta)$ eddy problem cannot be solved exactly and a numerical solution will be required. However, as it turns out, detailed knowledge of the structure of $h^{(1)}(r,\theta;T)$ and $\eta^{(1)}(r,\theta;T)$ is *not* required in order to determine either the leading-order cross-shelf translation speed or the complete structure of the exterior (i.e. $r > a^{(0)}$) topographic wave field.

The simplest and most direct way to obtain $c_y^{(1)}(T)$ is to work directly with (4.17) rewritten in the co-moving Cartesian variables (ξ,ζ) given by (4.1). In these coordinates (4.17) can be written in the form

$$[\nabla^2\eta^{(1)} + \eta^{(1)} + h^{(1)}]_\xi + \mu J(\eta^{(0)},\nabla^2\eta^{(1)}) + \mu J(\eta^{(1)},\nabla^2\eta^{(0)}) = c_y^{(1)}\,\nabla^2\eta_\xi^{(0)} - \nabla^2\eta_T^{(0)}, \tag{4.22}$$

$$\mu J(\eta^{(0)},h^{(1)}) + \mu J(\eta^{(1)},h^{(0)}) = c_y^{(1)}\,h_\xi^{(0)} - h^{(0)} - h_T^{(0)}. \tag{4.23}$$

In order to obtain solvability conditions on (4.22) and (4.23) we shall have to examine homogeneous solutions of the adjoint problem associated with (4.22) and (4.23). The procedure we describe here is similar to that developed by Flierl (1984*b*).

The governing equations for the adjoint problem can be obtained by multiplying (4.22) by a function $\phi_1(\xi,\zeta;T)$ and multiplying (4.23) by a function $\phi_2(\xi,\zeta;T)$, adding the two expressions together and then integrating over the eddy region. The result can be put into the form

$$\iint_R [\mu J(h^{(0)},\phi_2)-\mu\nabla^2 J(\eta^{(0)},\phi_1)-\mu J(\eta^{(0)}+h^{(0)},\phi_1)-\partial_\xi(\nabla+1)\,\phi_1)]\,\eta^{(1)}\,\mathrm{d}\xi\,\mathrm{d}\zeta$$

$$-\iint_R [\mu J(\eta^{(0)},\phi_2)+\partial_\xi\,\phi_1]\,h^{(1)}\,\mathrm{d}\xi\,\mathrm{d}\zeta$$

$$=\iint_R [c_y^{(1)}\,\nabla^2\eta_\zeta^{(0)}-\nabla^2\eta_T^{(0)}]\,\phi_1\,\mathrm{d}\xi\,\mathrm{d}\zeta+\iint_R [c_y^{(1)}\,h_\zeta^{(0)}-h^{(0)}-h_T^{(0)}]\,\phi_2\,\mathrm{d}\xi\,\mathrm{d}\zeta, \tag{4.24}$$

assuming where necessary that the boundary integrals are zero.

The homogeneous adjoint problem will therefore be given by

$$\mu J(h^{(0)},\phi_2)-\mu\nabla^2 J(\eta^{(0)},\phi_1)-\mu J(\eta^{(0)}+h^{(0)},\phi_1)-\partial_\xi(\nabla^2+1)\,\phi_1=0, \tag{4.25}$$

$$\mu J(\eta^{(0)},\phi_2)+\partial_\xi\,\phi_1=0. \tag{4.26}$$

We can find three immediate homogeneous solutions: $(\phi_1=0,\phi_2=1)$, $(\phi_1=1,\phi_2=0)$ and $(\phi_1=-\eta^{(0)},\phi_2=\zeta/\mu)$. There may be other solutions but we have not been able to find them. When these homogeneous solutions are substituted into (4.24) the left-hand side is identically zero and the right-hand side gives us the required solvability conditions. From the pair $(\phi_1=0,\phi_2=1)$ we find

$$\iint_R [h_T^{(0)}+h^{(0)}]\,\mathrm{d}\xi\,\mathrm{d}\zeta=0, \tag{4.27}$$

which is satisfied on account of (4.20). From the pair $(\phi_1=1,\phi_2=0)$ we find

$$\iint_R \nabla^2\eta_T^{(0)}\,\mathrm{d}\xi\,\mathrm{d}\zeta=0, \tag{4.28}$$

which is satisfied because $\eta_r^{(0)}(a^{(0)};T)=0$. From the pair $(\phi_1=-\eta^{(0)},\phi_2=\zeta/\mu)$ we find

$$c_y^{(1)}\iint_R h^{(0)}\,\mathrm{d}\xi\,\mathrm{d}\zeta=\mu\iint_R \eta^{(0)}\,\nabla^2\eta_T^{(0)}\,\mathrm{d}\xi\,\mathrm{d}\zeta,$$

which can be rearranged using (4.14) into the form

$$c_y^{(1)}(T)=\tfrac{1}{2}\mu\,\partial_T\int_0^{a^{(0)}} r[\boldsymbol{\nabla}\eta^{(0)}\cdot\boldsymbol{\nabla}\eta^{(0)}](r;T)\,\mathrm{d}r\Big/\int_0^{a^{(0)}} r\eta^{(0)}(r;T)\,\mathrm{d}r. \tag{4.29}$$

It follows from (4.29) that $c_y^{(1)}(T)>0$ since the numerator and denominator are both negative definite. This can be shown as follows. From (4.14) and the fact that $h^{(0)}(r;T)\geqslant 0$ for $0\leqslant r\leqslant a^{(0)}$ it will follow that the denominator in (4.29) is negative. If (4.21) is substituted into (4.13*a*) then it follows that $\eta^{(0)}(r;T)$ will have the general form $\eta^{(0)}(r;T)=\tilde{\eta}(r)\exp(-T)$, where $\tilde{\eta}(r)$ is a radially symmetric function that is determined in the calculation of (4.13*a*). Substituting this form for $\eta^{(0)}(r;T)$ into the numerator of (4.29) yields a term of the form $\partial_T[\eta_*\exp(-2T)]$ where the constant $\eta_*>0$ and hence the numerator will also be negative. Thus the effect of ventilation is to induce *upslope* motion in the propagating eddy. As the process of ventilation continues (4.29) implies that the magnitude of cross-shelf drift speed will decrease since the magnitude of $\eta^{(0)}$ decreases as T increases.

This qualitative result can be interpreted as a consequence of the conservation of potential vorticity in the upper layer. Equations (3.14*a*) and (3.14*b*) can be combined to state that to leading order in β the potential vorticity $\nabla^2\eta+h+y$ is conserved following the geostrophic motion. Since the leading-order 'averaged' geostrophic vorticity or circulation in the slope water above the eddy is identically zero, the effect of the diabatic heating (which implies $h_t<0$) can only be offset by a corresponding increase in the mean cross-shelf position of the cold-core eddy.

4.3.2. *Determination of the exterior topographic wave field*

The $O(\beta)$ geostrophic pressure in the exterior region for the upper layer must satisfy

$$\nabla^2\eta^{(1)}+\eta^{(1)}=0, \tag{4.30}$$

subject to the boundary condition

$$\eta^{(1)}=\mu^{-1}c_y^{(1)}\cos(\theta), \tag{4.31a}$$

evaluated on $r=a^{(0)}$, and the no-upstream-waves condition

$$r^{\frac{1}{2}}\eta^{(1)}(r,\theta;T)\rightarrow 0 \quad \text{as} \quad r\rightarrow\infty \quad \text{in} \quad \tfrac{1}{2}\pi<\theta<\tfrac{3}{2}\pi. \tag{4.31b}$$

The boundary condition (4.31*a*), which follows directly from (4.18*b*), will ensure that the geostrophic pressure and normal mass flux in the slope water is continuous across the eddy boundary $r=a^{(0)}$.

The method of solution we use is a modification of the procedures developed by Miles (1968) for a similar boundary-value problem (see also McKee 1971; McCartney 1975; Flierl 1984*a*). Following the arguments presented in Miles we construct the exterior geostrophic pressure field in the form

$$\eta^{(1)}=\sum_{n=0}^{\infty}a_n(T)\left[Y_n(r)\cos(n\theta)+\psi_n(r,\theta;T)\right]/Y_n(a^{(0)}), \tag{4.32a}$$

where

$$\psi_n(r,\theta;T)=\sum_{m=0}^{\infty}b_{n,m}J_m(r)\cos(m\theta). \tag{4.32b}$$

(The reason for our particular choice of normalization in (4.32*a*) will be given below.)

Recalling that the even (odd) $J_n(r)$ functions have the same asymptotic form as the odd (even) $Y_n(r)$ functions as $r\rightarrow\infty$ (see Abramowitz & Stegun 1965), the no-waves condition (4.31*b*) will imply that the coefficients $b_{n,m}$ must satisfy the constraints

$$\cos(2n\theta)=\sum_{m=0}^{\infty}(-1)^{m+n+1}b_{2n,\,2m+1}\cos[(2m+1)\theta], \tag{4.33a}$$

$$\cos[(2n+1)\theta]=\sum_{m=0}^{\infty}(-1)^{m+n}b_{2n+1,\,2m}\cos[2m\theta], \tag{4.33b}$$

for $n=0,1,2,\ldots$ in the sector $\frac{1}{2}\pi<\theta<\frac{3}{2}\pi$. Since the sets $\{\cos[(2m\text{-}1)\theta]\}_{m=0}^{\infty}$ and $\{\cos(2m)\}_{m=0}^{\infty}$ are both complete in the interval $\frac{1}{2}\pi<\theta<\frac{3}{2}\pi$, it follows from (4.33*a, b*) that

$$b_{n,m}=\begin{cases}(4/\pi)\,m(m^2-n^2)^{-1} & (n \text{ even, } m \text{ odd}),\\ (4/\pi)\,n(m^2-n^2)^{-1} & (n \text{ odd, } m \text{ even}),\\ 0 & (n-m \text{ even}).\end{cases} \tag{4.34}$$

The only quantities that are left to determine are the a_n. If the double summation in (4.32) is interchanged, $\eta^{(1)}(r,\theta;T)$ can be expressed in the form

$$\eta^{(1)}=\sum_{m=0}^{\infty}\left\{\sum_{n=0}^{\infty}a_n\,\Gamma_{nm}(r)\right\}\cos(m\theta), \tag{4.35a}$$

where
$$\Gamma_{nm}(r) = [\delta_{nm} Y_m(r) + b_{n,m} J_m(r)]/Y_n(a^{(0)}), \tag{4.35b}$$
δ_{nm} being the Kronecker delta function between n and m. If we apply the boundary condition (4.31 a) the a_n will be determined from the infinite set of linear equations
$$\sum_{n=1}^{\infty} a_n \Gamma_{nm}(a^{(0)}) = \mu^{-1} c_y^{(1)} a^{(0)} \delta_{m1}, \tag{4.36}$$
with $m = 0, 1, 2, \ldots$. With the $b_{n,m}$ and a_n known the exterior solution is complete. The expression (4.36) implies that as the cold eddy is progressively ventilated the excited wave field will decrease in magnitude because $c_y^{(1)}(T) \rightarrow 0$ as $T \rightarrow \infty$.

As it turns out, relatively few a_n values need to be computed to be able to give a very accurate approximation to the infinite sums in (4.32) or (4.35). If we recall the fact that $Y_n(a^{(0)}) \rightarrow -\infty$ and $J_n(a^{(0)}) \rightarrow 0$ as $n \rightarrow \infty$ (Abramowitz & Stegun 1965), then clearly (4.34 b) implies that $\Gamma_{n,m}(a^{(0)}) \approx \delta_{nm}$ for sufficiently large n or m. (This property motivated our normalization in (4.32 a).) In practice, we found that for $n, m \gtrsim 12$ this property held. As a result, when it came to solving for the a_n from (4.36) very good results were obtained by approximating the infinite system with the leading 20×20 finite system of linear equations. (For all cases examined we found $|a_n| < 10^{-7}$ for $n > 15$.)

5. An example calculation for a parabolic eddy

In this section we shall illustrate the theory that we have developed with an example where the leading-order eddy has a simple parabolic shape. Throughout this section we shall adopt constant parameter values of $s = 5 \times 10^{-2}$, $\mu = 0.5$ and $\beta = 5 \times 10^{-3}$.

The $O(1)$ eddy configuration is given by
$$h^{(0)}(r; T) = \exp(-T)[1 - (r/a^{(0)})^2]. \tag{5.1}$$
In figure 2 (a) we plot a cross-shelf section (along $x = 0$) of the *total* eddy height $h^{(0)}(r; 0) + sy$ versus y for $-10 < y < 10$.

Substitution of (5.1) into (4.13 a) can be evaluated analytically to yield
$$\eta^{(0)}(r; T) = \pi \exp(-T) \Bigg\{ Y_0(r) \Bigg[\Bigg(\left(\frac{r}{a^{(0)}}\right)^2 - 1 \Bigg) \tfrac{1}{2} r J_1(r) - \frac{r^2 J_2(c)}{(a^{(0)})^2} \Bigg]$$
$$+ J_0(r) \Bigg[\Bigg(1 - \left(\frac{r}{a^{(0)}}\right)^2 \Bigg) \tfrac{1}{2} r Y_1(r) + \frac{r^2 Y_2(a^{(0)})}{(a^{(0)})^2} - Y_2(a^{(0)}) \Bigg] \Bigg\}, \tag{5.2}$$
for the region $r < a^{(0)}$. In figure 2 (b) we plot a cross-shelf section (along $x = 0$) of $\eta^{(0)}(r; 0)$ versus y for $-10 < y < 10$.

As well, substitution of (5.1) into (4.13 c) can be evaluated analytically to imply that the allowed set of eddy radii will satisfy
$$J_2(a^{(0)}) = 0, \tag{5.3a}$$
which, for convenience, we choose to rewrite in the form
$$a^{(0)} = j_{2,n}, \tag{5.3b}$$
where $j_{2,n}$ is the nth (non-zero) zero of $J_2(*)$. Throughout this section we take the 'ground-state' radius $a^{(0)} = j_{2,1} \approx 5.136$.

Substitution of (5.2) into (4.29) implies that the time-dependent cross-shelf translation speed is given by
$$c_y^{(1)}(T) = 0.227 \exp(-T). \tag{5.3c}$$

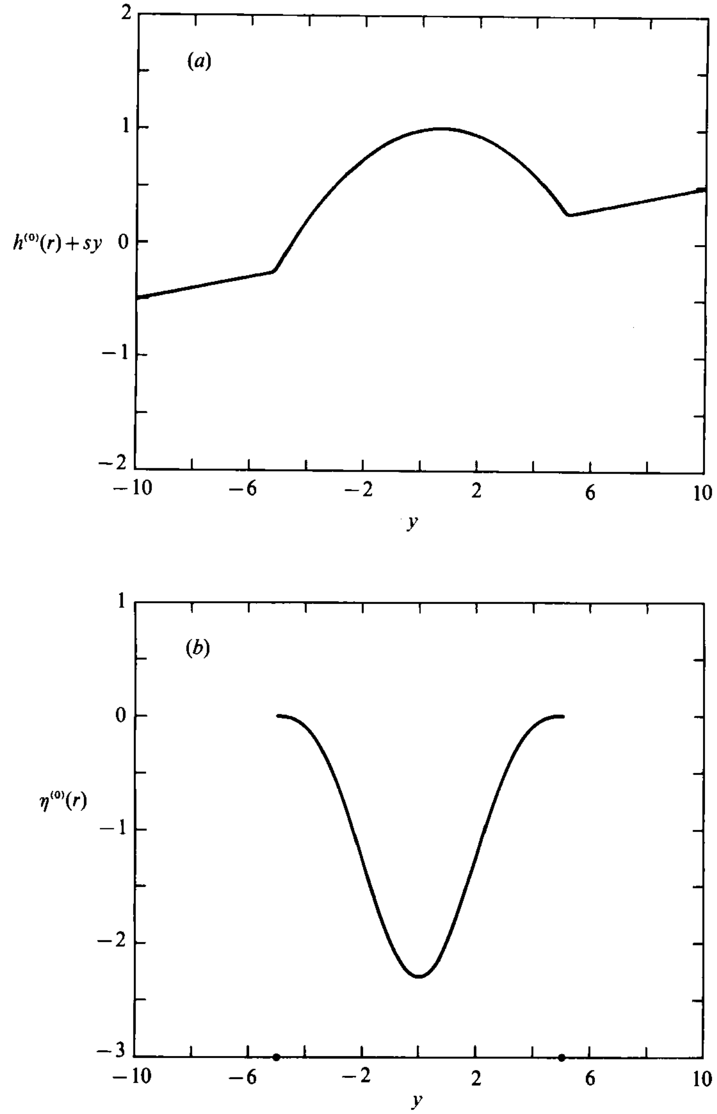


FIGURE 2. (*a*) A cross-shelf section of the parabolic eddy given by (5.1) on the sloping bottom at $T = 0$. (*b*) A cross-shelf section of the upper-layer $O(1)$ geostrophic pressure (immediately over the eddy depicted in (*a*) as determined by (5.2) at $T = 0$. The two dots on the y-axis denote the outer eddy boundary at $y = \pm a^{(0)}$.

We remark that the integral in the numerator of (4.29) was evaluated numerically.

The *total* leading-order pressure in the eddy (see (3.15*c*)) is given by

$$p^{(0)}(r;T) = y+\mu(\eta^{(0)}(r;T)+h^{(0)}(r;T)), \tag{5.4}$$

where the first term corresponds to the dynamic pressure associated with gravity and the sloping bottom, and the remaining terms are responsible for generating the swirl velocity (i.e. the leading-order eddy azimuthal velocity relative to the leading-order along-shelf translation speed) in the eddy interior (see (4.15)). For convenience we shall denote the last two terms on the right-hand side of (5.4) as the 'swirl pressure'. Recall that the *Stern integral constraint* (4.14) will imply that the swirl pressure in the

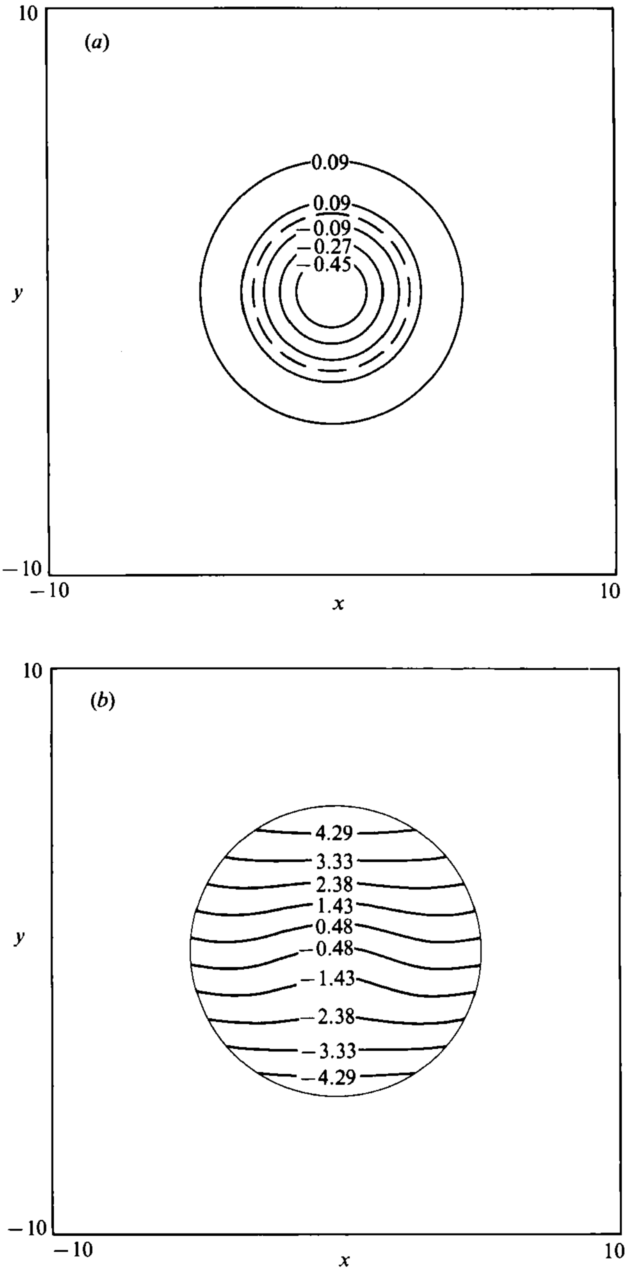


FIGURE 3. (*a*) A contour plot of the 'swirl' pressure $\mu[h^{(0)}(r;0)+\eta^{(0)}(r;0)]$ in the eddy. The dashed contour is the zero-value contour. Radially inward (outward) of the zero contour the swirl pressure anomaly is negative (positive) as demanded by the Stern integral constraint (4.14). (*b*) A contour plot of the *total* eddy geostrophic pressure field given by (5.4). The eddy geostrophic Eulerian velocity field is essentially the Nof speed (i.e. the along-shelf translation).

eddy must take on positive as well as negative values (if it is to be non-zero). In figure 3(*a*) we present a contour plot of the swirl pressure $\mu[h^{(0)}(r;0)+\eta^{(0)}(r;0)]$. The dashed contour is the 0-contour which occurs only once for the ground-state solution. The region of positive swirl pressure is located in $2.68 < r \leqslant a^{(0)}$. The region $0 < r < 2.68$

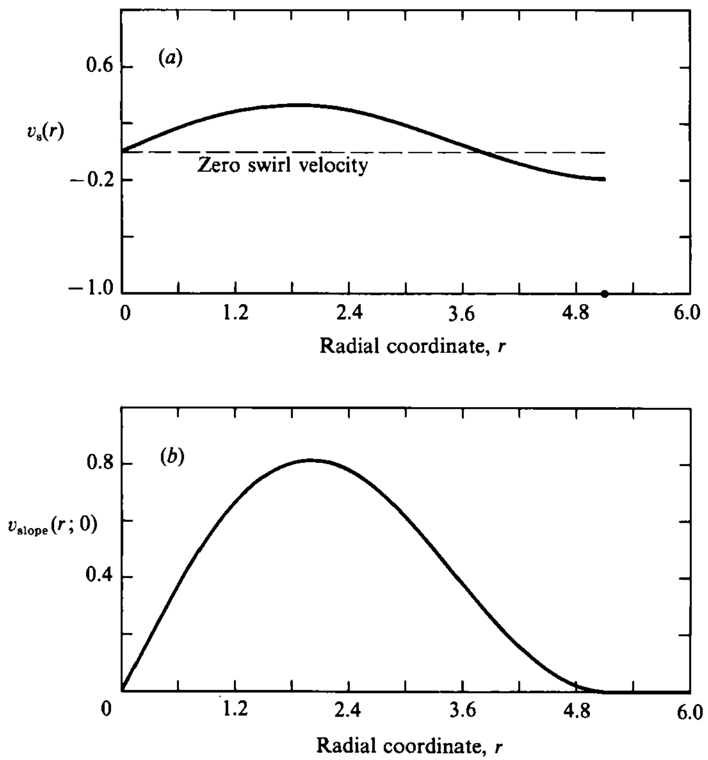


FIGURE 4. (*a*) The eddy swirl velocity as determined by (4.15). The velocities are non-dimensionalized with the Nof speed $g's/f \approx 2.5$ cm/s. The flow is mostly cyclonic with a small anticyclonic zone near the eddy boundary. The dashed horizontal line marks $v_s = 0$ and the dot on the r-axis denotes the eddy boundary $r = a^{(0)}$. (*b*) The cyclonic azimuthal velocity in the upper layer immediately over the eddy as determined by (5.5). The magnitude of the maximum dimensional azimuthal velocity is about twice as large as the maximum dimensional eddy swirl velocity.

has negative swirl pressure. The maximum swirl pressure is about 0.16 and occurs at $r \approx 3.83$. The minimum swirl pressure is located at $r = 0$ and has a value of approximately -0.65. These values scale linearly with the magnitude of the interaction parameter μ.

In figure 3(*b*) we present a contour plot of the total leading-order eddy pressure field $p^{(0)}(r;0)$ as given by (5.4). Because the magnitude of the swirl pressure is relatively small in comparison with the 'bottom slope' pressure contribution, the resulting contours are roughly speaking parallel to the isobaths. For realistic values of the interaction parameter (i.e. $\mu \approx 1$) there are no closed pressure contours. Consequently, to a stationary observer watching the cold-core eddy propagate, the relative velocity field in the eddy interior will appear almost negligible. This property has been observed in some rotating-tank experiments (J. A. Whitehead, personal communication). It is important to add, however, that the magnitude of the deflection in the pressure contours scales linearly with μ.

In figures 4(*a*) and 4(*b*) we present radial cross-sections of the leading-order swirl velocity and the azimuthal velocity in the upper layer, respectively, in the eddy region $r < a^{(0)}$ at $t = 0$. The eddy swirl velocity is given by (4.15) and the leading-order slope-water azimuthal velocity is given by

$$v_{\text{slope}}(r;T) = \eta_r^{(0)}(r;T). \tag{5.5}$$

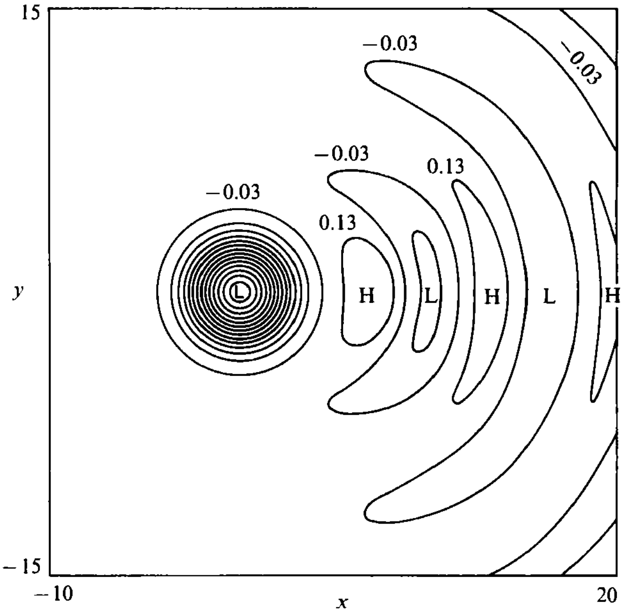


FIGURE 5. A contour plot of the complete leading-order geostrophic pressure field in the upper layer as determined by (5.6). The circular contours correspond to the strong cyclonic flow immediately over the eddy and the crescent-shaped contours correspond to the topographic Rossby wave field behind the propagating eddy. The H(L) symbols represent positive (negative) pressure anomalies and the contour interval is ±0.16.

Recall that it followed from the Stern integral constraint (4.14) that the swirl velocity in the eddy must take on cyclonic *and* anticyclonic values. For our parabolic eddy model, there is a broad region $r \lesssim 3.85$; see figure 4(*a*) where the swirl velocity is cyclonic. The maximum magnitude of $v_s(r;0)$ in this cyclonic region is about 0.33. This would correspond to about $\frac{1}{3}$ of the Nof translation speed in dimensional units, or about 0.8–1.0 cm/s if the Nof speed is about 2.5–3.0 cm/s. Near the outer interior edge of the eddy $(3.8 \lesssim r < a^{(0)})$, the swirl velocity becomes anticyclonic. The maximum swirl velocity in this region for $T = 0$ is located along the eddy boundary and is about 20% of the Nof translation speed.

We can estimate the maximum slope-water azimuthal velocities from figure 4(*b*). The azimuthal slope-water velocity is cyclonic over the eddy region and has a maximum non-dimensional value of about 0.8 near $r = 2.0$. From (2.4) the scale velocity in the slope water is $\delta fL \approx 3.0$ cm/s which implies that the maximum *dimensional* azimuthal velocity in the slope water is about 2.5 cm/s. Hence if we compare the swirl velocity in the eddy interior (recall that this is defined to be the azimuthal velocity in the eddy interior relative to the along-shelf motion) to the slope water azimuthal velocity we see that the velocities in the slope water are on the order of twice the eddy swirl velocities.

In figure 5 we present a contour plot of the complete leading-order geostrophic pressure in the upper layer including the topographic wave field in $r > a^{(0)}$ for $T = 0$. That is, we have plotted $\eta(r, \theta; 0)$ as determined by

$$\eta(r,\theta;0) = \begin{cases} \eta^{(0)}(r;0) & \text{if } r < a^{(0)}, \qquad (5.6a) \\ \beta\eta^{(1)}(r,\theta;0) & \text{if } r \geqslant a^{(0)}. \qquad (5.6b) \end{cases}$$

Even though we have not included the $O(\beta)$ solutions in the eddy region $r < a^{(0)}$, we believe that by choosing β small enough (recall that $\beta = 5 \times 10^{-3}$), figure 5 is a correct asymptotic representation of the leading-order slope-water pressure field in $0 < r < \infty$. The circular contours in figure 5 correspond to the eddy-region (i.e. $r < a^{(0)}$ $\eta^{(0)}(r; 0)$ solution and the remaining crescent-shaped contours correspond to the external (i.e. $r < a^{(0)}$) $\eta^{(0)}(r; 0)$ solution and the remaining crescent-shaped contours correspond to the external (i.e. $r > a^{(0)}$) $O(\beta)$ topographic wave field. The closed contours containing the H and L symbols correspond to regions of positive and negative pressure anomaly, respectively. The maximum wave amplitude occurs immediately behind the low-pressure region located over the eddy and has a magnitude of about 14% of the minimum in the main low.

6. Summary and concluding remarks

A theory has been presented to describe the propagation of coherent cold-core baroclinic eddies on a sloping bottom and their dynamic and thermodynamic (i.e. ventilated) interaction with the surrounding ocean. The theoretical study presented in this paper was motivated by oceanographic observations (e.g. Ou & Houghton 1982; Houghton *et al.* 1982; Armi & D'Asaro 1980) and rotating-tank experiments (e.g. Mory 1983; Mory *et al.* 1987) on steadily travelling baroclinic eddies on a sloping bottom. Some of the properties of the oceanographic data, particularly the along-shelf translation speed, agreed with a theory proposed by Nof (1983) for these eddies. However, the data from the rotating-tank experiments did not seem to agree with the Nof theory.

Mory *et al.* (1987) suggested that frictional forces or dynamical interaction between the cold-core eddies and the surrounding slope water (neglected in the Nof theory) may be important. The possible importance of dynamical interactions is suggested by the *Stern integral constraint* (Mory 1983, 1985; see (4.14)) which requires that the area-integrated geostrophic pressure in the slope-water balance the area-integrated buoyancy force in the cold dome if the eddy–slope-water configuration is to be isolated. If this balance does not hold, then there will be an excited topographic wave field behind the propagating eddy (see Flierl 1984*a*,*b*) and a concomitant cross-shelf drift.

In the oceanographic context, other processes are also important. For example, the process of ventilation between the cold eddy and the relatively warmer slope water was of importance during the evolution of the cold dome described by Houghton *et al.* (1982) and Ou & Houghton (1982).

The new model equations (see (3.14) and (3.15)) developed in this paper to study the dynamical and ventilatory interactions between a cold-core eddy and the surrounding ocean on a sloping bottom corresponded to strongly interacting 'hybrid' quasi-geostrophic, intermediate-lengthscale geostrophic dynamics (see Charney & Flierl 1981). Based on parameter values suggested by the oceanographic data, the dynamics of the surrounding slope water is a quasi-geostrophic with relative vorticity induced by the vortex-tube compression associated with the passage of the ventilating cold-core eddy. The eddy dynamics is geostrophic but is *not* quasi-geostrophic because eddy height changes are not small in relation to the scale height of the eddy (see figure 1). The model equations were derived in a formal asymptotic expansion based on the shallow-water equations for a two-layer fluid assuming a small (appropriately scaled) shelf slope parameter.

Because of the discrete stratification used in our model, the parameterization we adopted to describe the process of ventilation is the simple CI or cross-interfacial mass flux model of Dewar (1987, 1988*a*, *b*). This parameterization models the ventilation process as a continuous conversion of cold eddy water into relatively warmer slope water.

The model equations were solved using a multiple-scale asymptotic analysis valid in the limit of a relatively weak ventilation rate (but $O(1)$ dynamical interaction between the eddy and surrounding fluid) assuming an initially radially symmetric isolated eddy and slope-water configuration. The leading-order solution corresponds to a solitary baroclinic monopole configuration (see figure 2*a*, *b*) which propagates at the Nof speed and which satisfies the Stern integral constraint. For the simple parabolic eddy shape examined in §5, the swirl velocity in the eddy is about 30% of the Nof translation speed and about 50% of the Eulerian azimuthal velocity computed in the upper layer above the eddy. Consequently, to an external observer, the co-moving velocity field in the eddy would appear relatively quiescent in comparison to the slope-water velocity field above the eddy.

We are able to obtain an exact description of the induced topographic wave field behind the eddy (see figure 5) and to compute the associated cross-shelf translation velocity (see (4.29)). We find that the cross-shelf translation is positive (i.e. up the slope). If c_y^*, β^* and L denote the dimensional upslope speed, the dimensional ventilation rate (units of 1/s), and the lengthscale, then $c_y^* \sim 0.227\beta^* L$ initially. The fact that $c_y^* > 0$ can be explained as a straightforward consequence of slope-water potential vorticity conservation and the ventilating eddy (see the discussion after (4.29)).

It has been suggested (Mory *et al.* 1987) that the eddies of the type studied in this paper may result from the baroclinic instability of pycnobathic or bottom gravity currents (see also Smith 1976; Shaw & Csanady 1983; Griffiths, Killworth & Stern 1982). The analysis by Griffiths *et al.* (1982) is restricted to long wavelengths and only contains a single layer. It is not difficult to show that (in the absence of diabatic processes) our two-layer model equations (3.14) and (3.15) admit exact along-shelf gravity current solutions. The stratification characteristics of these gravity currents will resemble a cold-core coupled density front on a sloping bottom. Swaters (1991) presents a detailed baroclinic instability analysis of these gravity current solutions.

Another interesting problem is the effect of bottom friction on the propagation characteristics of the isolated-cold-eddy solutions examined in this paper. Mory *et al.* (1987) suggested that frictional spin-down may be important in the dynamics, particularly in the rotating-tank simulations. It turns out that if one includes a simple Rayleigh damping term in the non-dimensional eddy momentum equations (2.5*c*) with an $O(1)$ dissipation coefficient, and subsequently examines a weak dissipation limit similar to the weak ventilation limit examined here (for the appropriately modifed (3.14), (3.15) and (3.16)), it can be shown that $h^{(0)}(r;T)$ must satisfy a fully nonlinear parabolic equation and that the along-shelf translation speed is smaller than the Nof speed. We are currently examining this problem and hope to be able to report on this in the future.

This study was initiated when G.E.S. was a postdoctoral associate supported by National Science Foundation grants awarded to G.R.F. Final preparation of the manuscript was supported in part by an Operating Research grant awarded by the Natural Sciences and Engineering Research Council of Canada, and by a Science Subvention awarded by the Department of Fisheries and Oceans of Canada to G.E.S.

REFERENCES

ABRAMOWITZ, M. & STEGUN, I. A. 1965 *Handbook of Mathematical Functions*. Dover.

ARMI, L. & D'ASARO, E. 1980 Flow structures of the benthic ocean. *J. Geophys. Res.* **85** (C1), 469–483.

CHAPMAN, R. & NOF, D. 1988 The sinking of warm-core rings. *J. Phys. Oceanogr.* **18**, 565–583.

CHARNEY, J. G. & FLIERL, G. R. 1981 Oceanic analogues of large-scale atmospheric motions. In *Evolution of Physical Oceanography – Scientific Surveys in Honor of Henry Stommel* (ed. B. A. Warren & C. Wunsch), pp. 504–548. MIT Press.

DEWAR, W. K. 1987 Ventilating warm rings: theory and energetics. *J. Phys. Oceanogr.* **17**, 2219–2231.

DEWAR, W. K. 1988*a* Ventilating warm rings: structure and model evaluation. *J. Phys. Oceanogr.* **18**, 552–564.

DEWAR, W. K. 1988*b* Ventilating β-plane lenses. *J. Phys. Oceanogr.* **18**, 1193–1201.

FLIERL, G. R. 1984*a* Model of the structure and motion of a warm-core ring. *Austral. J. Mar. Freshw. Res.* **35**, 9–23.

FLIERL, G. R. 1984*b* Rossby wave radiation from a strongly nonlinear warm eddy. *J. Phys. Oceanogr.* **14**, 47–58.

GRIFFITHS, R. W., KILLWORTH, P. D. & STERN, M. E. 1982 Ageostrophic instability of ocean currents. *J. Fluid Mech.* **117**, 343–377.

HOUGHTON, R. W., SCHLITZ, R., BEARDSLEY, R. C., BUTMAN, B. & CHAMBERLIN, J. L. 1982 The middle Atlantic bight cool pool: evolution of the temperature structure during summer 1979. *J. Phys. Oceanogr.* **12**, 1019–1029.

KILWORTH, P. D. 1983 On the motion of isolated lenses on a β-plane. *J. Phys. Oceanogr.* **13**, 368–376.

MCCARTNEY, M. S. 1975 Inertial Taylor columns on a beta plane. *J. Fluid Mech.* **68**, 71–95.

MCKEE, W. W. 1971 Comments on 'A Rossby wake due to an island in an eastward current. *J. Phys. Oceanogr.* **1**, 287–289.

MILES, J. W. 1968 Lee waves in a stratified flow. Part 2. Semi-circular obstacle. *J. Fluid Mech.* **33**, 803–814.

MORY, M. 1983 Theory and experiment of isolated baroclinic vortices. *Tech. Rep.* WHOI-83-41, pp. 114–132. Woods Hole Oceanographic Institute.

MORY, M. 1985 Integral constraints on bottom and surface isolated eddies. *J. Phys. Oceanogr.* **15**, 1433–1438.

MORY, M., STERN, M. E. & GRIFFITHS, R. W. 1987 Coherent baroclinic eddies on a sloping bottom. *J. Fluid Mech.* **183**, 45–62.

NOF, D. 1983 The translation of isolated cold eddies on a sloping bottom. *Deep-Sea Res.* **30**, 171–182.

NOF, D. 1984 Oscillatory drift of deep cold eddies. *Deep-Sea Res.* **31**, 1395–1414.

OU, H. W. & HOUGHTON, R. 1982 A model of the summer progression of the cold-pool temperature in the middle Atlantic bight. *J. Phys. Oceanogr.* **12**, 1030–1036.

PEDLOSKY, J. 1987 *Geophysical Fluid Dynamics*, 2nd Edn. Springer.

SHAW, P. T. & CSANADY, G. T. 1983 Self-advection of density perturbations on a sloping continental shelf. *J. Phys. Oceanogr.* **13**, 769–782.

SMITH, P. C. 1976 Baroclinic instability in the Denmark strait overflow. *J. Phys. Oceanogr.* **6**, 355–371.

SWATERS, G. E. 1991 On the baroclinic instability of cold-core coupled density fronts on a sloping continental shelf. *J. Fluid Mech.* **224**, 361–382.

WHITEHEAD, J. A., STERN, M. E., FLIERL, G. R. & KLINGER, B. 1990 Experimental observations of a baroclinic eddy on a sloping bottom. *J. Geophys. Res.* **95**, 9585–9610.

J. Fluid Mech. (1991), *vol.* 223, *pp.* 589–601
Printed in Great Britain

On localized solutions in nonlinear Faraday resonance

By E. W. LAEDKE AND K. H. SPATSCHEK
Institut für Theoretische Physik I, Heinrich-Heine-Universität Düsseldorf,
D-4000 Düsseldorf, Germany

(Received 25 September 1989 and in revised form 30 July 1990)

The dynamics of a nonlinear modulated cross-wave of resonant frequency ω_1 and carrier frequency $\omega \approx \omega_1$ is considered. The wave is excited in a long channel of width b that contains water of depth d, which is subjected to a vertical oscillation of frequency 2ω. As has been shown by Miles (1984*b*), the complex amplitude satisfies a cubic Schrödinger equation with weak damping and parametric driving. The stability of its solitary wave solution is considered here in various parameter regions. We find that in a certain regime the solitary wave is stable. Completely new is the result of instability outside this parameter regime. The instability has also been verified numerically. It is shown that the final stage of solitary wave instability is a cnoidal-wave-type solution.

---

1. Introduction

This investigation was stimulated by a recent paper by Miles (1984*b*) who succeeded in developing a theory for the standing solitary wave observed by Wu, Keolian & Rudnick (1984). The wave appears owing to Faraday resonance, in which standing waves are parametrically excited in a basin that is subjected to a vertical oscillation at a frequency approximately twice the natural frequency of the dominant cross-wave. For the details of the theory we refer to the original articles of Miles (1984*a*, *b*) and Larraza & Putterman (1984). We shall use the notation of Miles in this paper.

The basic result of Miles is the cubic nonlinear Schrödinger equation

$$\mathrm{i}(r_t+\alpha r)+Br_{XX}+(\beta+A|r|^2)\,r+\gamma r^* = 0 \tag{1.1}$$

for the complex amplitude of the dominant cross-wave. Here, α is the linear damping ($\alpha > 0$) and the terms βr and νr^* appear because of the vertical oscillation $z = a_0 \cos 2\omega t$ in the gravitational field $-g\hat{z}$. Introducing a smallness parameter ϵ (which also characterizes the amplitude of the otherwise nonlinear oscillation), one defines

$$\gamma \equiv \frac{\omega^2 a_0}{\epsilon g} > 0 \tag{1.2}$$

and

$$\beta \equiv \frac{\omega^2-\omega_1^2}{2\epsilon\omega_1^2}. \tag{1.3}$$

The frequency ω approximates the natural frequency $\omega_1 = (gk \tanh kd)^{\frac{1}{2}}$, where k is the carrier wavenumber. β can have either sign: for $kd \rightarrow \infty, \beta < 0$ whereas for finite kd-values $\beta > 0$ is possible. In addition, since

$$B \equiv T+kd(1-T^2), \tag{1.4}$$

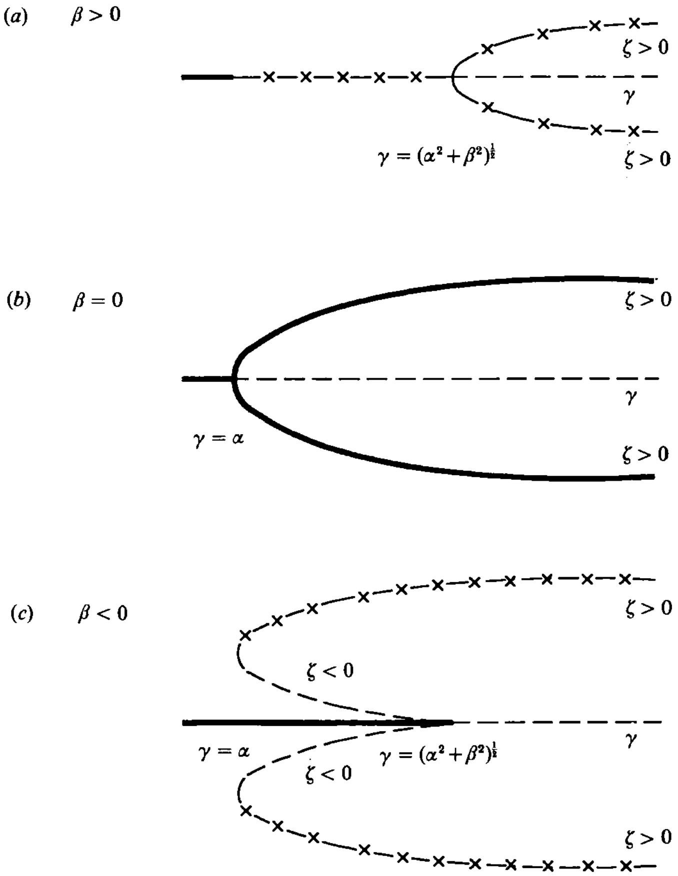


FIGURE 1. Bifurcation diagrams for X-independent stationary solutions for (*a*) $\beta > 0$, (*b*) $\beta = 0$, and (*c*) $\beta < 0$. Stable (——), unstable (----), and modulationally unstable (—×—×—×—) branches can be recognized.

with $T \equiv \tanh kd$, we always have $B > 0$. On the other hand,

$$A \equiv \tfrac{1}{8}(6T^4-5T^2+16-9T^{-2}) \tag{1.5}$$

is a monotonically increasing function of kd with negative values for $T \rightarrow 0$, and $A \rightarrow 1$ for $T \rightarrow \pm 1$. Thus A can have either sign, but for solitary wave solutions $A > 0$ is required since $B > 0$. Without loss of generality in the following we set $A = B = 1$. By rescaling the amplitude, the X-coordinate, as well as α, β, and γ, we can always obtain this simplification for $A, B > 0$.

When looking for X-independent solutions we can summarize some well-known results. First $r \equiv 0$ is always a solution. For $\alpha/\gamma < 1$ (necessary condition) additional solutions $|r| \exp(\mathrm{i}\varphi)$ can appear with

$$|r|^2 = -\beta \mp (\gamma^2-\alpha^2)^{\frac{1}{2}} \tag{1.6}$$

and

$$\cos 2\varphi = \pm\left(1-\frac{\alpha^2}{\gamma^2}\right)^{\frac{1}{2}}. \tag{1.7}$$

Thus, if $\beta < 0$ four solutions appear in the region $(\beta^2+\alpha^2)^{\frac{1}{2}} > \gamma > \alpha$, and only two of them remain for $\gamma > (\beta^2+\alpha^2)^{\frac{1}{2}}$. In the other case ($\beta \geqslant 0$), two X-independent solutions are always present for $\gamma > \alpha$. The situation is schematically shown in figure 1.

In that figure we have also included the stability results. The latter can be obtained in a straightforward manner within a linear stability analysis. We distinguish between stable, unstable, and modulationally unstable X-independent solutions. In our notation, stability and instability are first decided within a completely X-independent model. When the so-far stable solution becomes unstable with respect to X-dependent perturbations we call it modulationally unstable. To distinguish the various bifurcation branches it is advisable to introduce a new parameter

$$\zeta \equiv -\gamma \cos 2\varphi. \tag{1.8}$$

Using the latter, we find, for example, the instability criterion for the solution (1.6) and (1.7) in the form

$$\beta^2-[\kappa^2+\beta-2\zeta]^2 > 0, \tag{1.9}$$

where κ is the wavenumber of the (modulational) perturbations. From (1.6), i.e. $|r|^2 = -\beta+\zeta$, and (1.9) the corresponding conclusions summarized in figure 1 follow in a straightforward manner. Even simpler are the (in)stability arguments for the $|r| = 0$ solution. The rather trivial analysis leads to the instability criterion

$$\gamma^2-(\beta-\kappa^2)^2 > \alpha^2, \tag{1.10}$$

which completes the instability discussions for the X-independent solutions of (1.1). A final remark is appropriate with respect to the case $\beta = 0$. Note that for $\kappa \neq 0$ the two branches marked by $\zeta > 0$ correspond to a critical base in the sense of Lyapunov, where higher nonlinearities will determine the stability properties.

We continue this introductory part by posing the question of whether other X-independent solutions, i.e. limit cycles according to the Poincaré–Bendixson theorem (Guggenheimer & Holmes 1983), exist. Writing

$$r \equiv a+ib \tag{1.11}$$

we obtain from (1.1)

$$\dot{a} = -\alpha a-a^2b-b^3-\beta b+\gamma b, \tag{1.12}$$

$$\dot{b} = -\alpha b+ab^2+a^3+\beta a+\gamma a, \tag{1.13}$$

where the dot designates the time derivative. From these two equations obviously

$$\frac{\partial \dot{a}}{\partial a}+\frac{\partial \dot{b}}{\partial b} = -2\alpha < 0 \tag{1.14}$$

follows which, because of the logarithmic contraction for the area within a closed trajectory (Lichtenberg & Liebermann 1983; Miles 1984*b*), excludes the possibility of limit cycles.

One of the main conclusions of Miles (1984*b*) was that in the present system solitary waves can be parametrically excited. The existence of localized solutions was discussed and the bifurcation diagram was found. By some approximate stability method (Makhankov 1978; Whitham 1974) which we call the variation-of-action method (Laedke & Spatschek 1979) a completely stable solitary wave branch was predicted. As we shall elucidate in the next section, this soliton branch should be similar to the $\zeta > 0$ branches shown in figure 1.

In this paper we want to emphasize the following points: (i) An exact instability

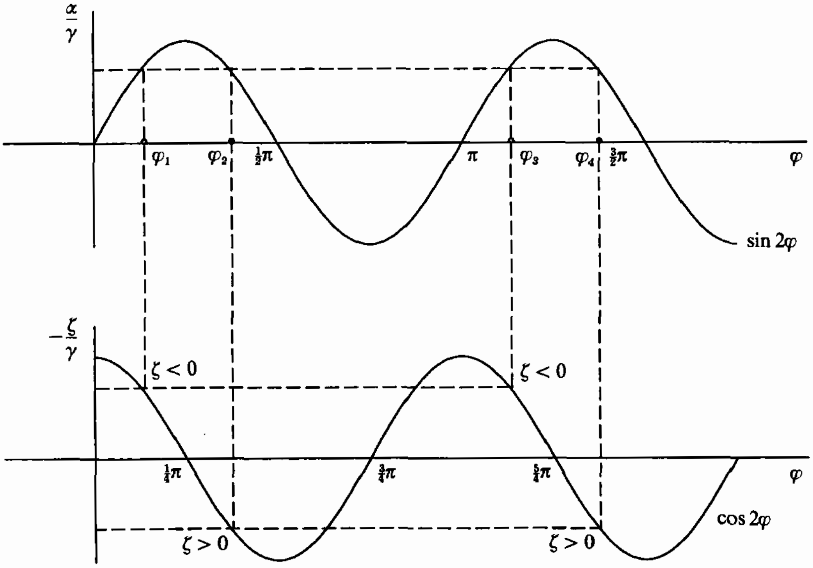


FIGURE 2. Construction of four possible phases at maximum from (1.8) and (2.3). Note that for $\beta \geqslant 0$ only the solutions φ_2 and φ_4, corresponding to $\zeta > 0$, are possible because of inequality (2.5).

calculation is possible in the region of interest. (ii) Existing solitary wave solutions are not stable in the whole parameter regime. (iii) An instability occurs which is expected to develop into a stable cnoidal-wave-type solution.

In order to demonstrate these conclusions, the paper is organized in the following way. In the next section we present the localized solitary wave solutions of (1.1). Their stability behaviour is analysed in §3 by variational principles. One special case cannot be treated by this method: its analysis is the main part of §4. All these analytical investigations are supplemented by a numerical solution of (1.1) in the relevant parameter regime. The numerics not only confirms the mathematical predictions; it also shows the nonlinear development of the instability which, at the present time, is beyond any analytical tractability.

2. Parametrically excited solitary waves

The non-trivial stationary solitary wave solutions of (1.1) (for $A = B = 1$), first presented by Miles (1984*b*), can be calculated in the following way. We substitute

$$r \equiv \psi_s(X) = G\,\mathrm{e}^{\mathrm{i}\varphi} \tag{2.1}$$

into (1.1) to obtain for G

$$\partial_X^2 G + G^3 + (\beta - \zeta)\, G = 0, \tag{2.2}$$

where ζ is as defined in (1.8). The imaginary part of (1.1) leads to the condition

$$\frac{\alpha}{\gamma} = \sin 2\varphi. \tag{2.3}$$

From (2.3) we immediately have the existence condition

$$\gamma > \alpha. \tag{2.4}$$

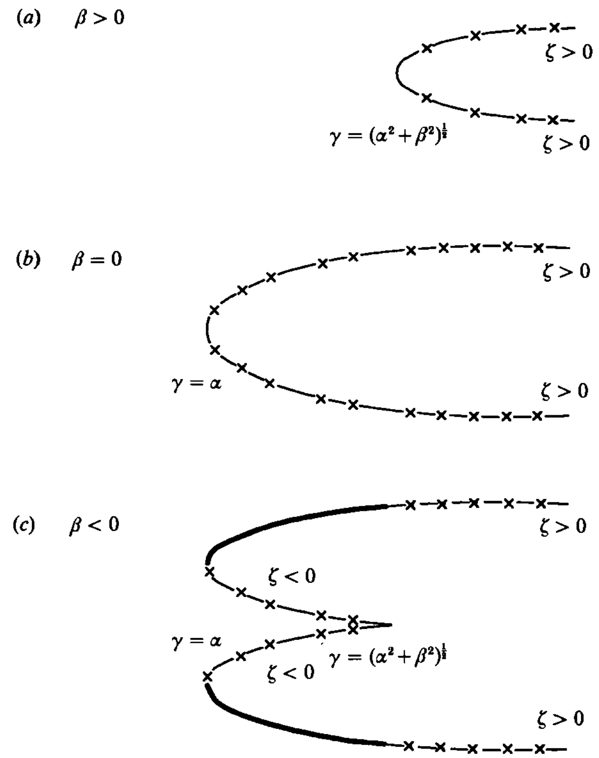


FIGURE 3. Bifurcation diagrams for solitary wave solutions. These should be compared to figure 1.

Demanding localized solutions of (2.2), the requirement

$$\beta < \zeta \tag{2.5}$$

is necessary. Then (2.2) has the well-known solitary wave solution

$$G = [2(\zeta-\beta)]^{\frac{1}{2}}\operatorname{sech}[(\zeta-\beta)^{\frac{1}{2}}X]. \tag{2.6}$$

Inequalities (2.4) and (2.5) lead to existence regions similar to those of the branches in figure 1 labelled by $\zeta > 0$ or $\zeta < 0$. Solutions of (1.8) and (2.3) are shown graphically in figure 2.

For $\beta \geqslant 0$ we have two principle solutions, whereas for $\beta < 0$ even four solutions are possible as long as inequality (2.5) is satisfied. The corresponding bifurcation diagrams are shown in figure 3 which should be compared to figure 1. The additional information contained in figure 3, i.e. the stability or the instability of the various branches, respectively, is not yet available. Its derivation is the contents of the following sections.

3. Instability by variational methods

We next perturb the solitary wave solution (2.1) in the form

$$r = (G+a+\mathrm{i}b)\,\mathrm{e}^{\mathrm{i}\varphi}, \tag{3.1}$$

to obtain the following dynamical equations for a and b:

$$\partial_t a = \mathrm{H}_+ b - 2\zeta b, \tag{3.2}$$

$$\partial_t b = -\mathrm{H}_- a - 2\alpha b. \tag{3.3}$$

Here, the Schrödinger operators

$$\mathrm{H}_+ \equiv -\partial_X^2 - G^2 + \zeta - \beta, \tag{3.4}$$

$$\mathrm{H}_- \equiv -\partial_X^2 - 3G^2 + \zeta - \beta \tag{3.5}$$

have been introduced. Their spectral properties are well-known: H_+ possesses the kernel function G, i.e.

$$\mathrm{H}_+ G = 0; \tag{3.6}$$

the continuum starts at $\eta^2 \equiv \zeta - \beta > 0$:

$$\mathrm{H}_+ \int^X G\,\mathrm{d}X' = \eta^2 \int^X G\,\mathrm{d}X'. \tag{3.7}$$

On the other hand, H_- has a negative eigenvalue,

$$\mathrm{H}_- G^2 = -3\eta^2 G^2, \tag{3.8}$$

and the kernel function $\partial_X G \equiv G_X \equiv \partial G/\partial X$, i.e.

$$\mathrm{H}_- \frac{\partial G}{\partial X} = 0. \tag{3.9}$$

For the following calculations it will be more appropriate to use

$$\mathrm{H} \equiv \mathrm{H}_+ - 2\zeta \tag{3.10}$$

instead of H_+. Then (3.2) is replaced by

$$\dot{a} = \mathrm{H} b. \tag{3.11}$$

We now summarize the basic properties of the Schrödinger operators H and H_-. For $\zeta < 0$, H is positive definite because of the property (3.6). In addition, H_- can be negative for even functions because (3.9) holds for an eigenfunction with one node. On the other hand, for $\zeta > 0$ the situation is different. H is negative for odd functions provided $\eta^2 - 2\zeta < 0$. This statement follows from the continuum limit (3.7). From the definition of $\eta^2 = \zeta - \beta$ we can write the latter condition in the form $\beta > -\zeta$. Thus for $\beta \geqslant 0$ the fact that H can be negative for odd functions is straightforward. In the case $\beta < 0$, we need the extra condition $\gamma^2 > \alpha^2 + \beta^2$ for a negative H. The operator H_- (for $\zeta > 0$) is positive definite for odd functions, being orthogonal to the kernel function G_X (see (3.9)).

These properties suggest combining (3.3) and (3.11) in the following forms:

(*a*) for $\zeta < 0$ $$\partial_t^2 a \equiv \ddot{a} = -\mathrm{H}\mathrm{H}_- a - 2\alpha\dot{a}, \tag{3.12}$$

(*b*) for $\zeta > 0$ $$\partial_t^2 b \equiv \ddot{b} = -\mathrm{H}_- \mathrm{H} b - 2\alpha\dot{b}. \tag{3.13}$$

On introducing $$\tilde{a} \equiv a\,\mathrm{e}^{\alpha t}, \tag{3.14}$$

and $$\tilde{b} \equiv b\,\mathrm{e}^{\alpha t}, \tag{3.15}$$

(3.12) and (3.13) become $$\ddot{\tilde{a}} = -\mathrm{HH}_{-}\tilde{a} + \alpha^2\tilde{a}, \tag{3.16}$$

$$\ddot{\tilde{b}} = -\mathrm{H}_{-}\mathrm{H}\tilde{b} + \alpha^2\tilde{b}, \tag{3.17}$$

respectively. In the following we use (3.16) for $\zeta < 0$ and even perturbations $\tilde{a}$ and (3.17) for $\zeta > 0$ and odd functions $\tilde{b}$ in the subspace perpendicular to G_X, i.e.

$$\int \tilde{b} G_X \,\mathrm{d}X \equiv \langle \tilde{b} \,|\, G_X \rangle = 0. \tag{3.18}$$

With these restrictions both equations can be considered to be of the same type:

$$\ddot{\tilde{\varphi}} = -\mathrm{PN}\tilde{\varphi} + \alpha^2\tilde{\varphi}, \tag{3.19}$$

where P is a positive operator and N has a negative eigenvalue. (For $\zeta < 0 : \mathrm{P} \equiv \mathrm{H}$, $\mathrm{N} \equiv \mathrm{H}_{-}$, and the functions $\tilde{\varphi}$ are even. For $\zeta > 0 : \mathrm{P} \equiv \mathrm{H}_{-}, \mathrm{N} \equiv \mathrm{H}$, and the functions $\tilde{\varphi}$ are odd and perpendicular to G_X, i.e. $\langle \tilde{\varphi} \,|\, G_X \rangle = 0$.) In Appendix A we discuss for the present case how a dynamical equation of the form (3.19) leads – with some restrictions for the test functions ψ – to the variational formulation (Blaha, Laedke & Spatschek 1987) for the exponential growth rate $\tilde{\Gamma}$,

$$\tilde{\Gamma}^2 = \alpha^2 + \sup_{\psi} \frac{-\langle \psi | \mathrm{N} | \psi \rangle}{\langle \psi | \mathrm{P}^{-1} | \psi \rangle}. \tag{3.20}$$

When applied to our original problem (3.12) and (3.13), we clearly find (for more details see Appendix B) that because of the transformations (3.14) and (3.15) formula (3.20) will predict instability for (i) $\zeta < 0$, as well as (ii) $\zeta > 0$ and $\beta > -\zeta$.

If we look at figure 3, we have thus proven (i) instability for the $\zeta < 0$ branches in all cases of β, and (ii) instability for the $\zeta > 0$ branches except for $\beta \leqslant 0$ in the region $\alpha^2 < \gamma^2 < \alpha^2 + \beta^2$. We have therefore to conclude that the parametrically excited solitary waves are not stable in the whole parameter regime. For example, for $\beta < 0$ and $\gamma^2 > \alpha^2 + \beta^2$ an instability occurs which, to the best of our knowledge, has not been discussed so far in the literature.

For the interpretation of that instability one cannot rely on the arguments for parametric instabilities of plane waves (see figure 1). If the physical picture for the latter were also to apply for solitary waves in general, the appearance of a stable solitary wave as shown in figures 3 and 4 would not be understandable.

4. Stability in the region $\beta < 0, \zeta > 0$, and $\alpha^2 < \gamma^2 < \alpha^2 + \beta^2$

Let us now consider the still-unsolved case of the existence of stable parametrically excited solitary waves in Faraday resonance. We demonstrate stability in the region $\beta < 0, \zeta > 0$, and $\alpha^2 < \gamma^2 < \alpha^2 + \beta^2$ in two ways: first by perturbation theory and secondly by numerics (see §5). In this section we present the simple but powerful perturbation scheme which has been successfully applied to other soliton problems by Zakharov, Kuznetsov & Rubenchik (1986).

Before going into the details of the calculation let us mention one important physical point. When comparing with the experimental results by Wu *et al.* (1984), we can recognize that upper and lower bounds for the driver amplitude are observed experimentally. They correspond to the limitation $\alpha^2 < \gamma^2 < \alpha^2 + \beta^2$. Furthermore, when capillary effects are included (Miles 1984*a*), the region $\beta < 0$ effectively means $\omega < \omega_1(1+\sigma)^{\frac{1}{2}}$, where σ is the surface tension. Thus the right boundary of the driving frequency observed in the experiment also agrees with the predicted existence region $\beta < 0$ for stable solitary waves.

Let us consider the region $|\zeta| \ll |\beta|$ which can be realized by $\gamma \approx \alpha$ and $|\beta| \gg \alpha$. In this region, we rewrite (3.2) and (3.3) in the form

$$\partial_t a = \hat{\mathrm{H}}_+ b - \zeta b, \tag{4.1}$$

$$\partial_t b = -\hat{\mathrm{H}}_- a - \zeta a - 2\alpha b, \tag{4.2}$$

where

$$\hat{\mathrm{H}}_+ \equiv -\partial_X^2 - G^2 - \beta, \tag{4.3}$$

$$\hat{\mathrm{H}}_- \equiv -\partial_X^2 - 3G^2 - \beta. \tag{4.4}$$

The idea is to use $\partial_t \sim \Gamma$, $|\zeta|$, and α as small quantities. When combining (4.1) and (4.2) in the form

$$\partial_t^2 b = \Gamma^2 b = -(\hat{\mathrm{H}}_- + \zeta)(\hat{\mathrm{H}}_+ - \zeta)\, b - 2\alpha\Gamma b \tag{4.5}$$

we obtain at lowest order the equation

$$0 = -\hat{\mathrm{H}}_- \hat{\mathrm{H}}_+ b_0. \tag{4.6}$$

Its odd solution is

$$b_0 = \hat{\mathrm{H}}_+^{-1} G_X, \tag{4.7}$$

where G is given by (2.6) for $\zeta = 0$. Within the scaling $\Gamma^2 \sim |\zeta| \sim \alpha\Gamma \sim \epsilon^2$, where ϵ is a smallness parameter, we develop the perturbation series for $b = b_0 + \epsilon^2 b_2 + \dots$. The contribution b_2 follows from

$$\Gamma^2 \hat{\mathrm{H}}_+^{-1} G_X = -\hat{\mathrm{H}}_- \hat{\mathrm{H}}_+ b_2 + \zeta \hat{\mathrm{H}}_- b_0 - \zeta G_X - 2\alpha\Gamma \hat{\mathrm{H}}_+^{-1} G_X; \tag{4.8}$$

this equation has the solvability condition

$$\Gamma^2 + 2\alpha\Gamma = -\zeta \frac{\langle G_X | G_X \rangle}{\langle G_X | \hat{\mathrm{H}}_+^{-1} | G_X \rangle}. \tag{4.9}$$

Note that

$$g \equiv \frac{\langle G_X | G_X \rangle}{\langle G_X | \hat{\mathrm{H}}_+^{-1} | G_X \rangle} > 0 \tag{4.10}$$

is a positive constant which can be evaluated without any difficulties. Thus the solution of (4.9),

$$\Gamma = -\alpha \pm (\alpha^2 - g\zeta)^{\frac{1}{2}}, \tag{4.11}$$

clearly tells us that (i) in accordance with the results of the previous section for $\zeta < 0$ an instability occurs, while (ii) for $\zeta > 0$ the stationary state is stable with respect to odd perturbations.

Let us now investigate the even solution

$$b_0 = G \tag{4.12}$$

of (4.6). Using the same scaling as in the previous case, we find instead of (4.8)

$$\Gamma^2 G = -\hat{\mathrm{H}}_- \hat{\mathrm{H}}_+ b_2 + \zeta \hat{\mathrm{H}}_- G - 2\alpha\Gamma G. \tag{4.13}$$

Its non-trivial solvability condition turns out to be

$$\Gamma^2 + 2\alpha\Gamma = \zeta \frac{\langle G | G \rangle}{\langle G | \hat{\mathrm{H}}_-^{-1} | G \rangle}. \tag{4.14}$$

Note that
$$g \equiv -\frac{\langle G|G\rangle}{\langle G|\hat{\mathrm{H}}_{-}^{-1}|G\rangle} > 0 \tag{4.15}$$

for the following reason. We have
$$\hat{\mathrm{H}}_{-}^{-1}G = -\frac{\partial G}{\partial(-\beta)} \tag{4.16}$$

and
$$\langle G|\hat{\mathrm{H}}_{-}^{-1}|G\rangle = -\frac{1}{2}\frac{\partial}{\partial(-\beta)}\langle G|G\rangle = -(-\beta)^{-\frac{1}{2}} < 0. \tag{4.17}$$

Thus, we again arrive at (4.11) and for $\zeta > 0$ the stability with respect to even perturbations follows.

We are aware of the fact that this is not a proof of stability for the whole region $\beta < 0, \zeta > 0$, and $\alpha^2 < \gamma^2 < \alpha^2+\beta^2$ in the strict mathematical sense. Going back to figure 3(*c*), we have just shown stability for the far left parts of the $\zeta > 0$ branches. Since we complete this investigation by a numerical procedure which clearly shows that (for $\beta < 0$) a transition from stable to unstable behaviour occurs in the $\zeta > 0$ branches at $\gamma = (\alpha^2+\beta^2)^{\frac{1}{2}}$ we do not aim here to construct a Lyapunov functional. Just to give the main idea of the latter procedure, we make the following remarks. One can prove that H is positive definite for odd functions whereas H_{-} is positive semidefinite for odd functions in the region being under consideration in this section. Thus, for odd functions a, multiply (3.12) by H^{-1} from the left in order to get a monotonically decreasing functional in time. The problem is that these considerations are restricted to odd functions whereas for even perturbations no successful procedure is known. We also shall not discuss the transition point from $\zeta < 0$ to $\zeta > 0$ which is a critical case in the sense of Lyapunov. It turns out that the critical case is nonlinearly unstable. It is straightforward to prove – starting from the basic equation (1.1) – that
$$\partial_t[(\mathrm{Re}\, r)^2+(\mathrm{Im}\, r)^2] = \gamma(\mathrm{Im}\, r)^2. \tag{4.18}$$

Thus solutions with $\mathrm{Im}\, r \neq 0$ initially will grow.

5. Numerical manifestation of the analytical predictions

Equation (1.1) has been solved numerically by a nonlinear semi-implicit unitary Crank–Nicholson scheme (Spatschek *et al.* 1989). This allowed us (i) to test the stability predictions of §4; (ii) to verify the unstable behaviour in the complementary parameter regime, as pointed out in §3, and (iii) to look for the nonlinear development of an unstable solitary wave.

First, we investigated the parameter regime where stability is predicted. For $\beta < 0$ and $\zeta > 0$ several runs were performed in the region $\alpha^2 < \gamma^2 < \alpha^2+\beta^2$. All of them confirmed the analytical predictions. A typical result is shown in figure 4. Secondly, we took solitary wave solutions as initial conditions in the unstable parameter regime. As expected, they all became unstable within a finite time ($t \approx 20$). Finally, and most interesting, we followed the (nonlinear) time-development of an unstable solitary wave. Typical runs are shown in figures 5 and 6. At the first stage ($t \lesssim 50$) the breakup of the unstable soliton is clearly visible (figure 5). At a later time, a nice spatially coherent structure in space develops. It starts from the centre ($X = 0$) and spreads to larger X-values. It can be interpreted as a stable cnoidal wave (see figure 6).

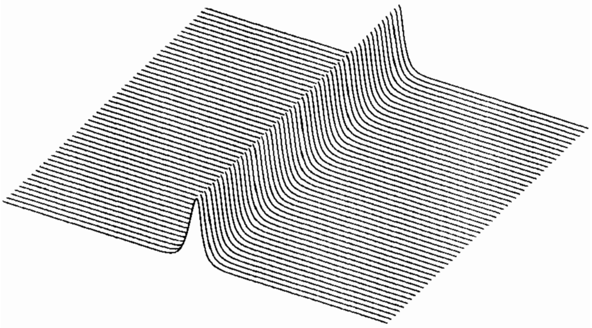


FIGURE 4. Space–time plot of a solitary wave (absolute value of the amplitude) for $\beta = -1, \alpha = 1, \gamma = 1.1$. The solitary wave is stable, at least for the time of computation ($0 \leqslant t < 100$).

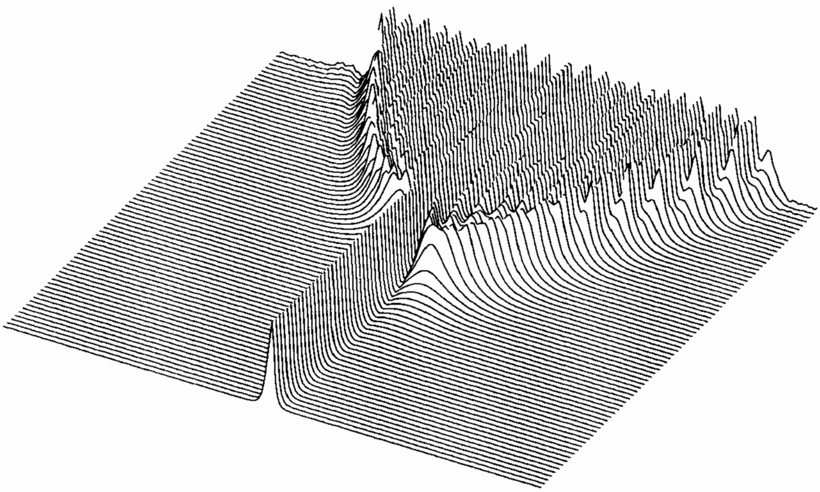


FIGURE 5. Space–time plot of the absolute value of the amplitude of an unstable solitary wave for $\beta = -1, \alpha = 1, \gamma = 1.6$. Initial phase of the instability for $-60 < X < 60$ and $0 < t < 75$.

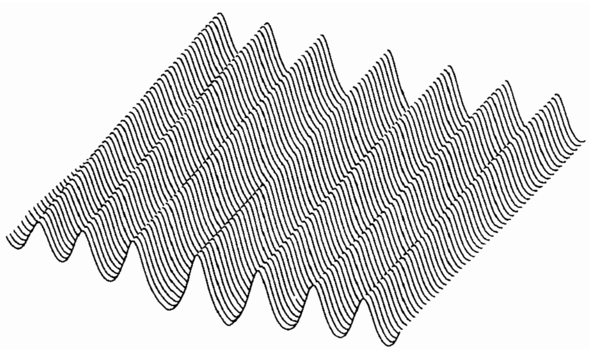


FIGURE 6. The same as figure 5 for large times and small X. A new state appears which can be interpreted as a cnoidal wave. Here the practically unchanged distribution for $-15 < X < 15$ is shown for the times $50 < t < 65$.

This work was supported by the Deutsche Forschungsgemeinschaft through SFB 237 and the Ministerium für Wissenschaft und Forschung des Landes Nordrhein-Westfalen. Discussions with Dr Funakoshi are gratefully acknowledged.

Appendix A. A variational principle for (3.19)

Since P is a positive operator, we rewrite (3.19) in the form

$$\mathrm{P}^{-1}\ddot{\tilde{\varphi}} = -(\mathrm{N}-\alpha^2\mathrm{P}^{-1})\tilde{\varphi}, \tag{A 1}$$

where the operator

$$\mathrm{F} \equiv -(\mathrm{N}-\alpha^2\mathrm{P}^{-1}) \tag{A 2}$$

is also positive. Multiplying

$$\mathrm{P}^{-1}\ddot{\tilde{\varphi}} = \mathrm{F}\tilde{\varphi} \tag{A 3}$$

from the left by $\dot{\tilde{\varphi}}$ and integrating over space we find

$$\langle\dot{\tilde{\varphi}}|\,\mathrm{P}^{-1}|\dot{\tilde{\varphi}}\rangle = \langle\tilde{\varphi}|\,\mathrm{F}|\tilde{\varphi}\rangle + \mathrm{const.} \tag{A 4}$$

The constant of (time-) integration can be set to zero if we choose appropriate initial conditions, i.e. at $t = 0$ we demand $\dot{\tilde{\varphi}} = \tilde{\Gamma}\tilde{\varphi}$. Since F is positive, the constant $\tilde{\Gamma}$ follows without problems from

$$\tilde{\Gamma}^2 = \frac{\langle\tilde{\varphi}_0|\mathrm{F}|\tilde{\varphi}_0\rangle}{\langle\tilde{\varphi}_0|\mathrm{P}^{-1}|\tilde{\varphi}_0\rangle}, \tag{A 5}$$

where $\tilde{\varphi}_0$ is the initial distribution.

Also multiplying (A 1) from the left by $\tilde{\varphi}$ and integrating over space leads to

$$\langle\tilde{\varphi}|\mathrm{P}^{-1}|\ddot{\tilde{\varphi}}\rangle = \langle\tilde{\varphi}|\mathrm{F}|\tilde{\varphi}\rangle. \tag{A 6}$$

Since all the operators are self-adjoint, we can combine (A 4) and (A 6) to give

$$\partial_t^2 L = 2\langle\tilde{\varphi}|\mathrm{F}|\tilde{\varphi}\rangle, \tag{A 7}$$

where

$$L \equiv \tfrac{1}{2}<\tilde{\varphi}|\mathrm{F}|\tilde{\varphi}\rangle. \tag{A 8}$$

The following rearrangement using the Schwarz inequality,

$$\partial_t^2 L = 2\langle\dot{\tilde{\varphi}}|\mathrm{P}^{-1}|\dot{\tilde{\varphi}}\rangle = \frac{2\langle\dot{\tilde{\varphi}}|\mathrm{P}^{-1}|\dot{\tilde{\varphi}}\rangle\langle\tilde{\varphi}|\mathrm{P}^{-1}|\tilde{\varphi}\rangle}{\langle\tilde{\varphi}|\mathrm{P}^{-1}|\tilde{\varphi}\rangle} \geqslant 2\frac{\langle\dot{\tilde{\varphi}}|\mathrm{P}^{-1}|\tilde{\varphi}\rangle^2}{\langle\tilde{\varphi}|\mathrm{P}^{-1}|\tilde{\varphi}\rangle}, \tag{A 9}$$

proves the result

$$\partial_t\left(\frac{\partial_t L}{L}\right) \geqslant 0. \tag{A 10}$$

It has the solution

$$L \geqslant L_0\,\mathrm{e}^{\tilde{\Gamma}t}. \tag{A 11}$$

Since P and P^{-1} are positive operators, unstable perturbations exist with exponential growth rate $\tilde{\Gamma}$. Going back to (A 5) we can maximize with respect to the initial distributions to obtain the largest exponential growth rate:

$$0 < \tilde{\Gamma}^2 \leqslant \sup_{\tilde{\psi}} \frac{\langle\tilde{\psi}|\mathrm{F}|\tilde{\psi}\rangle}{\langle\tilde{\psi}|\mathrm{P}^{-1}|\tilde{\psi}\rangle}. \tag{A 12}$$

When we use the definition (A 2), (3.20) follows. It should be mentioned that the right-hand side of inequality (A 12) actually represents the maximum exponential growth rate.

Appendix B. Growth rates in the unstable cases

In §3 we have argued that all the unstable cases lead to equations of the type (3.19) with the result (3.20). A few additional remarks are necessary since we have subsidiary conditions to obey. Let us start with $\zeta > 0$ when we use (3.17), i.e.

$$\ddot{\tilde{b}} = -\mathrm{H}_- \mathrm{H}\tilde{b} + \alpha^2 \tilde{b}. \tag{B 1}$$

Let us restrict $\tilde{b}$ to odd functions with $\langle \tilde{b} \,|\, G_X \rangle = 0$, so that H_- is positive definite. As discussed already, H is negative in the region of interest $\gamma^2 > \alpha^2 + \beta^2$. For the second-order differential equation (B 1) we start with

$$\langle \tilde{b} \,|\, G_X \rangle_{t=0} = \langle \dot{\tilde{b}} \,|\, G_X \rangle_{t=0} = 0. \tag{B 2}$$

Then the differential equation (B 1) and (3.9) tell us that $\langle \tilde{b} \,|\, G_X \rangle$ remains zero in time. Thus we can use $\langle \tilde{\varphi} \,|\, G_x \rangle = 0$ as a consistent subsidiary condition for all times. Next, we have to say a few words about the inversion procedure applied in (A 1). Of course, we can always add an arbitrary function from the kernel of $\mathrm{P} \equiv \mathrm{H}_-$, i.e. (A 1) reads in the present case

$$\mathrm{H}_-^{-1} \ddot{\tilde{b}} = -\mathrm{H}\tilde{b} + \alpha^2 \mathrm{H}_-^{-1} \tilde{b} + \mu G_X. \tag{B 3}$$

Here μ is a free parameter. We fix it, however, by the requirement that we remain in the subspace orthogonal to G_X. Thus

$$\mu = \frac{\langle G_X | \mathrm{H} | \tilde{b} \rangle}{\langle G_X \,|\, G_X \rangle}. \tag{B 4}$$

The rest follows along the lines outlined in Appendix A. The maximum growth rate for $\zeta > 0$ in the region $\gamma^2 > \alpha^2 + \beta^2$ is

$$\tilde{\Gamma}^2 \equiv (\Gamma + \alpha)^2 = \alpha^2 + \sup_{\substack{\psi\,\mathrm{odd} \\ \langle \psi | G_X \rangle = 0}} \frac{-\langle \psi | \mathrm{H} | \psi \rangle}{\langle \psi | \mathrm{H}_-^{-1} | \psi \rangle}. \tag{B 5}$$

On the other hand, for $\zeta < 0$ a similar calculation leads to

$$\tilde{\Gamma}^2 \equiv (\Gamma + \alpha)^2 + \sup_{\psi\,\mathrm{even}} \frac{-\langle \psi | \mathrm{H}_- | \psi \rangle}{\langle \psi | \mathrm{H}^{-1} | \psi \rangle}, \tag{B 6}$$

when (3.16) is used. Both formulae prove the instability results summarized in §3.

REFERENCES

Blaha, R., Laedke, E. W. & Spatschek, K. H. 1987 Maximum growth rates for packets of waves in water of finite depth. *Phys. Fluids* **30**, 264–266.

Guggenheimer, J. & Holmes, P. H. 1983 *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector fields*. Springer.

Laedke, E. W. & Spatschek, K. H. 1979 On the applicability of the variation of action method to some one-field solitons. *J. Math. Phys.* **20**, 1838–1841.

Larraza, A. & Puttermann, S. 1984 Theory of non-propagating surface-wave solitons. *J. Fluid Mech.* **148**, 443–449.

Lichtenberg, A. J. & Lieberman, M. A. 1983 *Regular and Stochastic Motion*, p. 383. Springer.

Makhankov, V. G. 1978 Dynamics of of classical solitons. *Phys. Rep.* **35**, 1–128.

Miles, J. W. 1984*a* Nonlinear Faraday resonance. *J. Fluid Mech.* **146**, 285–302.

Miles, J. W. 1984*b* Parametrically excited solitary waves. *J. Fluid Mech.* **148**, 451–460.

SPATSCHEK, K. H., PIETSCH, H., LAEDKE, E. W. & EICKERMANN, TH. 1989 On the role of soliton solutions in temporal chaos: examples for plasmas and related systems. In *Singular Behavior and Nonlinear Dynamics*, pp. 555–564. World Scientific.

WHITHAM, G. B. 1974 *Linear and Nonlinear Waves*, pp. 601–603. Wiley.

WU, J., KEOLIAN, R. & RUDNICK, I. 1984 Observation of a non-propagating hydrodynamic soliton. *Phys. Rev. Lett.* **52**, 1421–1424.

ZAKHAROV, V. E., KUZNETSOV, E. A. & RUBENCHIK, A. M. 1986 Soliton stability. In *Solitons*, pp. 503–554. Elsevier.

J. Fluid Mech. (1991), *vol.* 223, *pp.* 603–629
Printed in Great Britain

A numerical study of nonlinear energy fluxes due to wave–wave interactions
Part 1. Methodology and basic results

By D. RESIO[1] AND W. PERRIE[2]

[1] Department of Oceanography, Florida Institute of Technology, Melbourne, Florida, USA
[2] Physical and Chemical Sciences Scotia–Fundy Region, Department of Fisheries and Oceans, Bedford Institute of Oceanography, Dartmouth, Nova Scotia, Canada

(Received 21 February 1989 and in revised form 17 July 1990)

Nonlinear transfer due to wave–wave interactions was first described by the Boltzmann integrals of Hasselmann (1961) and has been the subject of modelling ever since. We present an economical method to evaluate the complete integral, which uses selected scaling properties and symmetries of the nonlinear energy transfer integrals to construct the integration grid. An important aspect of this integration is the inherent smoothness and stability of the computed nonlinear energy transfer. Energy fluxes associated with the nonlinear energy transfers and their behaviour within the equilibrium range are investigated with respect to high-frequency power law, peak frequency, peakedness, spectral sharpness and angular spreading. We also compute the time evolution of the spectral energy and the nonlinear energy transfers in the absence of energy input by wind or dissipated by wave breaking. The response of nonlinear iterations to perturbations is given and a formulation of relaxation time in the equilibrium range is suggested in terms of total equilibrium range energy and the nonlinear energy fluxes within the equilibrium range.

1. Introduction

The pioneering work of Hasselmann in the early 1960s (Hasselmann 1961, 1963*a*, 1963*b*) established a theoretical framework for estimating the net transfer of energy among different frequency-direction components in a wave spectrum. Unfortunately, the resulting integro-differential equation proved to be cumbersome in terms of its functional structure, necessitating that it be evaluated numerically. Early numerical integrations of this equation required very laborious efforts and were limited by computer systems available in the 1960s. With the Joint North Sea Wave Project (JONSWAP) of Hasselmann *et al.* (1973), in which the pattern of spectral evolution along a fetch was observed to agree at least qualitatively with that predicted by these 'wave–wave' interactions, substantial interest was focused on the accurate numerical solution of this equation.

Longuet-Higgins (1976) and Fox (1976) used a simplified approach to estimate the nonlinear wave–wave energy transfer in the vicinity of the spectral peak. Their results were not completely consistent with the earlier approximate computations of Sell & Hasselmann (1972). Moreover, the determination of essential features such as the central minimum and the transition from positive to negative transfer on the right of the spectral peak was unclear from either work. Subsequent studies by Webb

(1978) and Masada (1980) derived transformed versions of the original integro-differential equation which proved to be more adaptable to stable numerical solutions. The solutions of Webb (1978) and Masada (1980) supported the early computational results of Sell & Hasselmann (1972) and suggested that the narrow-band approximations invoked by Longuet-Higgins (1976) and Fox (1976) were somewhat limited in their applicability. More recently, Hasselmann & Hasselmann (1981) have completed a careful study of the nonlinear transfer. They exploited the symmetry of detailed balance (invariance with respect to permutations of all four wavenumbers of a quadruplet $\boldsymbol{k}_1+\boldsymbol{k}_2=\boldsymbol{k}_3+\boldsymbol{k}_4$). Their computations constitute a standard for evaluation of the nonlinear energy transfer, in terms of the detail and accuracy with which they were performed.

As mentioned previously, the need for a better understanding of the magnitude and structure of nonlinear wave–wave interaction energy transfers gained impetus following JONSWAP. Shortly thereafter, Hasselmann *et al.* (1976) considered the shape-stabilizing effect of wave–wave interactions on wave spectra and concluded that wave–wave interaction effects were so strong that wave spectra were effectively controlled, during periods of active wave generation, by a dynamic balance between wind inputs and the wave–wave interactions. Questions concerning the role of wave–wave interactions in governing spectral shapes have arisen since Hasselmann *et al.*'s (1976) study.

Postulating the stationary distribution corresponding to Kolmogorov's inertial subrange, Kitaigorodskii (1983) showed that energy fluxes due to wave–wave iterations should produce an f^{-4} equilibrium range in the spectrum, rather than the f^{-5} form assumed in the JONSWAP spectrum. This was also the result found by Zakharov & Filonenko (1966), as the exact stationary solution to the wave–wave iteration Boltzmann integral for an isotropic field of weakly nonlinear waves. However, Phillips (1985) suggested that a detailed balance of all source terms, including wind input, wave breaking and wave–wave interactions, could produce an f^{-4} equilibrium range and that knowledge of all source terms was necessary to understand the net scaling involved in establishing an equilibrium range. The numerical study of Komen, Hasselmann & Hasselmann (1984) examined the balances among all source terms in a 'fully developed' sea for frequencies extending from the vicinity of the spectral peak f_p and up to 2.5 times f_p, basing their nonlinear transfer calculation on Hasselmann & Hasselmann (1981).

Recently, Toba, Okada & Jones (1988) suggested that an energy flux must exist from high frequencies to low frequencies in the equilibrium range, as a result of their investigation of characteristics of the relaxation of a deep water wave spectrum under a decreasing wind. This is in addition to the more widely recognized flux from low frequencies to high frequencies in the Kolmogorov subrange. Therefore, a simple analogue to the one-dimensional cascade of energy in turbulence may not be appropriate for surface gravity waves.

Along with theoretical and conceptual developments related to the role of wave–wave interactions in wave generation and the influence of these interactions on spectral shape, a parallel continuing effort has been devoted to obtaining accurate parameterizations of the complete Boltzmann integral. Barnett (1968), Barnett & Sutherland (1968) and Ewing (1971) based parameterizations of complete integrals on the Neumann spectrum and Pierson–Moskowitz spectrum, respectively. Resio (1981) investigated parameterizations based on the scaling laws for f^{-5} spectra and also inherent in the complete integral. Hasselmann *et al.* (1985) and Hasselmann & Hasselmann (1985) examined approximations using empirical orthogonal functions and also parameterizations based on the superposition of a small number of discrete interaction configurations. The latter type of approximation has been adopted into the WAM model (Hasselmann *et al.* 1989). Since this model is the subject of considerable international investigation, the adequacy of the parameterization is of interest to most wave modellers.

It appears from the issues raised here that estimation of the effects of nonlinear wave–wave interactions has assumed an important role in wave generation and modelling research. In spite of this, most probably owing to the complexity of the numerical problem, few efforts have been made toward establishing a clearer formulation for some of the fundamental characteristics of the nonlinear energy transfers. The present paper will attempt to remedy this situation, at least in part. We begin by formulating an efficient numerical scheme which should assist in understanding some of the inherent scaling properties of the nonlinear energy transfer. This allows computation of the nonlinear transfer on a very fine integration grid and also achieves high numerical stability. This scheme will then be used to investigate the behaviour of nonlinear energy fluxes with respect to high-frequency power law, peak frequency, peakedness, spectral sharpness and angular spreading. We also consider spectral evolution with respect to time and the response of nonlinear wave–wave interactions to perturbations within the spectrum.

2. Evaluation of the nonlinear flux integral

To date most researchers have concentrated on solution of the 'source function' form for nonlinear wave–wave interactions. In this form, the collision integral for four resonantly interacting waves allows evaluation of the net rate of change of energy (or action) for a given wavenumber within the spectrum. Following Hasselmann (1961) the integral can be written as

$$\frac{\partial n(\boldsymbol{k}_1)}{\partial t} = \iiint \mathscr{C}^2(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3, \boldsymbol{k}_4)\, \mathscr{D}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3, \boldsymbol{k}_4)$$

$$\delta(\boldsymbol{k}_1 + \boldsymbol{k}_2 - \boldsymbol{k}_3 - \boldsymbol{k}_4)\,\delta(\omega_1 + \omega_2 - \omega_3 - \omega_4)\,\mathrm{d}\boldsymbol{k}_2\,\mathrm{d}\boldsymbol{k}_3\,\mathrm{d}\boldsymbol{k}_4, \quad (2.1)$$

where $\boldsymbol{k}_i$ is the ith interacting vector wavenumber, ω_i is the radial frequency of the ith wavenumber and $n(\boldsymbol{k}_1)$ is the action density at wavenumber $\boldsymbol{k}_1$. The coupling coefficient $\mathscr{C}^2$, is a complicated function of wavenumbers $\boldsymbol{k}_i$ and frequencies ω_i. The density function $\mathscr{D}$, varies cubically in the spectral densities and may be expressed as

$$\mathscr{D}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3, \boldsymbol{k}_4) = n(\boldsymbol{k}_1)\, n(\boldsymbol{k}_2)\, n(\boldsymbol{k}_3) + n(\boldsymbol{k}_2)\, n(\boldsymbol{k}_3)\, n(\kappa_4)$$

$$- n(\boldsymbol{k}_1)\, n(\boldsymbol{k}_3)\, n(\boldsymbol{k}_4) - n(\boldsymbol{k}_1)\, n(\boldsymbol{k}_2)\, n(\boldsymbol{k}_4). \quad (2.2)$$

An efficient form for integration is obtained by removing the delta functions from (2.1) through the transformation which results in

$$\frac{\partial n(\boldsymbol{k}_1)}{\partial t} = 2\int T(\boldsymbol{k}_1, \boldsymbol{k}_3)\, \mathrm{d}\boldsymbol{k}_3, \quad (2.3)$$

where

$$T(\boldsymbol{k}_1, \boldsymbol{k}_3) = \oint \mathscr{C}^2 \mathscr{D} \left| \frac{\partial \mathscr{W}}{\partial \boldsymbol{n}} \right| \theta(\boldsymbol{k}_1, \boldsymbol{k}_3, \boldsymbol{k}_4)\, \mathrm{d}s \quad (2.4)$$

is the nonlinear transfer integral, $\mathscr{W} = \omega_1+\omega_2-\omega_3-\omega_4$ and the frequency resonance condition is

$$\mathscr{W} = 0. \tag{2.5}$$

Unit vector $\boldsymbol{s}$ is along the interaction locus, defined in $\boldsymbol{k}_2$-space by the constraint $\mathscr{W} = 0$. Unit vector $\boldsymbol{n}$ is normal to that locus. The wavenumber resonance condition is $\boldsymbol{k}_1+\boldsymbol{k}_2-\boldsymbol{k}_3-\boldsymbol{k}_4 = 0$, and θ may be represented as

$$\theta(\boldsymbol{k}_1,\boldsymbol{k}_3,\boldsymbol{k}_4) = \begin{cases} 1 & \text{when} \quad |\boldsymbol{k}_1-\boldsymbol{k}_3| \leqslant |\boldsymbol{k}_1-\boldsymbol{k}_4| \\ 0 & \text{when} \quad |\boldsymbol{k}_1-\boldsymbol{k}_3| > |\boldsymbol{k}_1-\boldsymbol{k}_4|. \end{cases} \tag{2.6}$$

Webb (1978) showed that this provided a stable, efficient form for integration. However, it was still tedious to apply this formulation since for each different value of $\boldsymbol{k}_1$ and $\boldsymbol{k}_3$, the locus equation had to be solved and at each point along the locus in $\boldsymbol{k}_2$-space, the coupling coefficient, density term, Jacobian term and phase space volume had to be evaluated.

This problem was simplified by Tracy & Resio (1982) using a polar grid in wavenumber space with the radial coordinate spaced according to

$$\boldsymbol{k}_{m+1} = \lambda \boldsymbol{k}_m \tag{2.7}$$

where $m+1$ is the radial index shown in figure 1. It may be demonstrated that for any geometrically similar $\boldsymbol{k}_1$ and $\boldsymbol{k}_3$, for example $|\boldsymbol{k}_1'-\boldsymbol{k}_3'| = \lambda|\boldsymbol{k}_1-\boldsymbol{k}_3|$, the locus equation scales linearly in λ also. Specifically, for each point along the original locus, a geometrically similar point exists in a scaling locus such that $\boldsymbol{k}_2' = \lambda\boldsymbol{k}_2$. From the resonance condition for wavenumbers we obtain $\boldsymbol{k}_4' = \lambda\boldsymbol{k}_4$, and for each combination of four wavenumbers satisfying $(\boldsymbol{k}_1',\boldsymbol{k}_2',\boldsymbol{k}_3',\boldsymbol{k}_4') = \lambda\,(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3,\boldsymbol{k}_4)$ it follows that

$$\mathscr{C}^2(\boldsymbol{k}_1',\boldsymbol{k}_2',\boldsymbol{k}_3',\boldsymbol{k}_4') = \lambda^6\mathscr{C}^2(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3,\boldsymbol{k}_4), \tag{2.8}$$

$$|\partial\mathscr{W}'/\partial\boldsymbol{n}|^{-1} = \lambda^{\frac{1}{2}}|\partial\mathscr{W}/\partial\boldsymbol{n}|^{-1}, \tag{2.9}$$

and

$$\mathrm{d}\boldsymbol{s}' = \lambda\,\mathrm{d}\boldsymbol{s}. \tag{2.10}$$

Therefore, on the geometrically progressive polar grid of figure 1, where m_i is radial index and n_i is the angular index for the ith wavenumber, $\mathrm{d}\boldsymbol{s}$, $|\partial\mathscr{W}/\partial\boldsymbol{n}|^{-1}$ and $\mathscr{C}^2$ need only be calculated once for each different m_3-m_1 and $|n_3-n_1|$. Letting $\boldsymbol{k}_1 = (k_0, 0)$ and $\boldsymbol{k}_3$ vary over the entire grid, we initially construct a table of all possible values for $\mathrm{d}\boldsymbol{s}|\partial\mathscr{W}/\partial\boldsymbol{n}|^{-1}\mathscr{C}^2$. All other locus solutions, coupling coefficients, Jacobian terms, and phase space volumes can be obtained by appropriate rotation and multiplication of these results. For example if $|\boldsymbol{k}_1'-\boldsymbol{k}_3'| = \lambda^j|\boldsymbol{k}_1-\boldsymbol{k}_3|$, then

$$\mathrm{d}\boldsymbol{s}'\left|\frac{\partial\mathscr{W}'}{\partial\boldsymbol{n}}\right|^{-1}\mathscr{C}^2(\boldsymbol{k}_1',\boldsymbol{k}_2',\boldsymbol{k}_3',\boldsymbol{k}_4') = (\lambda)^{\frac{15}{2}j}\,\mathrm{d}\boldsymbol{s}\left|\frac{\partial\mathscr{W}}{\partial\boldsymbol{n}}\right|^{-1}\mathscr{C}^2(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3,\boldsymbol{k}_4). \tag{2.11}$$

These are exact scaling relationships inherent in the collision integral.

With the computation of $\mathscr{D}$ and integration around the locus $\boldsymbol{s}$, evaluation of the nonlinear transfer contour integral $\oint\mathscr{C}^2\mathscr{D}\,|\partial\mathscr{W}/\partial\boldsymbol{n}|\,\mathrm{d}\boldsymbol{s}$ is then complete. Integration over all $\boldsymbol{k}_3$ values gives the nonlinear transfer source function $\partial n(\boldsymbol{k}_1)/\partial t$, as indicated in (2.3). This is also usually denoted $\mathscr{S}_{\mathrm{nl}}$. The scaling geometry of figure 1 allowed Tracy & Resio (1982) to obtain integration times that were typically over an order of magnitude less than integration times on regularly spaced grids. Using additional

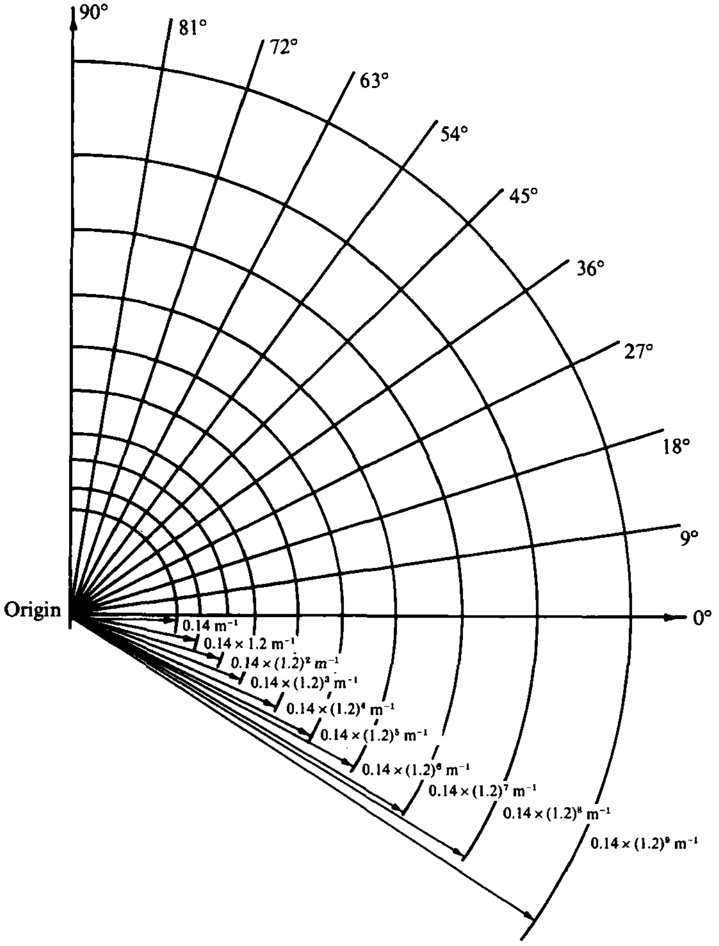


FIGURE 1. Polar grid in wavenumber space with radial intersection points spaced in the geometric progression $k_{m+1} = \lambda k_m$. This example uses $k_0 = 0.14\,\mathrm{ms}^{-1}$ and $\lambda = 1.2$.

symmetries, such as permutations in k_1 and k_3, detailed evaluations of the collision integral for the entire spectrum can be performed in 20 min run times on an IBM-PC with an accelerator board, on grid resolutions comparable to Hasselmann & Hasselmann (1981).

In the work of Kitaigorodskii (1983), Resio (1987) and Toba *et al.* (1988), it appeared informative to examine fluxes of action (or energy) past a specific frequency ω_A in addition to looking at a source function for the entire spectrum. The integral for these fluxes from high to low frequencies may be written as

$$\Gamma^-(\omega_\mathrm{A}) = \iint\oint \mathscr{C}^2\mathscr{D}\left|\frac{\partial \mathscr{W}}{\partial n}\right|^{-1} \mathrm{d}s\, H(|k_3| - k(\omega_\mathrm{A}))\, H(k(\omega_\mathrm{A}) - |k_1|)\, \mathrm{d}k_3\, \mathrm{d}k_1, \qquad (2.12)$$

where $H(x)$ is the Heaviside function, defined as

$$H(x) = \begin{cases} 1 & \text{for} \quad x \geqslant 0 \\ 0 & \text{for} \quad x < 0 \end{cases} \qquad (2.13)$$

and $k(\omega)$ is the wavenumber given by $k = \omega^2/g$. The corresponding nonlinear source function $\mathscr{S}_{\mathrm{nl}}$ can be obtained from a calculation of flux divergence whereas it is not possible to estimate the fluxes from the nonlinear source function. Flux estimates also provide a direct means of estimating all action transfers from one region of the spectrum to another. Consequently, they are useful in partitioning the percentage of action (or energy) that moves in various directions within the spectrum. The fact that we can compute positive fluxes from low to high frequencies as well as negative fluxes from high to low frequencies is very helpful in this regard.

3. Comparison of integration results to previous estimates

Most published results for the nonlinear wave–wave transfer have been restricted to spectra representable by the JONSWAP parameterization,

$$E(f) = \frac{\alpha g^2 f^{-5}}{(2\pi)^4} \exp\left(-1.25\frac{f}{f_{\mathrm{p}}}\right)\gamma^{\Xi}, \tag{3.1}$$

where

$$\Xi = \exp -\frac{1}{2}\left(\frac{f-f_{\mathrm{p}}}{\sigma f_{\mathrm{p}}}\right)^2. \tag{3.2}$$

These are typically converted to directional spectra by using the normalized $\cos^{2n}\theta$ form for angular spreading,

$$E_2(f,\theta) = E(f)\,A(n)\cos^{2n}\theta \tag{3.3}$$

where the normalization coefficient $A(n)$, satisfies

$$\int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} A(n)\cos^{2n}\theta\,\mathrm{d}\theta = 1. \tag{3.4}$$

Figures 2(*a*)–2(*d*) compare the nonlinear transfer due to wave–wave interactions obtained from our integration method with results of Hasselmann & Hasselmann (1981). Parameters for all comparative spectra described by (3.1)–(3.4) are given in Table 1. The integration resolution in our computation was selected to be comparable to Hasselmann & Hasselmann (1981) in the spectral peak region. Consequently, any apparent differences in jaggedness in figures 2(*a*)–2(*d*) cannot be attributed to differences in grid resolution in this region of the spectrum.

Since we have not made any simplifying assumptions, our integration accuracy is limited only by the resolution of the integration grid. In figure 3 we compute the nonlinear transfer for the Pierson–Moskowitz spectrum considered in figure 2(*c*) using integration grids $(\imath, \jmath, \ell)$, where $\imath$ is the number of wavenumber bins, $2\jmath$ is the number of angular bins from −120° to +120° (unless otherwise specified) and ℓ is the number of points on the locus specified in (2.5). Comparing integration grids (115, 60, 70), (78, 30, 50), (48, 20, 30) and (29, 10, 10), we find that (48, 20, 30) and (78, 30, 50) give results that are very close to those of (115, 60, 70). Not shown is the integration with 115 wavenumber bins, 360 angular bins from −180° to 180° (1° discretization) and 70 points on the locus, which is essentially the same as (115, 60, 70). The effect of lower resolution on our integration method is a slightly less precise representation of the nonlinear transfer. All results are smooth.

The action flux formulation (2.12) implies that the rate of change of energy due to

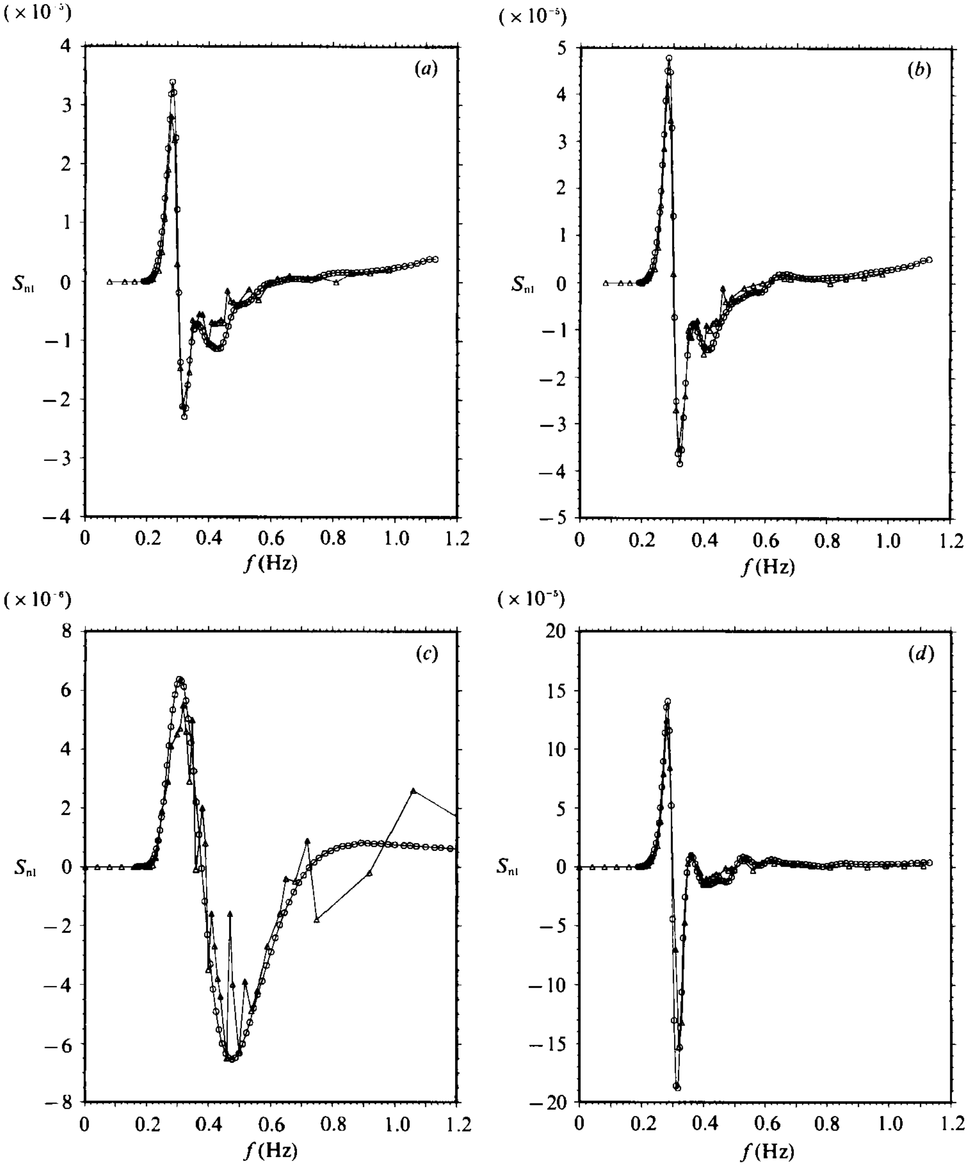


FIGURE 2. (*a*) Nonlinear transfer due to wave–wave interactions obtained from the integration method of this paper: ○, compared to Hasselmann & Hasselmann (1981); △, for case 2 in the latter study with $\cos^2\theta$ angular spreading. (*b*) As in (*a*) for Hasselmann & Hasselmann (1981) case 3 with angular spreading $\cos^4\theta$. (*c*) As in (*a*) for Hasselmann & Hasselmann (1981) case 13 corresponding to Pierson–Moskowitz spectrum. (*d*) As in (*a*) for Hasselmann & Hasselmann (1981) case 15 with $\gamma = 7$.

nonlinear transfer $\mathscr{S}_{\mathrm{nl}}(f)$ may be written as the one-dimensional divergence of energy flux,

$$\mathscr{S}_{\mathrm{nl}}(f) = \frac{\partial[\Gamma_{\mathrm{E}}^{+}(f) + \Gamma_{\mathrm{E}}^{-}(f)]}{\partial f}, \qquad (3.5)$$

where $\Gamma_{\mathrm{E}}^{+}(f)$ is the energy flux past f from low to high frequencies, and $\Gamma_{\mathrm{E}}^{-}(f)$ is the energy flux past f from high to low frequencies. Numerically, we may compare this

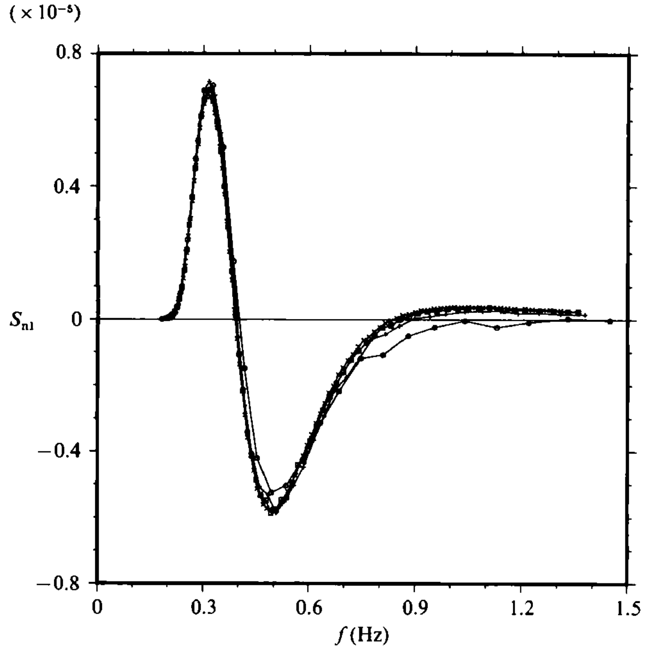


FIGURE 3. As in figure 2(c) for Pierson–Moskowitz spectrum, comparison the nonlinear transfer for integration grids (i, j, k), using ×, (115, 60, 70); □, (78, 30, 50); +, (48, 20, 30); ○, (29, 10, 10), where i is the number of wavenumber bins, $2j$ is the number of angular bins from −120° to +120° and k is the number of points on the locus.

Case†	Spreading function	Peakedness
2	$\frac{2}{\pi}\cos^2\theta$	3.3
3	$\frac{8}{3\pi}\cos^4\theta$	3.3
13	$\frac{2}{\pi}\cos^2\theta$	1.0
15	$\frac{2}{\pi}\cos^2\theta$	7.0

† From Hasselmann & Hasselmann (1981).

TABLE 1. Parameters for comparative spectra

flux divergence with the Boltzmann integral (2.3) for nonlinear transfer. Using the radial polar geometry of figure 1, we estimate the flux divergence centred between radial grid points where nonlinear transfer source term estimates are made,

$$\frac{\partial E(f)_{m+\frac{1}{2}}}{\partial t} = \frac{[\Gamma_E^+(f)_m - \Gamma_E^+(f)_{m+1}] + [\Gamma_E^-(f)_{m+1} - \Gamma_E^-(f)_m]}{f_{m+1} - f_m}. \tag{3.6}$$

Consequently, in regions of rapid nonlinear variations, the two calculations deviate

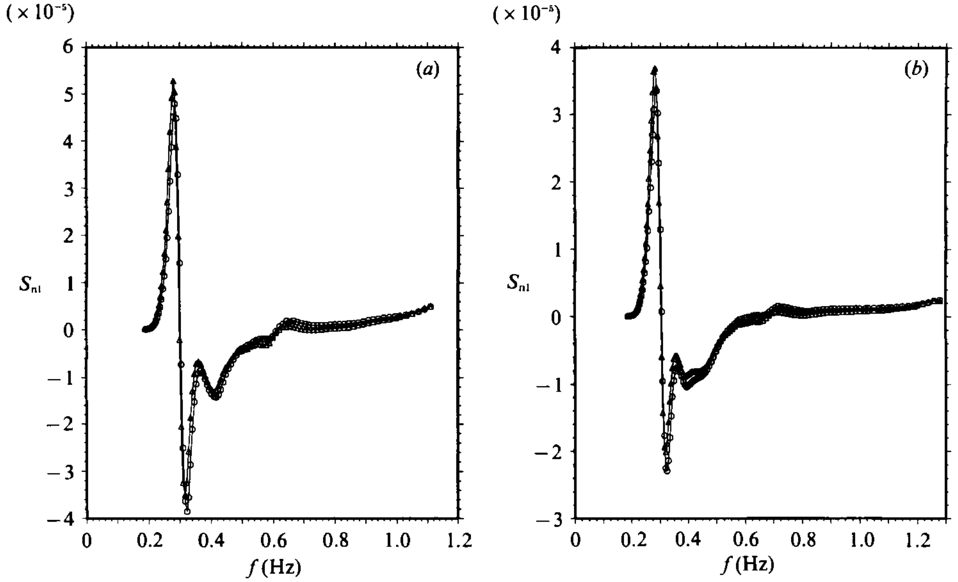


FIGURE 4. (*a*) A comparison between △, flux divergence, calculated by (3.5), and ○, the Boltzmann integral of (2.3), for Hasselmann & Hasselmann (1981) case 2. (*b*) As in (*a*) for Hasselmann & Hasselmann (1981) case 3.

slightly. As seen in figures 4(*a*) and 4(*b*), flux divergence estimates are a good approximation to estimates obtained directly from the Boltzmann integral. This confirms that the integration method for energy fluxes is properly posed and that the numerical technique has adequate detail.

Two-dimensional computations of $\partial n(\boldsymbol{k})/\partial t$, comparable with Webb (1978), are shown in figure 5. Two-dimensional energy fluxes through the wave spectrum can also be computed at each grid point and may be represented in terms of a flux density. For example, the action flux density into an element of phase space centred on $\boldsymbol{k}_1$ is

$$\Gamma_d(\boldsymbol{k}_1) = \int T(\boldsymbol{k}_1, \boldsymbol{k}_3)\, \hat{\boldsymbol{k}}_{31}\, \mathrm{d}\boldsymbol{k}_3, \tag{3.7}$$

expressed in terms of a unit vector $\hat{\boldsymbol{k}}_{31}$ in the direction $\boldsymbol{k}_1 - \boldsymbol{k}_3$ and the transfer integral $T(\boldsymbol{k}_1, \boldsymbol{k}_3)$ defined in (2.4). The action flux in the positive x-direction into an element of phase space $\mathrm{d}\boldsymbol{k}_1$ centred on $\boldsymbol{k}_1$ is therefore

$$\Gamma_x^+(\boldsymbol{k}_1) = \mathrm{d}\boldsymbol{k}_1 \int T(\boldsymbol{k}_1, \boldsymbol{k}_3) \cos\theta_{31}\, \mathrm{d}\boldsymbol{k}_3, \tag{3.8}$$

and so forth for other components. Only contributions to the integral are allowed for which $T(\boldsymbol{k}_1, \boldsymbol{k}_3)\cos\theta_{31}$ is positive, where $\theta_{13} = \arctan[(k_{x_1} - k_{x_3})/(k_{y_1} - k_{y_3})]$. The usual Green's relation relates flux divergence to the energy change due to nonlinear wave–wave transfer. We plot the action flux density vectors in figure 6. These are for the high-frequency region of the spectrum considered in figure 2(*a*) using a 'moderate' resolution grid (50, 31, 46): 50 frequency bins, 4° discretization within the angular domain (−120°, +120°) and 46 points on the locus-resonance condition (2.5). It is evident that there is little nonlinear transfer outside the angular domain (−120°, +120°) and no flux across the x-axis.

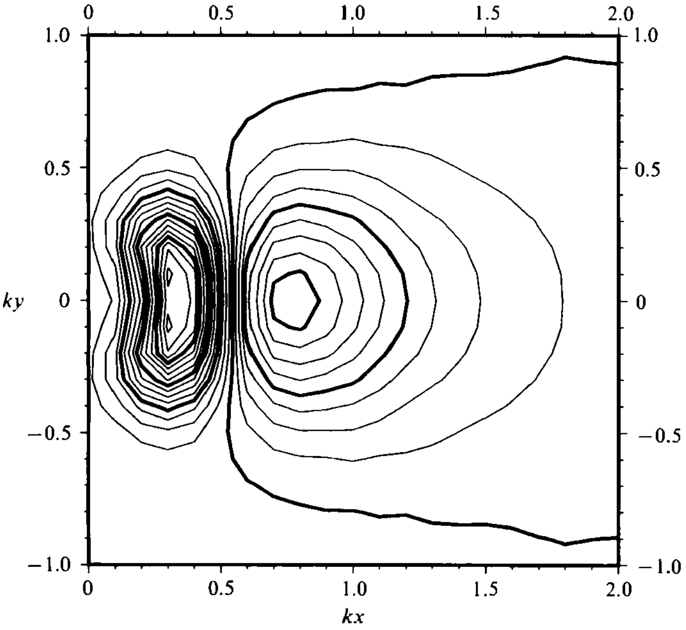


FIGURE 5. Two-dimensional nonlinear transfer $\partial n(\boldsymbol{k})/\partial t$ as a function of wavenumber, as in Webb (1978).

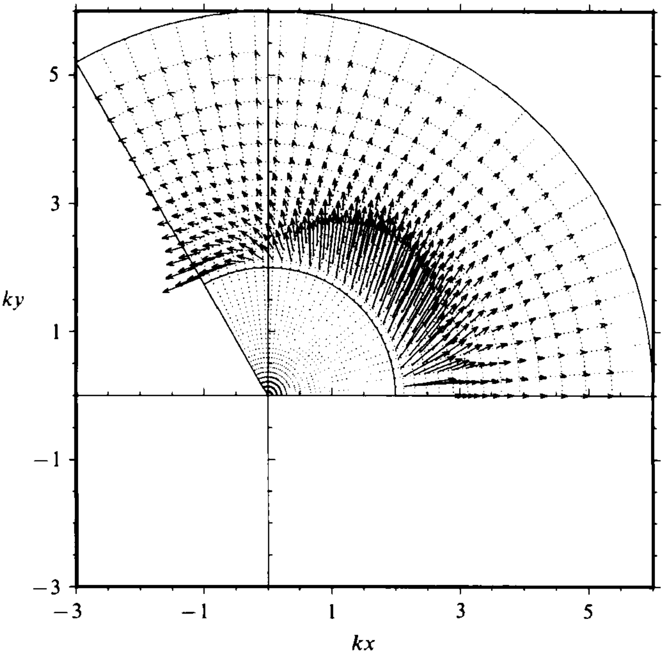


FIGURE 6. The two-dimensional nonlinear energy fluxes through the spectrum in the high-frequency region of the spectrum in figure 2(*a*). The magnification factor is 1.4×10^7.

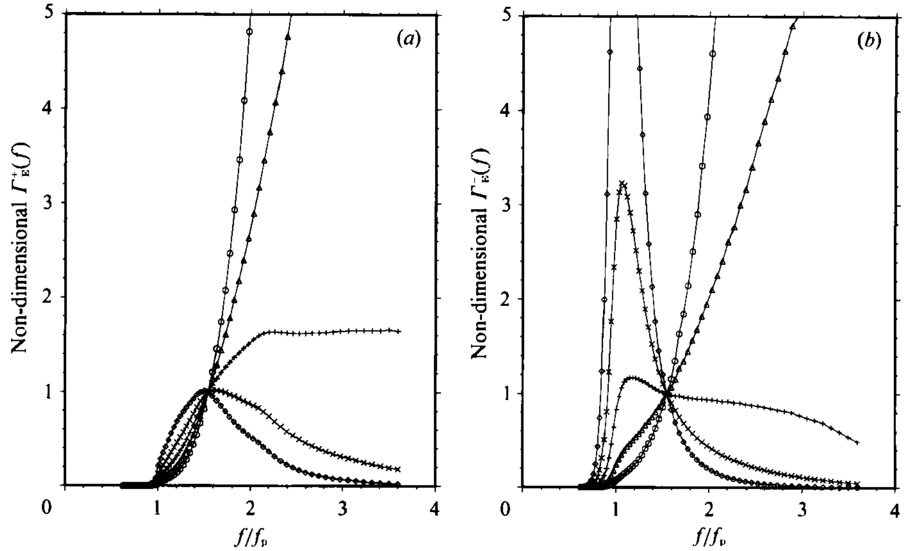


FIGURE 7. (*a*) Variations of energy flux to high frequencies Γ^+_E with different equilibrium range power laws normalized by the value each has at $1.6 f_p$: ○, for f^{-2}, △, for f^{-3}; +, for f^{-4}; ×, for f^{-5}; ◇, for f^{-6}. Other parameters are $f_p = 0.3$, $\gamma = 1.214$, $\alpha_4 u = 0.01$, $\sigma_a = 0.07$, $\sigma_b = 0.09$ and directional spreading is $\cos^2\theta$. (*b*) Variations of energy flux to low frequencies Γ^-_E as a function of equilibrium range power law as in (*a*).

4. Basic scaling behaviour of nonlinear fluxes

4.1. *Flux dependence on power laws in the equilibrium range*

To examine the energy flux behaviour for different equilibrium range power laws, we consider simple specta of the form

$$E(f,\theta) = A\alpha_4 u g f^{-m}\, \psi\left(\frac{f}{f_p}\right)\cos^2\theta, \tag{4.1}$$

where m is a positive integer, $\alpha_4 u$ is a dimensional constant with units of length/time, A is a directional normalization constant satisfying

$$\int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} A\cos^2\theta\, d\theta = 1 \tag{4.2}$$

and ψ is a non-dimensional shape function specified by the usual JONSWAP-type parameters (3.1)–(3.4), and prescribed in subsequent sections. Experimental evidence suggests that the equilibrium range exists in a subrange of the spectrum from approximately $1.6 f_p$ to $2.6 f_p$. To accentuate the divergence aspects of energy fluxes, we normalize all fluxes for a given power law by its value at the low-frequency limit of the equilibrium range (i.e. at about $1.6\, f_p$). Figures 7(*a*) and 7(*b*) show the behaviour of these normalized fluxes to high frequencies $\Gamma^+_E(k)$ and low frequencies $\Gamma^-_E(k)$ for various power laws. Unless otherwise specified, parameters are $f_p = 0.3$, $\gamma = 1.214$, $\alpha_4 u = 0.01$, $\sigma_a = 0.07$, $\sigma_b = 0.09$ and directional spreading is $\cos^2\theta$ in this and subsequent sections. For a constant energy flux through the equilibrium range of the spectrum, the normalized fluxes should retain a value of approximately 1. However, as can be seen in figure 7, the fluxes are approximately constant only for an f^{-4} spectrum. Spectra with f^{-2} and f^{-3} equilibrium ranges have fluxes which

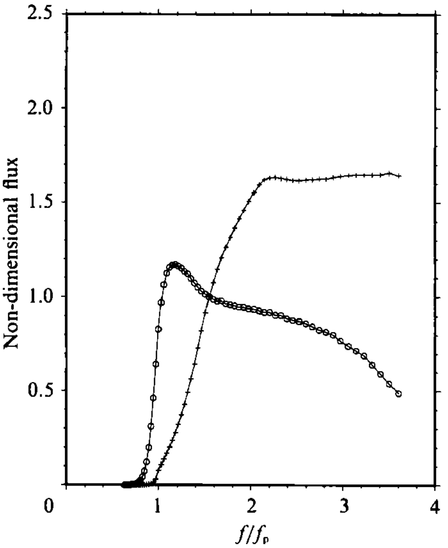


FIGURE 8. Variations of energy flux to high frequencies Γ_E^+, and to low frequencies Γ_E^-, as a function of different peak frequencies: 0.2, 0.3, 0.4 Hz.

increase with increasing frequency. Spectra with f^{-5} and f^{-6} equilibrium ranges have fluxes which decrease with increasing frequency.

The calculations shown here provide additional support for the existence of an f^{-4} equilibrium range in wave spectra. For the remainder of this study, we restrict ourselves to analyses of spectra with f^{-4} equilibrium ranges. The equation for this class of spectra is

$$E(f,\theta) = \Lambda\alpha_4 ugf^{-4}\psi\left(\frac{f}{f_p}\right)\cos^2\theta. \tag{4.3}$$

As discussed in Resio & Perrie (1989), a judicious selection of parameters makes such a form equivalent to the JONSWAP spectral form.

4.2. *Flux dependence on variations in peak frequency*

To investigate the behaviour of energy fluxes through the equilibrium range as a function of peak frequency, we consider spectra of the form (4.3) with ψ given by

$$\psi\left(\frac{f}{f_p}\right) = \begin{cases} \left(\frac{f}{f_p}\right)^4 \exp\left(1-\left(\frac{f_p}{f}\right)^4\right) & \text{for} f \leqslant f_p, \\ 1 & \text{for} f \geqslant f_p. \end{cases} \tag{4.4}$$

Integrations for peak frequencies f_p of 0.2, 0.3 and 0.4 show that energy fluxes are identical when plotted as a function of f/f_p. Figure 8, which appears to contain only two curves, actually contains plots for the fluxes to both high and low frequencies through all three spectra.

This may seem surprising at first, since Hasselmann *et al.* (1973) established that nonlinear transfer due to wave–wave interactions is dependent on f_p. However, the frequency axis in figure 8 is scaled by f_p. The nonlinear transfer source function for these spectra has the form

$$\mathscr{S}_{nl} \sim f_p^{-1}, \tag{4.5}$$

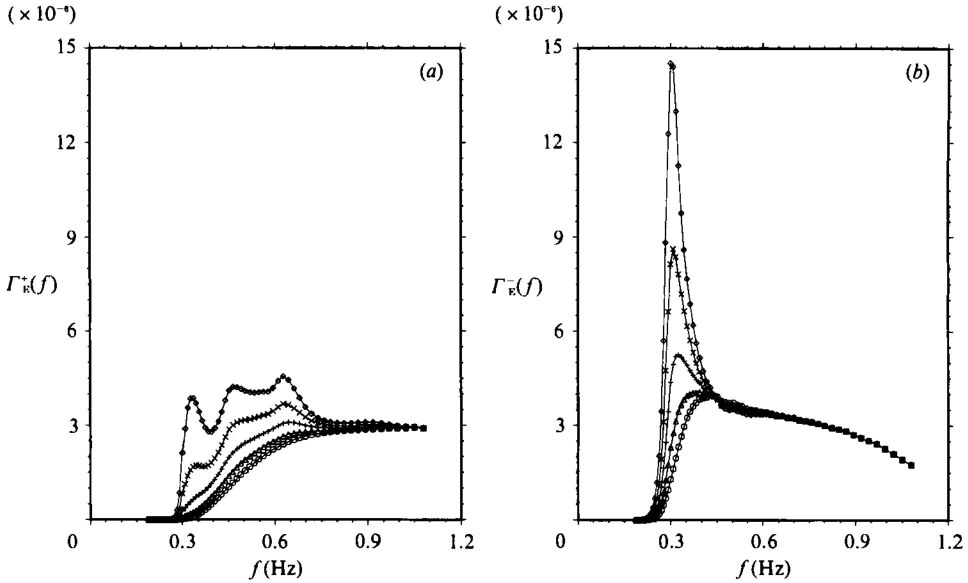


FIGURE 9. (*a*) Variations of energy flux to high frequencies Γ^+_E with peakedness: ○, for $\gamma = 0.65$; △, for $\gamma = 1.0$; +, for $\gamma = 1.55$; ×, for $\gamma = 2.30$; ◇, for $\gamma = 3.25$. (*b*) Variations of energy flux to low frequencies Γ^-_E with peakedness as in (*a*).

whereas for an f^{-5} spectrum

$$\mathcal{S}_{\mathrm{nl}} \sim f_{\mathrm{p}}^{-4}. \tag{4.6}$$

4.3. *The influence of the spectral peak on fluxes*

In this section we consider the influence of the sharp energy cutoff below the spectral peak as well as the manner in which variations in spectral peakedness affect energy fluxes through the spectrum. Our spectra are of the form (4.2), with ψ given by

$$\psi\left(\frac{f}{f_{\mathrm{p}}}\right) = \begin{cases} \left(\dfrac{f}{f_{\mathrm{p}}}\right)^4 \exp\left(1-\left(\dfrac{f_{\mathrm{p}}}{f}\right)^4\right)\gamma^{\Xi} & \text{for} f \leqslant f_{\mathrm{p}}, \\ \gamma^{\Xi} & \text{for} f > f_{\mathrm{p}}, \end{cases} \tag{4.7}$$

where

$$\Xi = \exp -\frac{1}{2}\left(\frac{f-f_{\mathrm{p}}}{\sigma f_{\mathrm{p}}}\right)^2, \tag{4.8}$$

and γ is a non-dimensional peakedness parameter.

The variation in energy fluxes for $\gamma = 0.65$, 1.0, 1.55, 2.30 and 3.25 is shown in figure 9. This range of γ corresponds to the measurements of Donelan, Hamilton & Hui (1985). Figure 10 show the corresponding nonlinear energy transfer for these spectra. A notable feature is the shift of the positive lobe toward higher frequencies as γ decreases, particularly when γ becomes less than 1. The associated transfers of action and momentum to the forward face also vary markedly as a function of γ. This is an important mechanism in controlling wave growth and possibly the evolution into a fully-developed spectral form.

Approximations by Kitaigorodskii (1983) and Resio (1987) concerning equilibrium range fluxes neglected consideration of the extent to which energy transfers are influenced by the spectral peak and the associated cutoff energy on the forward face

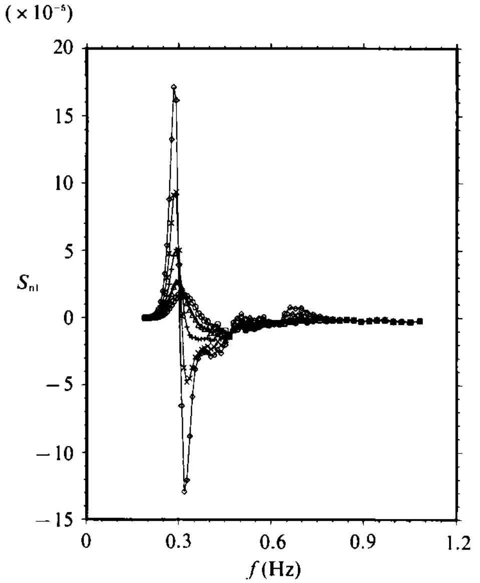


FIGURE 10. As in figure 9 for corresponding nonlinear transfer.

of the spectrum. Experimental results suggest large departures from equilibrium range for frequencies less than $1.6f_p$ (figure 15 in Donelan *et al.* 1985). To examine the influence of the spectral peak, we consider normalized energy fluxes through the spectrum as a function of non-dimensional frequency f/f_p. The normalization is that proposed for equilibrium range fluxes by Resio (1987) where it was shown that energy fluxes in the equilibrium range should vary as

$$\Gamma_E(k) \approx \epsilon \frac{g^2F^3(k)\,k^9}{\omega^3} \tag{4.9}$$

letting $F(k)$ be the one-dimensional energy density in wavenumber space and ϵ, a non-dimensional constant. We therefore define normalized energy fluxes to be

$$\hat{\Gamma}_E(k) = \frac{\Gamma_E(k)\,\omega}{g^2F^3(k)\,k^9} \tag{4.10}$$

Figure 11 shows the behaviour of $\hat{\Gamma}^{\pm}_E(k)$ as a function of f/f_p for different values of γ. From figure 11(*a*), the normalized flux to high frequencies $\hat{\Gamma}^{+}_E(k)$ attains equilibrium range values that are essentially independent of γ by $f/f_p \approx 2.3$. Figure 11(*b*) implies that the normalized flux to low frequencies $\hat{\Gamma}^{-}_E(k)$ attains equilibrium range values that are essentially γ independent by $f/f_p \approx 1.5$.

Finally, we consider the magnitude of nonlinear energy fluxes and their associated source terms as a function of spectral peakedness. Hasselmann *et al.* (1973) suggest that as the peakedness of the spectrum increases, the source functions become larger. While this may be true, the mechanism behind it may not be the 'sharpness' of the spectrum. It may be due to the fact that, in the spectral parameterization used in this study and by Hasselmann *et al.* (1973), the absolute magnitude of the energy densities in the region of the peak increases with increasing γ. As these energy densities increase, their contributions to the nonlinear fluxes increase by $F^3(k)$ and the divergences of the fluxes also increase.

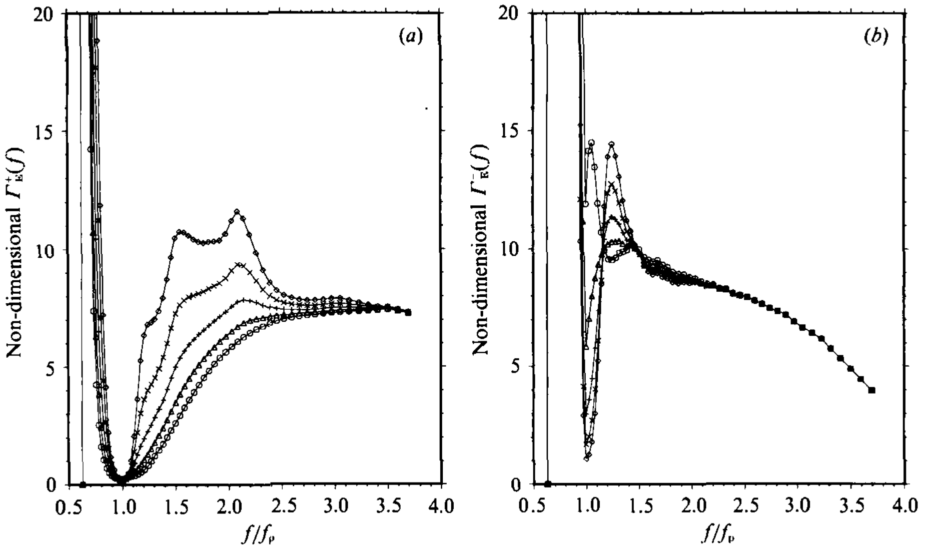


FIGURE 11. (*a*) Variations in non-dimensional flux to high frequencies $\hat{\Gamma}^+_E(k)$ with respect to peakedness as in figure 9, plotted against f/f_p. $\hat{\Gamma}^+_E(k)$ is non-dimensionalized according to (4.10). (*b*) Variations in non-dimensional flux to low frequencies $\hat{\Gamma}^-_E(k)$ with respect to peakedness as in (*a*).

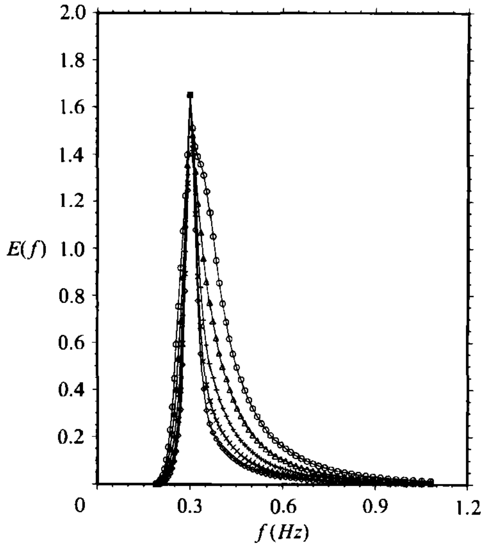


FIGURE 12. Variation of spectra with respect to peakedness as in figure 9. All spectra are normalized to have the same maximum.

A related concern is therefore whether or not source terms and fluxes become larger for different peakedness values, given the same energy density at the spectral peak. To answer this, we performed integrations for spectra defined by (4.3), (4.7)–(4.8) and normalized to have the same spectral energy at the peak. Figure 12 shows the spectral shapes for various γ generated in this manner. The fluxes and associated nonlinear energy transfer terms are shown in figures 13 and 14

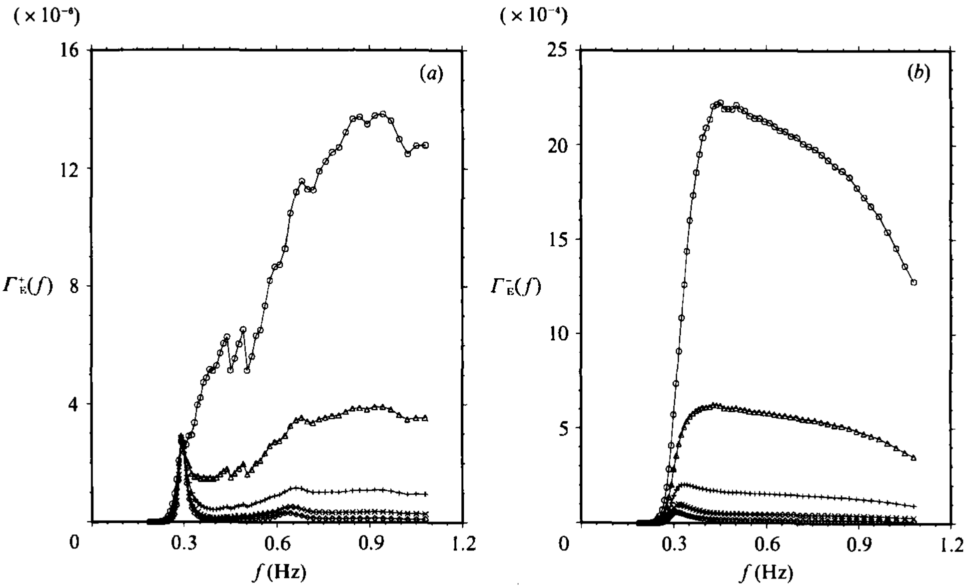


FIGURE 13. (*a*) Energy fluxes to high frequencies $\Gamma^{+}_{\mathrm{E}}(k)$ corresponding to spectra of figure 12 as a function of peakedness as in figure 9. (*b*) Energy fluxes to low frequencies $\Gamma^{-}_{\mathrm{E}}(k)$ as a function of peakedness as in (*a*).

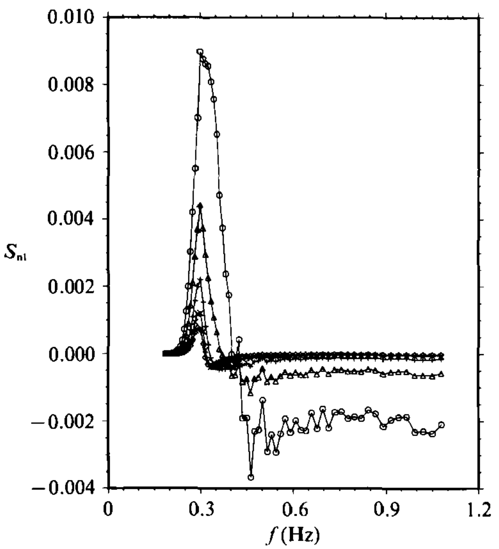


FIGURE 14. Nonlinear transfer corresponding to the spectra of figure 12.

respectively. It is evident that, with the same energy density at the spectral peak, the magnitude of both the fluxes and the source terms are larger for broader spectra than they are for more sharply peaked spectra.

4.4. *Effects of angular distribution of energy on energy fluxes*

The parameterization of spectral peakedness presented above gives an indication of the manner in which the gradient of energy density as a function of frequency f, can affect energy fluxes through a spectrum. For a given frequency, the gradient of energy density with angle θ can also affect energy fluxes through a spectrum. Thus

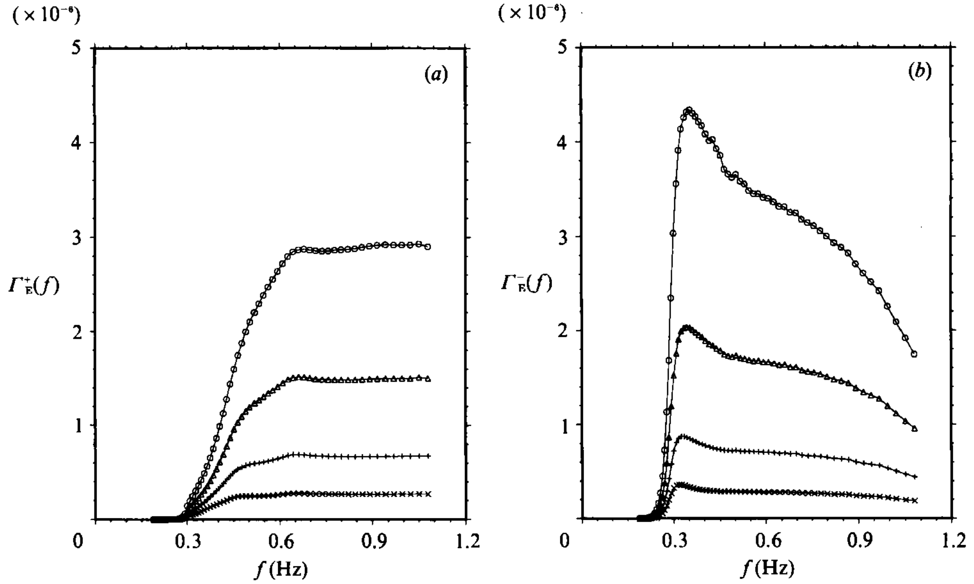


FIGURE 15. (*a*) Energy fluxes to high frequencies $\Gamma_E^+(k)$ as a function of angular spreading $\cos^{2n}\theta$: ×, for $n = 1$; +, for $n = 2$; △, for $n = 4$; ○, for $n = 8$. (*b*) Energy fluxes to low frequencies $\Gamma_E^-(k)$ as a function of angular spreading as in (*a*).

far, we have only considered spectra with a $\cos^2\theta$ angular distribution. Although it may be instructive to analyse spectra with f- and θ-dependent spreading functions as presented by Mitsuyasu *et al.* (1975), Hasselmann, Dunkel & Ewing (1980) and Donelan *et al.* (1985), for simplicity we examine only spreading functions which are independent of f.

Using the spectrum (4.3) and (4.7)–(5.8) with $\alpha_4 u = 0.01$, $f_p = 0.3$, $\gamma = 1.214$, $\sigma_a = 0.07$ and $\sigma_b = 0.09$, integrations were made with normalized $\cos^{2n}\theta$ spreading functions, letting n equal 1, 2, 4 and 8. As shown in figure 15, energy fluxes are dependent on the angular spreading function and increase as n increases. The dependency is not as strong as in the case of spectral peakedness although the range of variation was taken to cover the range that could be expected to occur in nature. The equilibrium range first occurs at $f/f_p \approx 2.0$ for the flux to high frequencies $\Gamma_E^+(k)$ and somewhat earlier at $f/f_p \approx 1.5$ for the flux to low frequencies $\Gamma_E^-(k)$.

4.5. *Variations in nonlinear fluxes due to the equilibrium range coefficient*

It is apparent from the algebraic structure of the density function $\mathscr{D}$ as shown in (2.2), that any multiplicative factor introduced into a spectral density manifests itself as the cube of that factor in $\mathscr{D}$. As this is used later, we numerically demonstrate this. Figure 16 shows the calculated energy fluxes for a reference f^{-4} spectrum with parameter $\alpha_4 u = 0.01, f_p = 0.3, \gamma = 1.21, \sigma_a = 0.07$ and $\sigma_b = 0.09$. Energy fluxes are also presented for spectra with identical parameters except that $\alpha_4 u = 0.03$. Dividing the fluxes of the second computation by 27 makes the two curves exactly the same.

4.6. *Numerical evaluation of the non-dimensional flux coefficients*

Figure 9 showed that the flux to high frequencies $\Gamma_E^+(k)$ attains equilibrium range values that are independent of peakedness γ by $f/f_p \approx 2.3$ and the flux to low frequencies $\Gamma_E^-(k)$ attains equilibrium range values that are γ independent by $f/f_p \approx$

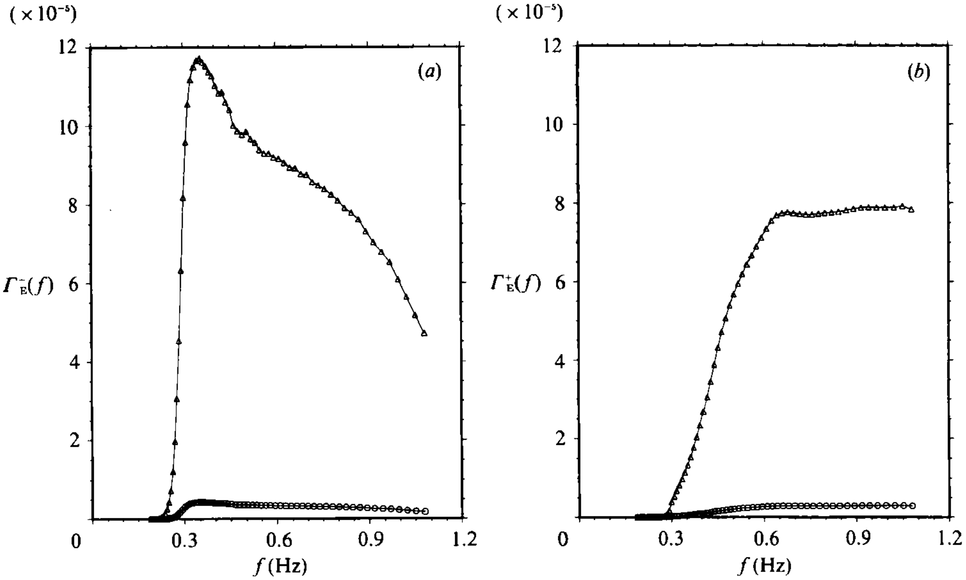


FIGURE 16. (*a*) Variation of energy fluxes to high frequencies $\Gamma_E^+(k)$ with $\alpha_4 u$: ○, for $\alpha_4 u = 0.01$; △, for $\alpha_4 u = 0.03$. The plots differ by a factor of 27. (*b*) Variation of energy fluxes to low frequencies $\Gamma_E^-(k)$ as a function of $\alpha_4 u$ as in (*a*).

1.5. Figure 8 shows that these fluxes are independent of the location of f_p. Therefore, for a given angular spreading function, whether or not frequency dependent, energy fluxes in these equilibrium ranges are only dependent on the equilibrium range coefficient. This supports the estimates for energy fluxes through the spectrum made by Kitaigorodskii (1983) and Resio (1987), based on arguments that the fluxes should approximately balance energy input by wind.

Energy flux estimates (4.9) expressed in terms of frequency, may be represented as

$$\Gamma_E^{\pm}(f) = \frac{\epsilon^{\pm}(2\pi)^9 f^{12} E^3(f)}{8g^4}\phi^{\pm}\left(\frac{f}{f_p}\right) \tag{4.11}$$

where ' + ' refers to fluxes from low to high frequencies, ' — ' refers to fluxes from high to low frequencies, $\phi^{\pm}$ is a non-dimensional shape function and $\epsilon^{\pm}$ is a non-dimensional constant. As $\phi^{\pm} = 1$ in the equilibrium range, $\epsilon^{\pm}$ has the same meaning as ϵ in (4.9).

Substituting the appropriate equilibrium form for $E(f)$ from (4.3) into (4.11) yields,

$$\Gamma_E^{\pm}(f) = \epsilon^{\pm}\frac{\alpha_4^3 u^3 (2\pi)^9 A^3}{8g} \tag{4.12}$$

which, as shown in figure 9, represent good approximations for all spectra of the form (4.3). Since $\alpha_4 u$ is a known quantity in our integrations, we can explicitly evaluate $\epsilon^{\pm}$. We find

$$\epsilon^{\pm} \approx 60 \tag{4.13}$$

making the evaluation at about the midpoint of the equilibrium range. This is consistent with the earlier estimates of Resio (1987).

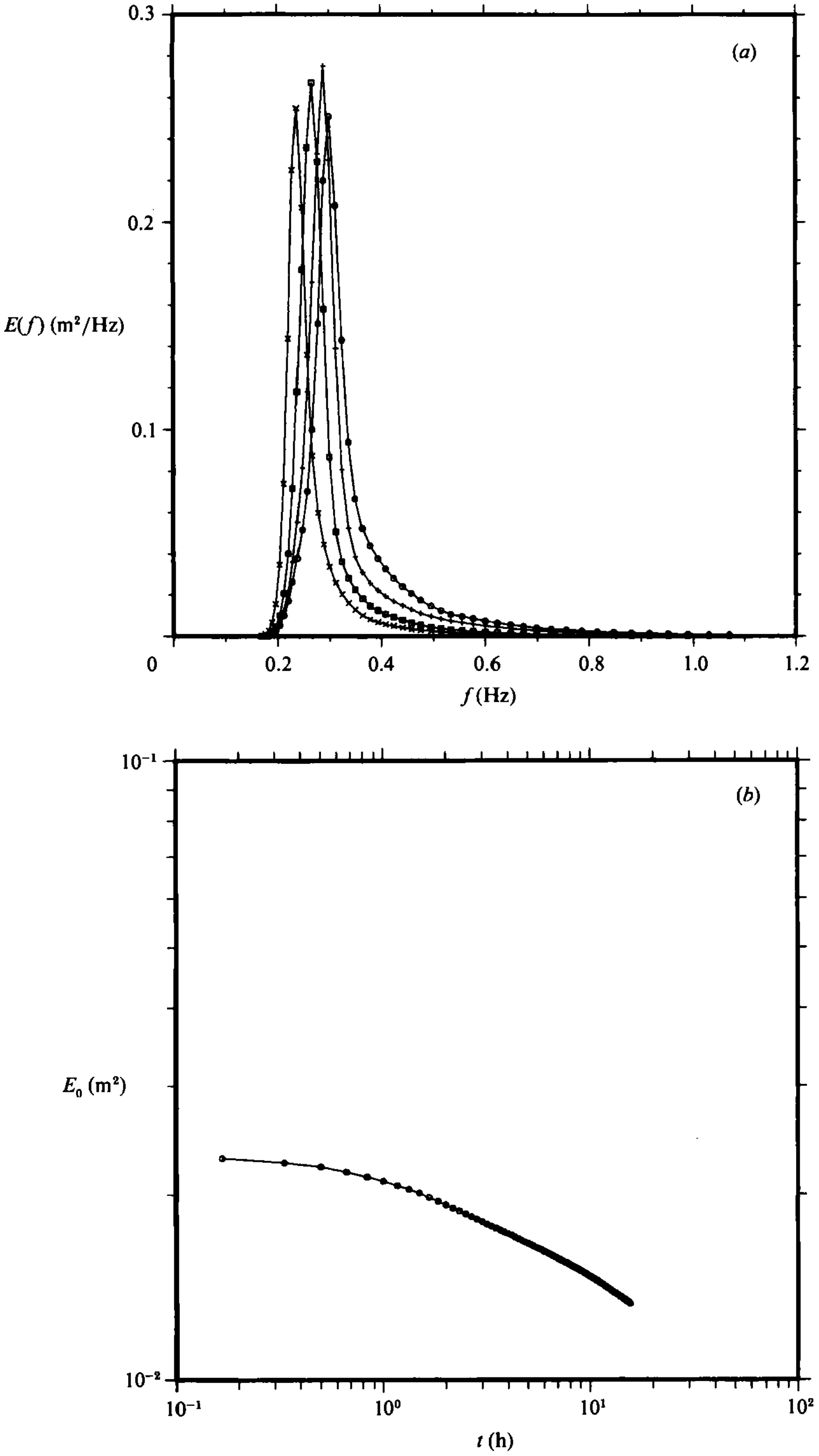


FIGURE 17. For caption see next page.

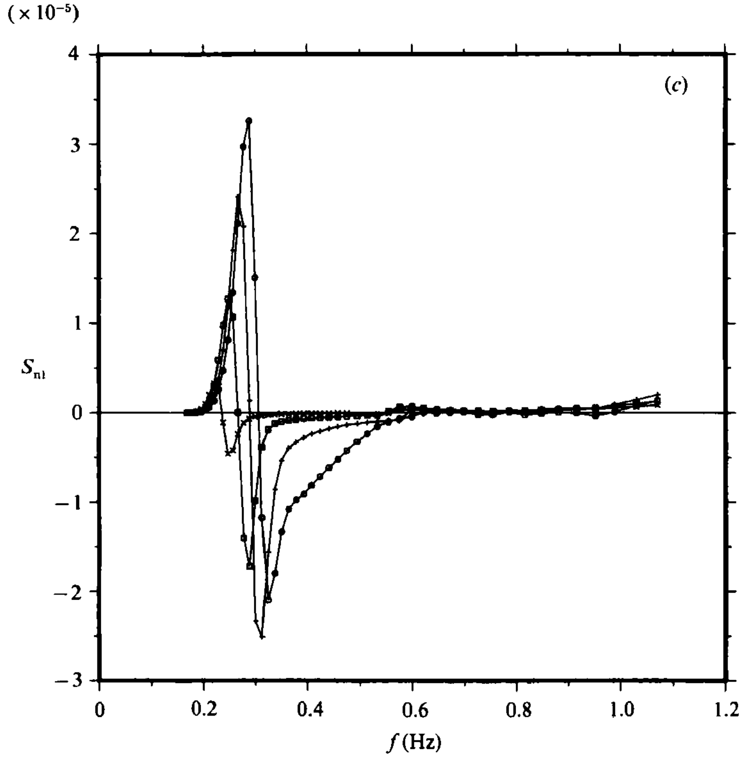


FIGURE 17. (*a*) The time evolution of one-dimensional energy $E(f)$ in the absence of $\mathscr{S}_{\text{in}}$ and $\mathscr{S}_{\text{ds}}$. The integration grid is (48, 20, 30) with 15 s timesteps. ○, 15 s; +, 1 h; □, 3 h; ×, 11 h. (*b*) As in (*a*), the total energy E_0 variation with time. (*c*) As in (*a*), the variation of the nonlinear energy transfer $\mathscr{S}_{\text{nl}}$ with time.

5. Spectral evolution in time

5.1. *Evolution of energy and nonlinear transfer in time*

With respect to space and time, the spectral energy density $E(f,\theta)$ from (4.3) in deep water evolves as

$$\frac{\partial E}{\partial t}(f,\theta)+\boldsymbol{C}_{\text{g}}\cdot\nabla E(f,\theta)=\mathscr{S}_{\text{in}}+\mathscr{S}_{\text{nl}}+\mathscr{S}_{\text{ds}} \tag{5.1}$$

where $\mathscr{S}_{\text{in}}$ is the wind input spectral energy, $\mathscr{S}_{\text{nl}}$ is the nonlinear transfer due to wave–wave interactions and $\mathscr{S}_{\text{ds}}$ is the wave breaking dissipation. We compute the time evolution of one-dimensional energy $E(f)$, total energy E_0 and nonlinear transfer $\mathscr{S}_{\text{nl}}$ in figures 17(*a*), 17(*b*) and 17(*c*), in the absence of $\mathscr{S}_{\text{in}}$ and $\mathscr{S}_{\text{ds}}$. Our integration grid is (48, 20, 30), with 30 s timesteps and an initial spectrum as shown in figure 17(*a*). We model the spectrum above $2f_{\text{p}}$ with an f^{-4} tail to reduce computer requirements.

The time progression of the spectrum is presented in figure 17(*a*). Influenced only by wave–wave interactions, the spectrum initially steepens and the peak migrates to lower frequencies. However, because the integration grid extends over only a finite range on the frequency axis, total energy slowly decreases as energy is fluxed to the high-frequency boundary of the grid and is lost to the next timestep of the integration. Thus, the peak of the spectrum decreases with time after about an hour. The time evolution of total energy E_0 in figures 17(*b*) is a further reflection of the loss

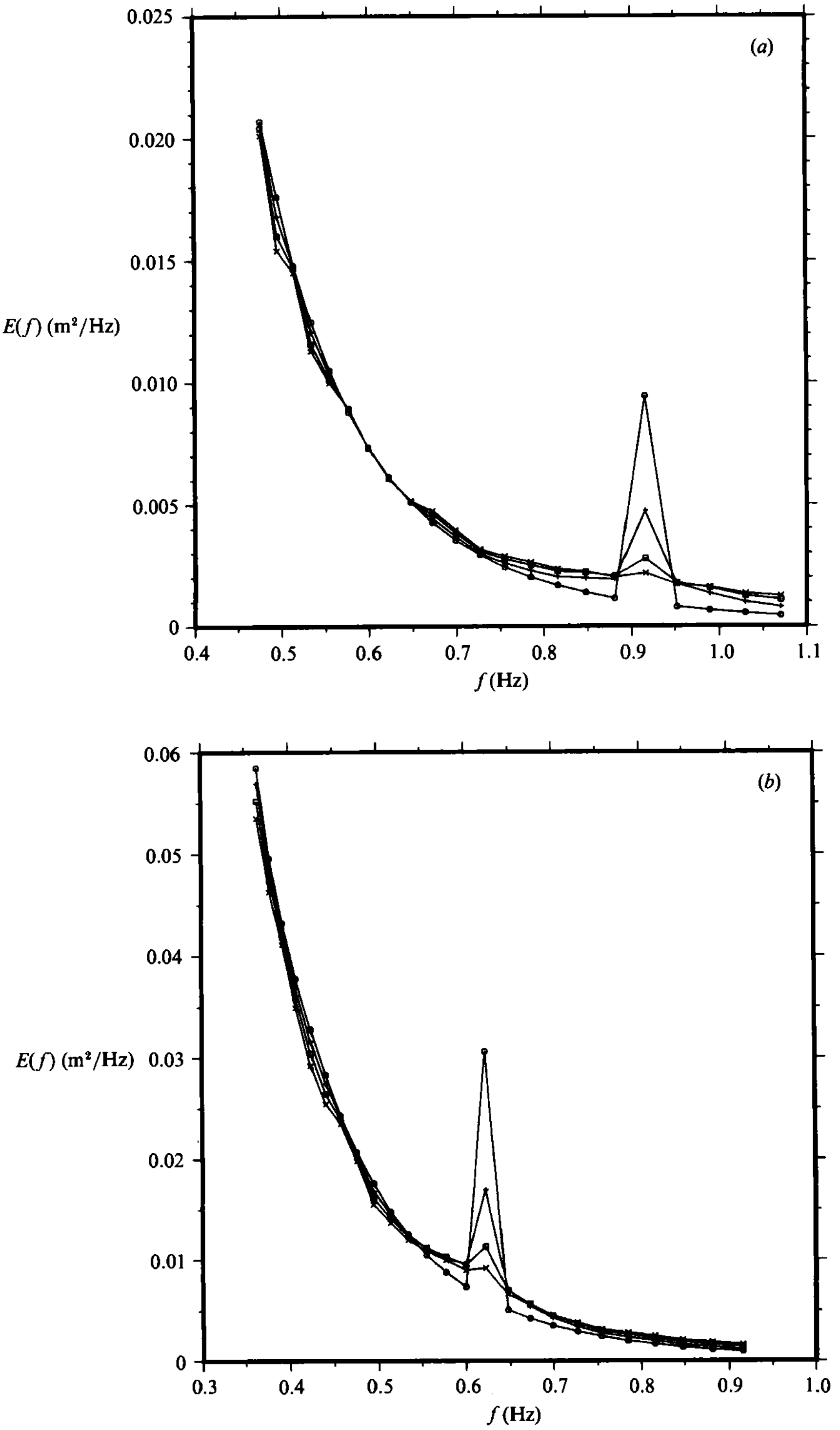
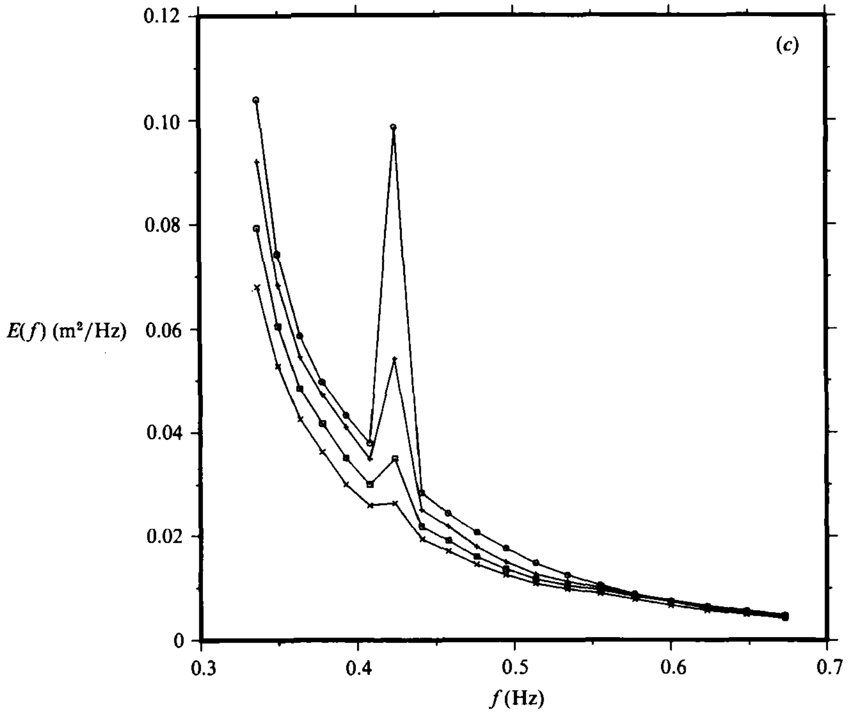


FIGURE 18. For caption see next page.



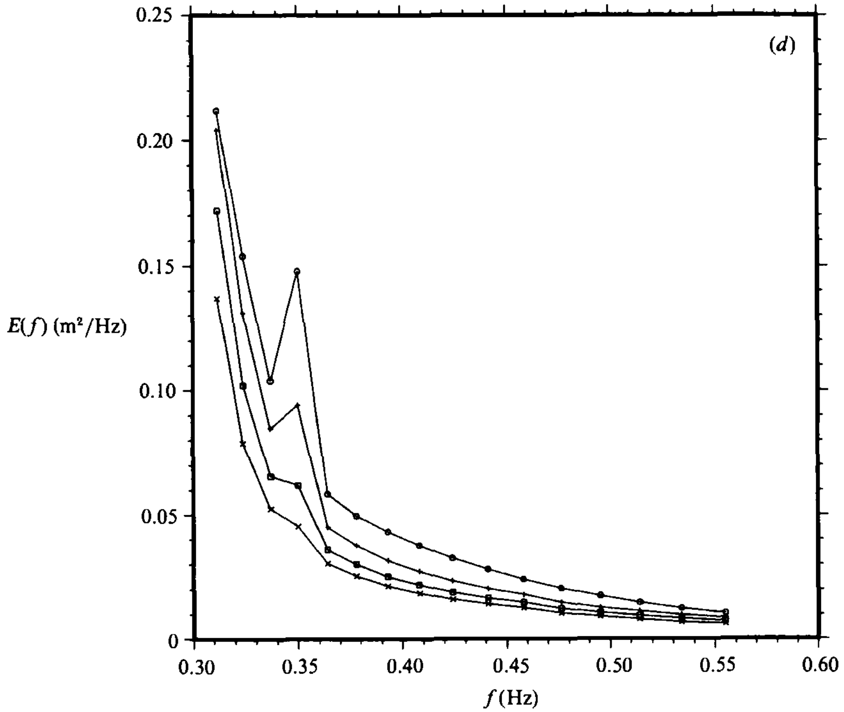


FIGURE 18. (*a*) Evolution of the energy spectrum in response to a 10 × perturbation at $3f_p$ where the spectrum is denoted by: ○, initially; +, after 1 min; □, after 2 min; ×, after 3 min. (*b*) As in (*a*) in response to a 5 × perturbation at $2f_p$ where the spectrum is denoted by: ○, initially; +, after 2 min; □, after 4 min; ×, after 6 min. (*c*) As in (*a*) in response to a 3 × perturbation at $1.5f_p$ where the spectrum is denoted by: ○, initially; +, after 10 min; □, after 20 min; ×, after 30 min. (*d*) As in (*a*) in response to a 2 × perturbation at $\frac{7}{8}f_p$ where the spectrum is denoted by: ○, initially; +, after 20 min; □, after 40 min; ×, after 60 min.

of energy due to finiteness of the grid. Finally, in figure 17 (*c*) we see the evolution of the corresponding nonlinear transfer $\mathcal{S}_{\mathrm{nl}}$. Owing to a diminished spectrum, $\mathcal{S}_{\mathrm{nl}}$ decreases drastically as time increases and migrates to lower frequencies following the migration of the spectrum.

5.2. *Response to perturbations in the spectrum*

It is important to investigate the response of the nonlinear energy transfer to perturbations within the spectrum in order to assess the rate at which a spectrum relaxes toward a quasi-stationary state. We compute the time evolution of one-dimensional energy $E(f)$ using a (48, 20, 30) grid with initial spectral parameters as in §5.1. In figure 18 (*a*), we introduce a 10 × perturbation at $3f_{\mathrm{p}}$. Using 15 s timesteps and integrating over the entire frequency domain (without using an f^{-4} tail as in §5.1), we see that within 3 min the nonlinear interactions have reduced the perturbation to less than 90 % of its original magnitude and distributed the energy among neighbouring spectral energy bins. Figure 18 (*b*) presents a 5 × perturbation at $2f_{\mathrm{p}}$. In this situation, the nonlinear interactions essentially remove the perturbation within 6 min. In figure 18 (*c*) a 3 × perturbation at $1.5f_{\mathrm{p}}$ is removed in 30 min. Finally in figure 18 (*d*) a 2 × perturbation at $\frac{7}{8}f_{\mathrm{p}}$ requires 1 h before nonlinear interactions have removed approximately 90 % of it.

From figure 3, our investigations show that the nonlinear energy transfer should be smooth, even for very coarse integration grids. We have demonstrated that nonlinear wave–wave interactions work to smooth perturbations introduced into the spectrum as spikes. Clearly the time taken for the nonlinear interactions to respond to any perturbation depends strongly on where it occurs within the spectrum.

5.3. *Relaxation times within the equilibrium range*

An alternate view of the spectral response to perturbations within the equilibrium range may be presented in terms of relaxation times within the equilibrium range. In a manner typical of many others. Kitaigorodskii (1983) estimated relaxation times for wave–wave interactions in terms of energy density and rate of change of energy, at a given frequency

$$\mathcal{T} = \frac{E(f)}{\dfrac{\partial E(f)}{\partial t}}. \tag{5.2}$$

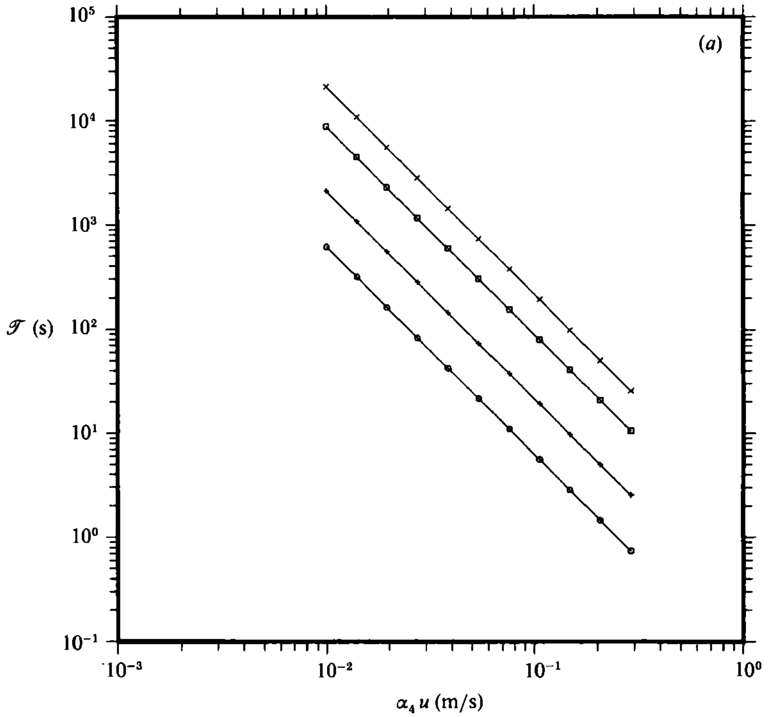
Unfortunately, the denominator of (5.2) becomes very small in the equilibrium range and the estimated relaxation time becomes arbitrarily large. Therefore we propose a different form based on energy fluxes

$$\mathcal{S} = \frac{\mathscr{E}}{\Gamma_{\mathrm{E}}^{+} + \Gamma_{\mathrm{E}}^{-}}, \tag{5.3}$$

where $\mathscr{E}$ is the total energy in the region of the spectrum being considered,

$$\mathscr{E} = \int_{f_1}^{f_2} E(f)\,\mathrm{d}f, \tag{5.4}$$

and f_1 and f_2 are the appropriate upper and lower frequency limits of the equilibrium range. This definition relates relaxation time to the time required for nonlinear energy fluxes to remove all the energy from a specific region of the spectrum. This is a physically consistent approach to estimating the relative strength of the



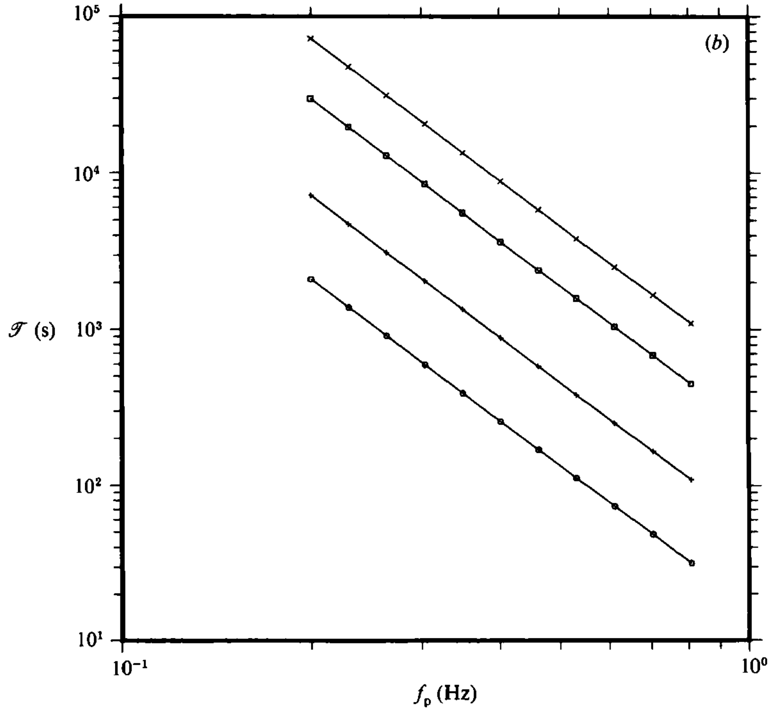


FIGURE 19. (a) Relaxation time defined by (5.3) as a function of $\alpha_4 u$: ○, $\gamma = 1$; +, $\gamma = 2$; □, $\gamma = 5$; ×, $\gamma = 10$. (b) As in (a) as a function of f_p.

nonlinear fluxes and is motivated by (3.5) relating the nonlinear transfer $\mathcal{S}_{\mathrm{nl}}$ to spectral fluxes.

To investigate the sensitivity of relaxation times defined by (5.3) with respect to peakedness γ, peak frequency f_p and $\alpha_4 u$, we combine (4.3), (4.12), and (5.3) and infer

$$\mathcal{T} = \frac{\mathcal{D} g^2 f_p^{-3}}{(\alpha_4 u)^2}, \tag{5.5}$$

where $\mathcal{D}$ is a non-dimensional constant depending on f_1, f_2 and spectral peakedness γ. Assuming an equilibrium range between $1.6 f_p$ and $2.5 f_p$, figure 19(*a*) shows the variation of $\mathcal{T}$ with $\alpha_4 u$ and γ, computed in the middle of the equilibrium range from (5.3). Similarly, figure 19(*b*) shows the variation of $\mathcal{T}$ with respect to f_p and γ. The agreement between (5.5) and figures 19(*a*) and 19(*b*) is remarkable. The exponents for f_p and $\alpha_4 u$ in (5.5) are obtained to 3 decimal places with correlation coefficients that are very near one. Relaxation times in these computations are seen to correspond to the results of the previous section. The mid-range abscissa in figure 19(*a*) is appropriate for wind speeds from 12 to 20 m s^{-1} with $f_p = 0.3$.

6. Conclusions

A different perspective for nonlinear energy transfer due to wave–wave interactions in a spectrum has been suggested. Formulating an efficient numerical integration scheme for the nonlinear energy transfers first described by Hasselmann (1961), we calculated energy fluxes through the spectrum and the spectral evolution with time. The divergence of these fluxes is the conventional 'source term' formulation used in past parameterizations of nonlinear transfer due to wave–wave interactions. Moreover, fluxes appear to provide an important understanding of overall energy exchanges among various regions of the spectrum. We make the following conclusions:

(i) As described by Zakharov & Filonenko (1968), Kitaigorodskii (1983) and Resio (1987), and seen in figure 7, energy fluxes through the equilibrium range of a spectrum are approximately constant only for an f^{-4} spectrum, at frequencies sufficiently above the spectral peak.

(ii) Fluxes through a spectrum are independent of f_p, as seen in figure 8; nonlinear energy transfers for an f^{-4} spectrum therefore scale as $\alpha^3 f_p^{-1}$ rather than $\alpha^3 f_p^{-4}$ as in the case of an f^{-5} spectrum.

(iii) Energy fluxes through the equilibrium range are independent of spectral peakedness and depend only on the local energy densities. This is evident in figure 9 and also the non-dimensional curves of figure 11. The lobes of the corresponding nonlinear energy transfers shift to higher frequencies with decreasing peakedness, particularly when peakedness is less than 1.0.

(iv) Decreasing peakedness while holding the energy of the peak constant leads to decreased sharpness, broader spectra, enhanced energy fluxes and nonlinear energy transfers. Concomitantly, the peaks of the lobes migrate to higher frequencies. This is shown in figures 12–14.

(v) Variations in peakedness produce more significant changes in energy fluxes than do variations in angular spreading. This compares figure 9 to figure 15. In either case, the range considered was taken to cover what could be expected to occur in nature.

(vi) The nonlinear energy transfer should be smooth, as shown in figure 3, even for very coarse integration grids. Section 5.2 demonstrated that wave–wave interactions work to smooth perturbations introduced into the spectrum as spikes. The time taken for the nonlinear interactions to respond to any perturbation depends strongly on where it occurs within the spectrum.

(vii) Computations of relaxation time, expressed as the quotient of the equilibrium range energy by the sum of energy fluxes, agree well with equilibrium range flux parameterizations, as shown in figure 19. The more usual expression for relaxation time is the quotient of the equilibrium range energy by the nonlinear energy transfer $\mathscr{S}_{\mathrm{nl}}$.

We are motivated by these results to expect that it is possible to include the complete Boltzmann integral for nonlinear wave–wave interactions in a research wave model using the integration method of this paper. This approach should be an improvement over the present parameterizations of nonlinear transfer due to wave–wave interactions. Although the computer time required for this may still be prohibitive for operational wave modelling, it should be possible to investigate simple fetch- and duration-limited wave growth situations.

We want to thank Klaus Hasselmann for comments and criticisms over the past year and a half. These forced us to try to clarify our ideas and directed us in ways to improve our work.

We are thankful to Andrew McLeod who did much of the plotting under the supervision of Bechara Toulany; his work was supported by a Natural Sciences and Engineering Research Council Strategic Grant to John Walsh of the Centre for Cold Ocean Resource Engineering, Memorial University of Newfoundland.

This work was also supported by the Federal Panel on Energy Research and Development (PERD) of Canada.

REFERENCES

Barnett, T. P. 1968 On the generation, dissipation, and prediction of wind waves. *J. Geophys. Res.* **73**, 513–529.

Barnett, T. P. & Sutherland, S. 1968 A note on an overshoot effect in wind-generated waves. *J. Geophys. Res.* **73**, 6879–6885.

Donelan, M. A., Hamilton, J. & Hui, W. H. 1985 Directional spectra of wind-generated waves. *Phil. Trans. R. Soc. Lond.* A **315**, 509–562.

Ewing, J. A. 1971 A numerical wave prediction method for the North Atlantic Ocean. *Deutsche Hydrograph. Z.* **24**, 241–261.

Fox, M. J. H. 1976 On the nonlinear transfer of energy in the peak of a gravity-wave spectrum II. *Proc. R. Soc.* A **348**, 467–483.

Hasselmann, D. E., Dunkel, M. & Ewing, J. A. 1980 Directional wave spectra observed during JONSWAP 1973. *J. Phys. Oceanogr.* **10**, 1264–1280.

Hasselmann, K. 1961 On the nonlinear energy transfer in a gravity-wave spectrum. Part 1. General Theory. *J. Fluid Mech.* **12**, 481–500.

Hasselmann, K. 1963*a* On the nonlinear energy transfer in a gravity-wave spectrum. Part 2. Conservation theorems; wave-particle analogy; irreversibility. *J. Fluid Mech.* **15**, 273–281.

Hasselmann, K. 1963*b* On the nonlinear energy transfer in a gravity-wave spectrum. Part 3. Evaluation of the energy flux and swell-sea interaction for a Neumann spectrum. *J. Fluid Mech.* **15**, 385–398.

Hasselmann, K., Barnett, T. P., Bouws, E., Carlson, H., Cartwright, D. E., Enke, K., Ewing, J. A., Gienapp, H., Hasselmann, D. E., Kruseman, P., Meerburg, A., Müller, P., Olbers, D. J., Richter, K., Sell, W. & Walden, H. 1973 Measurements of wind–wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP). *Deutsche Hydrograph. Z. Suppl. Ser.* A (8), no. 12.

HASSELMANN, K., ROSS, D. B., MÜLLER, P. & SELL, W. 1976 A parametric wave prediction model. *J. Phys. Oceanogr.* **6**, 200–228.

HASSELMANN, S. & HASSELMANN, K. 1981 A symmetrical method of computing the nonlinear transfer in a gravity-wave spectrum. *Hamburger Geophys. Einzelschr.* A **52**, 163 pp.

HASSELMANN, S. & HASSELMANN, K. 1985 Computation and parameterizations of the nonlinear energy transfer in a gravity wave spectrum. Part I: A new method for efficient computations of the exact nonlinear transfer integral. *J. Phys. Oceanogr.* **15**, 1369–1377.

HASSELMANN, S., HASSELMANN, K., ALLENDER, J. H. & BARNETT, T. P. 1985 Computation and parameterizations of the nonlinear energy transfer in a gravity wave spectrum. Part II: Parameterizations of the nonlinear energy transfer for application in wave models. *J. Phys. Oceanogr.* **15**, 1378–1391.

HASSELMANN, S., HASSELMANN, K., KOMEN, G. K., JANSSEN, P., EWING, J. A. & CARDONE, V. 1989 The WAM model – A third generation ocean wave prediction model. *J. Phys. Oceanogr.* **18**, 1775–1810.

KITAIGORODSKII, S. A. 1983 On the theory of the equilibrium range in the spectrum of wind-generated gravity waves. *J. Phys. Oceanogr.* **13**, 816–827.

KOMEN, G. J., HASSELMANN, S. & HASSELMANN, K. 1984 On the existence of a fully developed windsea spectrum. *J. Phys. Oceanogr.* **14**, 1271–1285.

LONGUET-HIGGINS, M. S. 1976 On the nonlinear transfer of energy in the peak of a gravity-wave spectrum: a simplified model. *Proc. R. Soc.* A **347**, 311–328.

MASADA, A. 1980 Nonlinear energy transfer between wind waves. *J. Phys. Oceanogr.* **15**, 1369–1377.

MITSUYASU, H., TASAI, F., SUHARA, T., MIZUNO, S., OHKUSO, M., HONDA, T. & RIKIISHI, K. 1975 Observations of the directional spectra of ocean waves using a cloverleaf buoy. *J. Phys. Oceanogr.* **5**, 750–760.

PHILLIPS, O. M. 1985 Spectral and statistical properties of the equilibrium range in wind-generated gravity waves. *J. Fluid Mech.* **156**, 505–531.

RESIO, D. T. 1981 The estimation of a wind–wave spectrum in a discrete spectral model. *J. Phys. Oceanogr.* **11**, 510–525.

RESIO, D. T. 1987 Shallow-water waves. I: theory. *J. Waterway Port, Ocean Engng* **113** (3), 264–281.

RESIO, D. T. & PERRIE, W. A. 1989 Implications of an f^{-4} equilibrium range for wind-generated waves. *J. Phys. Oceanogr.* **19**, 193–204.

SELL, W. & HASSELMANN, K. 1972 Computations of nonlinear energy transfer for JONSWAP and empirical wind-wave spectra. *Rep. Inst. Geophys.* University of Hamburg.

TOBA, Y., OKADA, K. & JONES, I. S. F. 1988 The response of wind–wave spectra to changing winds. Part I: Increasing winds. *J. Phys. Oceanogr.* **18**, 1231–1240.

TRACY, B. A. & RESIO, D. T. 1982 Theory and calculations of the nonlinear energy transfer between sea waves in deep water. *WES Rep.* 11, US Army Engineer Waterways Experiment Station, Vicksburg, MS.

WEBB, D. J. 1978 Nonlinear transfer between sea waves. *Deep-Sea Res.* **25**, 279–298.

ZAKHAROV, V. E. & FILONENKO, N. N. 1966 Energy spectrum for stochastic oscillations of the surface of a liquid. *Dokl. Akad. Nauk SSSR* **160** (6), 1292–1295.

J. Fluid Mech. (1991), *vol.* 223, *pp.* 631–661
Printed in Great Britain

Buoyant inhibition of Ekman transport on a slope and its effect on stratified spin-up

By PARKER MacCREADY and PETER B. RHINES
School of Oceanography, WB-10, University of Washington, Seattle, WA 98195 USA

(Received 5 September 1989 and in revised form 3 August 1990)

The unsteady boundary layer of a rotating, stratified, viscous, and diffusive flow along an insulating slope is investigated using theory, numerical simulation, and laboratory experiment. Previous work in this field has focused either on steady flow, or flow over a conducting boundary, both of which yield Ekman-type solutions. After the onset of a circulation directed along constant-depth contours, Ekman-type flux up or down the slope is opposed by buoyancy forces. In the unsteady, insulating case, it is found that the cross-slope transport decreases in time as $(t/\tau)^{-\frac{1}{2}}$ where

$$\tau = \frac{1}{S^2 f \cos\alpha}\left(\frac{1/\sigma+S}{1+S}\right),$$

may be called the 'shut-down' time. Here $S = (N\sin\alpha/f\cos\alpha)^2$, f is the Coriolis frequency, α is the slope angle, N is the buoyancy frequency, and σ is the Prandtl number. Subsequently the along-slope flow, $\hat{v}$, approximately obeys a simple diffusion equation

$$\frac{\partial \hat{v}}{\partial t} = \nu\left(\frac{1/\sigma+S}{1+S}\right)\frac{\partial^2 \hat{v}}{\partial \hat{z}^2},$$

where t is time, ν is the kinematic viscosity, and $\hat{z}$ is the coordinate normal to the slope. By this process the boundary layer diffuses into the interior, unlike an Ekman layer, but at a rate that may be much slower than would occur with simple non-rotating momentum diffusion. The along-slope flow, $\hat{v}$, is nevertheless close to thermal wind balance, and the much-reduced cross-slope transport is balanced by stress on the boundary. For a slope of infinite extent the steady asymptotic state is the diffusively driven 'boundary-mixing' circulation of Thorpe (1987). By inhibiting the cross-slope transport, buoyancy virtually eliminated the boundary stress and hence the 'fast' spin-up of classical theory in laboratory experiments with a bowl-shaped container of stratified, rotating fluid.

1. Introduction

1.1. *Ekman layer on a slope with stratification*

The Ekman layer has a cross-isobar transport (the 'Ekman transport'), which, if horizontally divergent, drives a vertical velocity out of the boundary layer. This vertical velocity stretches or shrinks vortex lines in the interior, and the boundary layer may thereby affect large-scale atmospheric or oceanic flows, a process called 'spin-up'. The review article by Benton & Clark (1974) gives the early history of spin-up in many different contexts. If the fluid is stratified and the boundary is sloping, then buoyancy forces may impede the Ekman transport, lessening the vertical velocity, and significantly decreasing the effect of the boundary on the interior. In this paper we derive a theory of the boundary layer for rotating, stratified flow along a slope and then compare this theory with numerical and laboratory experiments.

Several authors have considered the flow of a rotating, stratified, viscous fluid over topography. By 'topography' we mean any boundary whose normal is not parallel to gravity. Fluid over such terrain may have isopycnals intersecting the slope, and hence there will be gradients of density on the slope. Above a flat boundary, scaling arguments (Pedlosky 1987, pp. 360–362) suggest that, if buoyancy and Coriolis forces are comparable in a depth H (the vertical scale of the interior flow), then buoyancy may be negligible within a boundary layer much thinner than H. This argument is often used (e.g. Pedlosky 1987, equation 6.6.9) to justify the use of simple Ekman theory for oceans or atmospheres with sloping lower boundaries. With a sloping boundary, however, persistent advection of buoyancy contrasts may eventually introduce significant new effects. The key effect is the excursion of fluid particles within the boundary layer, which tends to be far larger than that in the interior. During spin-up, for example, the classic solution without stratification shows excursion of near-boundary fluid through a distance $L(\epsilon H/\delta)$, which may readily exceed L, the lateral scale of the flow. Here δ is the thickness of the Ekman layer, $(\nu/\Omega)^{\frac{1}{2}}$, where ν is the kinematic viscosity, and Ω is the rotation rate of the fluid. The Rossby number, ϵ, is $U/\Omega L$, where U is the scale of the horizontal velocity. Although ϵ is typically small for large-scale geophysical flows, H/δ is typically large.

Holton (1967) first solved for a 'buoyant' Ekman layer when studying atmospheric flow over the Great Plains of the United States, a region of gradually sloping terrain. The density in Holton's model boundary layer has diurnal radiative forcing, with a specified temperature (and hence density) at the ground. An along-slope, geostrophic wind is specified in the interior. He finds that the boundary layer is a modified Ekman spiral, plus a thermally driven diurnal oscillation. The effect of the slope combined with the stratification is to create a buoyancy force which decreases the magnitude of the diurnal oscillation. Yet there is also a steady cross-slope transport associated with the mean along-slope wind. This steady transport is possible because, for example, as the Ekman transport drives cold, heavy air up-slope, this air is warmed by internal diabatic heating, forced by the temperature boundary condition.

Hsueh (1969) also solved for a buoyant Ekman layer, again specifying the flow in the interior and the temperature at the boundary. His analysis is for shallow slopes ($\alpha \leqslant 1$, where α is the slope angle from horizontal) and allows horizontal variation of the slope. His solution is similar to an Ekman layer, but of reduced thickness: $\delta(1+\sigma(\alpha N/f)^2)^{-\frac{1}{4}}$, where σ is the Prandtl number, ν/κ, and κ is the density diffusivity. N is the buoyancy frequency, and f is the Coriolis frequency, 2Ω. As in Holton's solution a steady up-slope transport is allowed by the diffusion of heat to or from the boundary.

If the sloping boundary is insulating instead of conducting, the buoyancy can no longer adjust as fluid moves up- or down-slope, except in the presence of diffusion to the interior. The cross-slope buoyancy flux must nevertheless enter into the force balance. Siegmann (1971), considering stratified spin-up in a spherical container with conducting walls, suggests that if the walls are insulating rather than conducting then there may be no order-one fast spin-up, since cross-slope boundary layer transport will be suppressed by the buoyancy. Our laboratory experiments (§5) show this to be largely correct in the limit of strong stratification and steep walls, but even in this limit there were unexpected results. For example, we found that the boundary layer was no longer confined to the narrow Ekman layer thickness, but instead diffused far into the interior.

Weatherly & Martin (1978) developed a numerical model of the turbulent boundary layer along a sloping ocean bottom, comparing it with data from the Western Atlantic on the continental slope. They account for buoyancy forces in their calculation, and one may observe in their model results (see their figure 9) a gradual slowing of the up-slope Ekman transport, although the calculation is not carried very far in time.

Above an insulating slope, such as the ocean floor, the boundary conditions for temperature and salinity require that isopycnal surfaces lie normal to the boundary. This is accomplished by diffusion, but the resulting tilted isopycnals are not in balance with buoyancy forces. For a non-rotating fluid the result, Phillips (1970), is a steady up-slope boundary layer current wherein up-slope advection of the density gradient exactly balances the diffusive down-slope density flux. Oceanographic attention has been focused recently on this class of 'self-propelled' boundary-layer flows (Phillips, Shyu & Salmun 1986; Thorpe 1987, and Garrett 1990) because of the possibility that enhanced mixing occurs at ocean boundaries. Diffusively driven flows on sloping boundaries may both mix stratified fluid and exchange fluid with the interior.

Thorpe (1987) discussed the boundary layer of a rotating, stratified, viscous, diffusive flow along an insulating slope of constant angle and infinite extent. His solution plays an important role in the time-dependent model developed below in §2. It is steady, with a vertical structure much like Hsueh's (1969) solution. Yet on application of the insulating boundary condition Thorpe finds that the interior flow far from the boundary is specified as a part of the solution. Thus, while steady solutions exist for any interior flow if density is specified at the boundary (as we see in Holton 1967 and Hsueh 1969), there is only one interior flow that has a steady boundary layer in the insulating case. Thorpe, like Holton, also presents an oscillatory solution, but it must oscillate about the steady solution, and the time-averaged properties are unchanged. For uniform ν and κ, the steady flow in Thorpe's solution is one that leads to an up-slope transport. This up-slope flow is allowed because it is balanced by a diffusive down-slope density flux, driven by the density boundary condition, as in the non-rotating solution of Phillips (1970). The boundary layer is, like the Ekman layer, confined to a thin region, and is unable to alter the interior except through meridional circulation.

The diversity of steady solutions in the literature, corresponding to different boundary conditions and values of σ, suggests the need for a theory with more than oscillatory time-dependence. What happens, for example, when an interior along-slope flow is 'switched-on' to a value different than Thorpe's? At one extreme (early time) the problem yields Ekman theory, in which the interior flow controls the boundary layer, which then feeds back on the interior by Ekman pumping. At the other extreme (late time) lies Thorpe's solution, where a steady boundary layer requires a particular value of the interior velocity, although how this state comes about is unclear. In more general applications, the presence of other insulating boundaries will cause the fluid eventually to be well-mixed and at rest, in the absence of sources or sinks of momentum or buoyancy. This suggests that it is crucial to know the rate of establishment of the quasi-steady Ekman- and boundary-mixing solutions, for they must compete with the external forcing affects that maintain the circulation and stratification of the fluid. We can anticipate a strong dependence on σ, and in particular a weakness of the boundary-mixing circulations for the large values of σ typical of laminar conditions in laboratory experiments.

1.2. *Classic spin-up*

While the boundary layer is accessible to theory and numerical solution, its small size makes visualization difficult in laboratory experiments. But we may readily observe the large-scale effects of boundary-layer transport in spin-up experiments. In the classic spin-up problem (Greenspan & Howard 1963) a right circular cylinder filled with homogenous, incompressible, viscous fluid is rotating initially at angular velocity Ω. The cylinder is then impulsively accelerated to a slightly greater angular velocity $\Omega+\Delta\Omega$. In a time $O(\Omega^{-1})$ an Ekman layer forms on the bottom boundary (we take the top to be a free surface). The ensuing Ekman transport drives a meridional circulation involving radial and vertical velocities in the interior. The interior fluid never directly feels the effects of viscosity, and approaches the new rotation rate in a timescale $E^{-\frac{1}{2}}\Omega^{-1}$, where E is the Ekman number, $(\delta/H)^2$, and H is the height of the cylinder. For small Ekman number this spin-up timescale is much faster than that for penetration of viscous effects into the interior, $E^{-1}\Omega^{-1}$. For typical laboratory parameters ($\nu = 0.01\ \mathrm{cm^2\,s^{-1}}$, $\Omega = 1\ \mathrm{s^{-1}}$, $H = 10$ cm) the 'fast' spin-up occurs in 100 s, whereas the viscous timescale is nearly three hours, and this contrast is magnified in flows of geophysical scale.

It is essential to incorporate stratification if spin-up is to apply to atmosphere and ocean flows. Holton (1965) divides the equations of motion into interior and boundary-layer parts, scaling vertical derivatives in the boundary layer as $E^{-\frac{1}{2}}$ greater than those in the interior. He takes the timescale of the problem to be the 'fast' spin-up time, $E^{-\frac{1}{2}}\Omega^{-1}$, based on the results of Greenspan & Howard (1963), and expands all dependent variables in powers of $E^{\frac{1}{2}}$, the obvious small parameter of the problem.

The primary result of Holton's analysis is that the 'fast' spin-up process no longer extends through the entire depth of the fluid, but is confined by the buoyancy within a 'Prandtl scale' $H_{\mathrm{P}} = fL/N$ above the bottom boundary (the subscript P is for Prandtl). L is the horizontal lengthscale of the forcing, typically the tank radius. Holton also finds that stratified spin-up is faster than in the homogenous case, owing to the reduced height of penetration, H_{P}. At the end of the 'fast' spin-up process the fluid in the interior has considerable vertical shear, which is removed by viscosity. Diffusive effects at $O(E^{\frac{1}{2}})$ and $O(E)$ were considered by St-Maurice & Veronis (1975), who find both a gradual migration to a diffusively controlled interior, and modification of the 'fast' spin-up by viscosity.

Spin-up theory has provided justification for a simple parameterization of bottom boundary friction of geophysical flows, essentially as a surface drag that is linear in the geostrophic velocity. Numerical and analytical models of large-scale flow thus often assume that relative vorticity decays exponentially with timescale $E^{-\frac{1}{2}}\Omega^{-1}$, rather than resolve the boundary layer itself. The exact timescale is, in practice, difficult to determine because δ and H_{P} are hard to measure, but their variation is not so great as to make the theory unworkable.

In §2 we develop the boundary-layer equations, and derive approximate time-dependent solutions for an insulating boundary with constant ν and κ, and an impulsively started interior flow. We compare these to numerical solutions to the full one-dimensional problem in §3. The balance of the paper is a description and discussion of laboratory experiments. We carried out stratified spin-up experiments in a container with a sloping bottom boundary, using stratification large enough to inhibit Ekman transport well before 'fast' spin-up could influence the interior flow.

2. The time-dependent Ekman layer with stratification

2.1. *Development of the equations*

The equations of motion and mass conservation for a Boussinesq, incompressible fluid, in a reference frame rotating at angular velocity $\Omega \boldsymbol{k}$ may be written as

$$\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t}+f(\boldsymbol{k}\times\boldsymbol{u}) = -\frac{\nabla p}{\rho_0}-\frac{g\rho\boldsymbol{k}}{\rho_0}+\nu\nabla^2\boldsymbol{u}, \tag{2.1a}$$

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \kappa\nabla^2\rho, \tag{2.1b}$$

$$\nabla\cdot\boldsymbol{u} = 0. \tag{2.1c}$$

Here $\boldsymbol{u}$ is the velocity vector (u,v,w) in Cartesian coordinates (x,y,z), ∇ is the gradient $(\partial/\partial x, \partial/\partial y, \partial/\partial z)$, $\mathrm{d}/\mathrm{d}t$ is the material derivative $\partial/\partial t+\boldsymbol{u}\cdot\nabla$, $\boldsymbol{k}$ is the vertical unit vector $(0,0,1)$, and g is gravity.

The density ρ has been separated into three parts such that

$$\rho = \rho_0+\bar{\rho}(z)+\rho'(x,y,z,t). \tag{2.2}$$

We wish to consider a system with small density variations from the mean, ρ_0, hence $\rho' \ll \rho_0$ and $\bar{\rho} \ll \rho_0$. We allow ρ' to be as large as $\bar{\rho}$, so that the time-dependent density variation could, for example, overwhelm the static stability of the stratification. We define the buoyancy frequency by

$$N^2 = -\frac{g}{\rho_0}\frac{\partial\bar{\rho}}{\partial z}. \tag{2.3}$$

We take N to be constant in our analysis.

The pressure p is separated into two parts:

$$p = \bar{p}(x,z)+p'(x,y,z,t). \tag{2.4}$$

The z-dependence of $\bar{p}$ is taken to be hydrostatic, hence

$$\frac{\partial\bar{p}}{\partial z} = -g(\rho_0+\bar{\rho}). \tag{2.5}$$

The x-dependence of $\bar{p}$ is used to introduce an along-slope geostrophic velocity, V, in the interior, given by

$$V = \frac{1}{f\rho_0}\frac{\partial\bar{p}}{\partial x}. \tag{2.6}$$

V is constant in both time and space (after time $t = 0$), and is specified as an initial condition of the problem.

We want to describe the development of the boundary layer for flow along a slope. Following Phillips (1970) we simplify the problem by considering flow along a boundary of constant slope, $\tan\alpha$. While this one-dimensional geometry also eliminates boundary-layer divergence, parametric variations of interior velocity or slope can later incorporate this 'pumping', which is crucial to classic spin-up. We then rotate the equations into the slope frame of reference, as defined in figure 5(*b*). All variables in the new frame of reference will be denoted by $\hat{}$. The rotated velocity vector components are the up-slope velocity $\hat{u}$, the along-slope velocity $\hat{v}$, and the velocity normal to the slope, $\hat{w}$, in the corresponding Cartesian coordinate system, $\hat{x}$,

$\hat{y}$, $\hat{z}$. We assume $\hat{u}$, $\hat{v}$, $\hat{w}$, ρ', and p' have no $\hat{x}$ or $\hat{y}$ variation, which is reasonable if the initial condition has no such variation. This assumption makes the problem inherently linear, and disallows velocities normal to the boundary, i.e. $\hat{w} = 0$. Hence the solutions cannot represent the overturning of unstable stratifications. In geophysical boundary flows such overturning is often parameterized by increased eddy viscosity and diffusivity in the unstable region. For the purposes of our analysis we shall assume that ν and κ are constant, and simply point out where the solutions are not statically stable.

Writing the equations in the rotated frame of reference, and adding $\cos\alpha$ times the $\hat{x}$-momentum equation to $\sin\alpha$ times the $\hat{z}$-momentum equation to eliminate the pressure, we find

$$\frac{\partial \hat{u}}{\partial t} - f\cos\alpha(\hat{v}-V) = -B\sin\alpha + \nu\frac{\partial^2 \hat{u}}{\partial \hat{z}^2}, \tag{2.7a}$$

$$\frac{\partial \hat{v}}{\partial t} + f\cos\alpha\,\hat{u} = \nu\frac{\partial^2 \hat{v}}{\partial \hat{z}^2}, \tag{2.7b}$$

$$\frac{\partial B}{\partial t} = N^2\sin\alpha\,\hat{u} + \kappa\frac{\partial^2 B}{\partial \hat{z}^2}. \tag{2.7c}$$

We have written B for the buoyancy, $g\rho'/\rho_0$, to simplify the notation. The boundary conditions are

$$\hat{u} = \hat{v} = 0 \quad \text{at} \quad \hat{z} = 0, \tag{2.8a}$$

$$\frac{\partial B}{\partial \hat{z}} = N^2\cos\alpha \quad \text{at} \quad \hat{z} = 0, \tag{2.8b}$$

$$\hat{u} \quad \text{and} \quad B \rightarrow 0 \quad \text{as} \quad \hat{z} \rightarrow \infty, \tag{2.8c}$$

and

$$\hat{v} \rightarrow V \quad \text{as} \quad \hat{z} \rightarrow \infty. \tag{2.8d}$$

Thus there is a no-slip velocity boundary condition, and the slope is insulating. When $\kappa = 0$, (2.8*b*) should be replaced by

$$B = 0 \quad \text{at} \quad \hat{z} = 0. \tag{2.8e}$$

When $\alpha = 0$, (2.7) gives rise to a standard Ekman layer. When $N = 0$ the solution is a modified Ekman layer with thickness

$$\delta_s = (2\nu/f\cos\alpha)^{\frac{1}{2}}. \tag{2.9}$$

The subscript s indicates that this is in the slope frame of reference.

Buoyancy becomes important to the momentum balance through the term $B\sin\alpha$ in (2.7*a*), which grows in magnitude initially by advection of the stratification, and by diffusion of the boundary condition. For the non-diffusive ($\kappa = 0$) case we may simply estimate when buoyancy will first become important. Assuming that buoyancy is initially unimportant the solution will be approximately the modified Ekman layer of thickness δ_s described above. In the boundary layer we may then make the scale estimates:

$$(\hat{v}-V) \quad \text{and} \quad \hat{u} \sim -V. \tag{2.10}$$

Integrating (2.7*c*) in time and using the scale estimate, the buoyancy within the boundary layer is approximately

$$B \sim -VN^2\sin\alpha\,\tau_0, \tag{2.11}$$

at a time τ_0. Substituting this expression into the $\hat{x}$-momentum equation (2.7*a*) we find that the buoyancy term becomes as large as the Coriolis term when

$$\tau_0 \sim \frac{f\cos\alpha}{(N\sin\alpha)^2}. \tag{2.12}$$

The above derivation was based on the assumption that Ekman theory was workable for some time before buoyancy became important, which implies $(N/f)\tan\alpha \ll 1$. If this were not the case then buoyancy would presumably become important while the Ekman layer was forming. Walin (1969) suggests that Ekman-layer theory remains valid if $(N/f)\tan\alpha$ is small. Our scale analysis (and full theory to be presented below) in fact predicts that buoyancy forces eventually become substantial unless $(N/f)\tan\alpha$ is actually zero.

2.2. *Steady solutions*

Thorpe's (1987) steady solution to (2.7) subject to boundary conditions (2.8) is

$$\hat{u} = \frac{2\kappa\cot\alpha}{\delta_{\mathrm{T}}}\exp(-\hat{z}/\delta_{\mathrm{T}})\sin(\hat{z}/\delta_{\mathrm{T}}), \tag{2.13a}$$

$$\hat{v} = V_{\mathrm{T}}(1-\exp(-\hat{z}/\delta_{\mathrm{T}})\cos(\hat{z}/\delta_{\mathrm{T}})), \tag{2.13b}$$

where

$$\delta_{\mathrm{T}} = \left(\frac{4\nu^2}{(f\cos\alpha)^2+\sigma(N\sin\alpha)^2}\right)^{\frac{1}{4}}, \tag{2.14a}$$

and

$$V_{\mathrm{T}} = -2(\kappa\cot\alpha)\,\delta_{\mathrm{T}}/\delta_{\mathrm{s}}^2. \tag{2.14b}$$

This solution has roughly the form of an Ekman layer of thickness δ_{T}. The subscript T is for Thorpe. The along-slope velocity in the interior is fixed by the density boundary condition to a constant value, V_{T}. The cross-slope transport is completely determined by the diffusivity and the slope angle, as seen by the integral

$$\int_0^\infty \mathrm{d}\hat{z}\,\hat{u} = \kappa\cot\alpha. \tag{2.15}$$

This result, which comes directly from integration of the steady form of (2.7*c*), with boundary condition (2.8*b*), expresses the fact that for the steady problem there must be a global balance of advective and diffusive buoyancy fluxes. Thorpe also extends his solution to the case where the viscosity and diffusivity vary away from the slope, to represent flows with a mixed layer at the boundary. Still, it remains difficult to apply these solutions to geophysical situations where the along-slope velocity is arbitrary. Garrett (1990) has addressed this problem by suggesting that the thickness and diffusivity of the mixed layer may adjust to conform to the interior flow. He finds that it may be possible to have such a solution for arbitrary interior velocity V of positive sign.

Even if the diffusivity and viscosity are variable, steady solutions still strongly limit the cross-slope transport. Thorpe shows that the integrated transport in this case (still holding N constant) is always given by $\kappa_\infty\cot\alpha$, where κ_∞ is the value of κ as $\hat{z}\rightarrow\infty$. This would have drastic consequences for geophysical flows, through the fast spin-up process, if the steady solutions were always in force. Taking this as a caution about the applicability of steady solutions we shall explore the time-dependent case analytically below.

2.3. *Approximate unsteady solutions*

Consider a situation where, initially, $\hat{u} = B = 0$, and $\hat{v} = V$ everywhere except at the boundary. In the absence of stratification, an Ekman layer of thickness δ_s will develop in a time $O(\Omega^{-1})$. With stratification, buoyancy will become important in a time τ_0, (2.12).

Let us assume for the moment that the buoyancy has become sufficiently important that the along-slope velocity $\hat{v}$ has come approximately into 'thermal wind' balance with the density field. This would be expressed by an approximate version of the $\hat{x}$-momentum equation (note that (2.16) is actually a $\hat{z}$-integral of the thermal wind equation):

$$\frac{\hat{v}}{V} - 1 = B\frac{\sin\alpha}{Vf\cos\alpha} + O(R_t, E_s)\frac{U}{V}, \tag{2.16}$$

where U is the scale of the $\hat{u}$-velocity,

$$R_t = (\tau_u f\cos\alpha)^{-1}, \tag{2.17}$$

a temporal Rossby number, and

$$E_s = (\delta_s/D_u)^2, \tag{2.18}$$

an Ekman number. Here τ_u is the timescale of the temporal variations of $\hat{u}$, and D_u is the thickness of the boundary layer for $\hat{u}$. Below we shall be able to make more meaningful estimates of when and where these are small, and hence assess the validity of the thermal wind approximation in (2.16).

Assuming that R_t and E_s are negligibly small, we take $\partial/\partial t$ and $\partial^2/\partial\hat{z}^2$ of (2.16) and substitute the results into the buoyancy equation (2.7*c*) to find

$$\frac{f\cos\alpha}{\sin\alpha}\frac{\partial\hat{v}}{\partial t} = N^2\sin\alpha\,\hat{u} + \kappa\frac{f\cos\alpha}{\sin\alpha}\frac{\partial^2\hat{v}}{\partial\hat{z}^2}. \tag{2.19}$$

Solving this for $\hat{u}$ and substituting into the $\hat{y}$-momentum equation (2.7*b*) the result may be written as

$$\frac{\partial\hat{v}}{\partial t} = \nu\left(\frac{1/\sigma + S}{1+S}\right)\frac{\partial^2\hat{v}}{\partial\hat{z}^2}, \tag{2.20}$$

where

$$S = \left(\frac{N\sin\alpha}{f\cos\alpha}\right)^2, \tag{2.21}$$

a Burger number. Thus the along-slope momentum dynamics have been reduced to a simple diffusion equation, despite the presence of both rotation and stratification. We shall refer to (2.20) as the 'slow diffusion' equation, since for $\sigma > 1$ it predicts that the boundary layer will diffuse inward more slowly than the usual non-rotating boundary layer. The cross-slope flow leads to Coriolis forces which oppose the diffusion of along-slope momentum into the fluid. This is the effect that slows the diffusion and, in the case of an Ekman layer, brings it to a halt. Yet the growing buoyancy force does not allow a steady Ekman balance, and so the boundary layer continues to thicken.

The slow-diffusion equation (2.20) is, in the limit $S \ll 1$ (and $S \sim O(1/\sigma)$), identical to a result derived previously by Gill (1981) for the evolution of the density field during the spin-down of a frontal region in the ocean interior (see also Garrett 1982). In their context slow diffusion is seen as an enhanced lateral diffusion.

Using the slow-diffusion equation (2.20) we may also estimate the cross-slope transport. Taking the $\hat{z}$-integral of the buoyancy equation (2.7*c*) and applying the insulating boundary condition (2.8*b*) we form

$$\int_0^\infty \mathrm{d}\hat{z}\frac{\partial B}{\partial t} = MN^2\sin\alpha - \kappa N^2\cos\alpha, \tag{2.22}$$

where
$$M = \int_0^\infty \mathrm{d}\hat{z}\,\hat{u}, \tag{2.23}$$

the cross-slope transport. Substituting $\hat{v}$ for B as before from the approximate thermal wind equation (2.16) (with R_t and E_s equal to zero) and rearranging we find

$$M = \frac{1}{Sf\cos\alpha}\int_0^\infty \mathrm{d}\hat{z}\frac{\partial \hat{v}}{\partial t} + \kappa\cot\alpha. \tag{2.24}$$

Hence, if the flow does attain a steady state, the cross-slope transport is given by Thorpe's result (2.15). Assuming that the slow-diffusion equation (2.20) adequately describes the time rate of change of $\hat{v}$ over most of the boundary layer, we may use it to obtain a scale estimate of the integral term in (2.24). By time t, $\hat{v}$ will have changed by an amount comparable with $-V$ (assuming $\hat{v}=0$ is the proper bottom boundary condition for (2.20)) in a region of thickness D_v estimated from (2.20) as

$$D_v \sim \left[\nu t\left(\frac{1/\sigma+S}{1+S}\right)\right]^{\frac{1}{2}}. \tag{2.25}$$

Hence we may form the scale estimate

$$\int_0^\infty \mathrm{d}\hat{z}\frac{\partial \hat{v}}{\partial t} \sim D_v\left(\frac{-V}{t}\right). \tag{2.26}$$

Using this result the transport equation may be rewritten as

$$\frac{M}{|M_0|} = C\left(\frac{\tau}{t}\right)^{\frac{1}{2}} + \frac{\kappa\cot\alpha}{|M_0|}, \tag{2.27}$$

where
$$\tau = \frac{1}{S^2 f\cos\alpha}\left(\frac{1/\sigma+S}{1+S}\right). \tag{2.28}$$

We shall call τ the 'shut-down time' because it gives the timescale over which the cross-slope transport relaxes to the steady limit of Thorpe's solution, $\kappa\cot\alpha$. C is an $O(1)$ constant to be determined empirically. We have normalized the equation by $|M_0| = |-\frac{1}{2}V\delta_s|$, the magnitude of the steady Ekman transport when $N=0$, since this will be the scale of the transport before buoyancy becomes important.

For $\sigma=\infty$ and $S \ll 1$ the shut-down time, τ, is equal to τ_0, the timescale we determined in (2.12) for the onset of buoyancy effects in the boundary layer. Thus, at least for this simple case, the shut-down time is the time it takes for cross-slope advection to significantly alter the force balance within the boundary layer.

When $\sigma = 1$ the shut-down time varies as α^{-4}, indicating that care should be taken when applying the theory to regions of non-constant slope. In particular, the horizontal lengthscale of variation in τ should be much greater than the boundary-layer thickness.

The analysis above hinged upon having very small values of R_t and E_s, the scales of the inertial and viscous terms in the $\hat{x}$-momentum equation (2.16). While the boundary layer is initially forming, the timescale for changes in $\hat{u}$ will be $(f\cos\alpha)^{-1}$, and R_t will be $O(1)$, so clearly our results do not apply for this early time. After this time the transport equation (2.27) suggests that the timescale for changes in $\hat{u}$ will be τ, the shut-down time. Thus we expect $R_t \ll 1$ if (i) $t \gg (f\cos\alpha)^{-1}$, and (ii) $\tau \gg (f\cos\alpha)^{-1}$. We shall see in §3 that violation of this constraint on τ does not significantly alter our results. The viscous term E_s will also initially be at least $O(1)$ for $t < (f\cos\alpha)^{-1}$, since the boundary layer is thinner than δ_s before that time. If the boundary layer for $\hat{u}$ later thickens at the same rate as slow diffusion (2.20) predicts for $\hat{v}$ then E_s will be small when $D_v \gg \delta_s$. In general we expect that $E_s \ll 1$ if (i) $t \gg (f\cos\alpha)^{-1}$, and (ii) $D_u \gg \delta_s$. We explore the validity of these expectations numerically below.

3. Numerical solutions

The set of coupled equations (2.7) describing the boundary layer was solved numerically, using forward-differencing in time, and central-differencing in space. Each run was initialized with $\hat{u}$ and $\hat{v}$ set equal to the steady $N = 0$ Ekman-layer solution on a slope, and with $B = 0$. This was done to minimize inertial oscillations in the solutions, similar to the gradual 'switching on' of the boundary condition as seen, for example, in Weatherly & Martin (1978). We are thus necessarily concentrating only on boundary-layer behaviour for $t > (f\cos\alpha)^{-1}$. Runs started with undisturbed initial conditions, $\hat{u} = 0$, $\hat{v} = V$, and $B = 0$, had the same general behaviour as the solutions shown below, but had larger inertial oscillations, making the results more difficult to see. There were approximately six grid points within the initial boundary-layer thickness, δ_s (typically 0.14 cm), and 2400 time steps per period of revolution. Integrations covered a depth of at least $70\delta_s$, and the boundary layer never significantly interacted with the upper boundary. The numerical scheme was checked against known behaviour (e.g. final steady velocity profiles, and timescale for decay of transients) of the unstratified case.

3.1. *Numerical integrations with little or no diffusivity*

Our laboratory experiments were salt stratified, so density diffusion was essentially negligible over the timescales of interest. To compare with these experiments we first present results of numerical integrations for the case $\kappa = 0$. Throughout all the numerical results, $f = 1\ \mathrm{s}^{-1}$, $N = 2\ \mathrm{s}^{-1}$, and $\nu = 0.01\ \mathrm{cm}^2\ \mathrm{s}^{-1}$, values typical of the laboratory experiments.

Figure 1(*a*) shows $\hat{u}$- and $\hat{v}$-profiles versus $\hat{z}$ at three different dimensionless times for $V = -1\ \mathrm{cm\ s}^{-1}$ (corresponding to up-slope boundary-layer transport). Note that the $V = \pm 1\ \mathrm{cm\ s}^{-1}$ solutions are symmetrical in their velocity fields. The dashed lines are solutions to the slow diffusion equation (2.20) at the same times (also initialized with the steady $N = 0$ Ekman velocity profile, and with a no-slip lower boundary condition). The $\hat{v}$ boundary layer thickened almost exactly as predicted by the slow diffusion equation, especially for large t/τ. The up-slope flow decreased in magnitude over time, and extended over roughly the same thickness as the $\hat{v}$ boundary layer. In this instance, lacking density diffusion, it was the cross-slope velocity $\hat{u}$ that had to advect the density field to bring $\hat{v}$ into thermal wind balance. The accuracy of the slow diffusion equation in predicting the $\hat{v}$-velocity is an indication that the thermal wind assumption was valid over almost the whole depth of the boundary layer, particularly for larger t/τ (inspection of the individual terms in (2.7*a*) during the integration also showed this to be true).

As $\hat{u}$ advects the stratification up- or down-slope, there is the clear possibility that the resulting density field may not be statically stable. Figure 1(*b*) shows the density

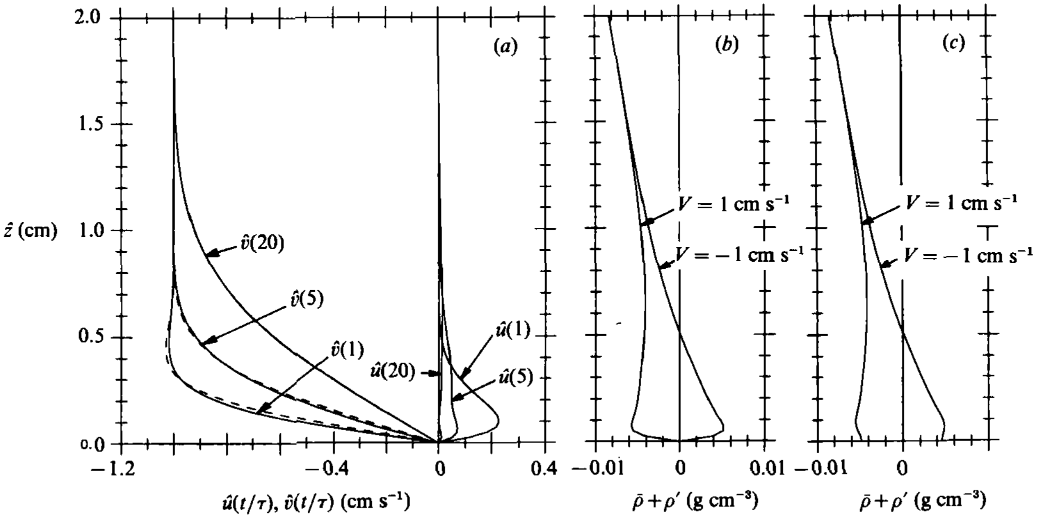


FIGURE 1. Numerical solutions of (2.7) versus $\hat{z}$, with $\alpha = 10°$, $f = 1\ \mathrm{s}^{-1}$, $N = 2\ \mathrm{s}^{-1}$, and $\nu = 0.01\ \mathrm{cm}^2\ \mathrm{s}^{-1}$. Along-slope ($\hat{v}$) and cross-slope ($\hat{u}$) velocity profiles are shown in (*a*) for $\kappa = 0$ and $V = -1\ \mathrm{cm\ s}^{-1}$, at three dimensionless times: $t/\tau = 1$, 5, and 20 ($\tau = 7.3$ s). The $\hat{v}$-profiles are compared with solutions (dashed lines) of the slow-diffusion equation (2.20) for the same times (at $t/\tau = 20$ the profiles are nearly identical). The density perturbation, $\bar{\rho}+\rho'$, at $t/\tau = 20$ is shown in (*b*) with all parameters as in (*a*) except that $V = \pm 1\ \mathrm{cm\ s}^{-1}$. The effect of a small density diffusivity, $\kappa = 10^{-5}\ \mathrm{cm}^2\ \mathrm{s}^{-1}$, on the profiles in (*b*) is shown in (*c*).

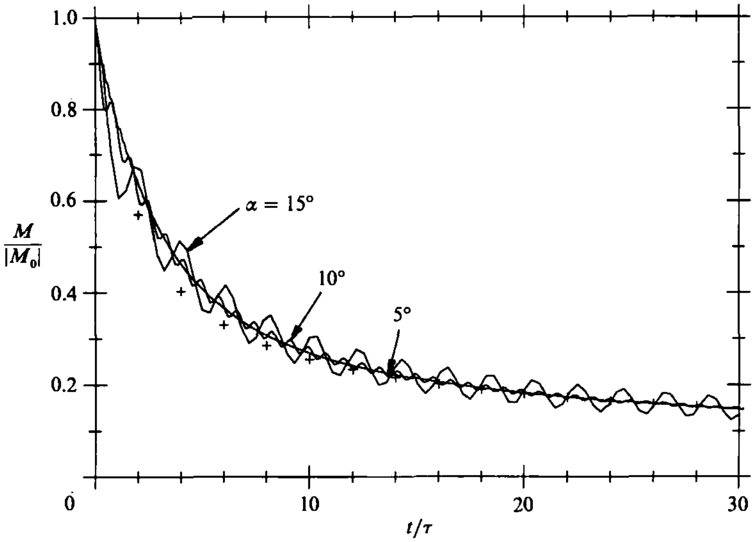


FIGURE 2. Normalized cross-slope transport versus t/τ from three numerical solutions to (2.7). All parameters were as in figure 1 (*a*) except the slope angle, which was varied as shown. Also shown (+) is a fit to the data given by $M/|M_0| = 0.8072\ (t/\tau)^{-\frac{1}{2}}$.

perturbation, $\rho+\rho'$, at $t/\tau = 20$, for two different interior along-slope velocities, $V = +1\ \mathrm{cm\ s}^{-1}$ (causes a down-slope transport), and $V = -1\ \mathrm{cm\ s}^{-1}$ (causes an up-slope transport). Figure 1 (*c*) shows the density perturbation for the same situations as in (*b*) except with $\sigma = 10^3$, representative of the actual molecular diffusivity of salt. The up-slope favourable case ($V = -1\ \mathrm{cm\ s}^{-1}$) was statically unstable only in a

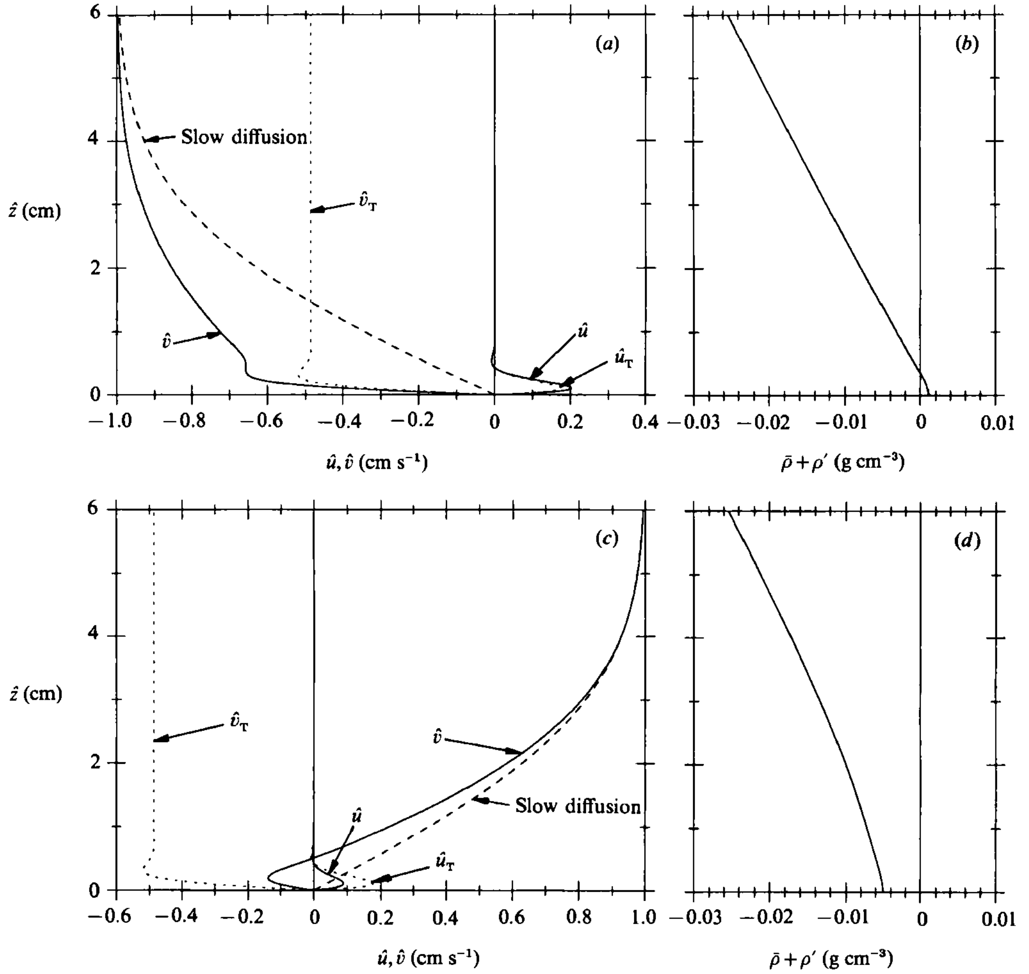


FIGURE 3. Numerical solutions of (2.7) versus $\hat{z}$, with $\alpha = 15°$, $f = 1$ s^{-1}, $N = 2$ s^{-1}, and $\nu = \kappa = 0.01$ cm^2 s^{-1}; all at $t/\tau = 20$ ($\tau = 12.6$ s). In (a) and (b) $V = -1$ cm s^{-1}, whereas in (c) and (d) $V = 1$ cm s^{-1}. The $\hat{v}$-profiles are compared with solutions (dashed lines) of the slow diffusion equation (2.20). Also shown (dotted lines, marked $\hat{u}_T$ and $\hat{v}_T$) are the $\hat{u}$- and $\hat{v}$-profiles for Thorpe's steady solution (2.13) with the same parameters (except V).

region very close to the boundary, and most of this unstable region was removed by the small diffusivity. In all of the laboratory experiments presented below this was the sense of the interior velocity field, hence we ignore static instability in our analysis of the experiments. By contrast, the down-slope case ($V = +1$ cm s^{-1}) was marginally unstable over much of the boundary layer, and the diffusivity did little to change this situation. Thus we expect that for $\sigma \gg 1$ our theory may need to be modified for interior velocities of positive sign, to account for possible static instability. Indeed, some laboratory experiments (not presented here) involving down-slope flow did show signs of instability in the boundary layer.

Figure 2 shows the normalized cross-slope transport $M/|M_0|$, versus non-dimensional time t/τ, for runs with three different slope angles, $\alpha = 5°$, $10°$, and $15°$, with corresponding shut-down times 31.8, 7.3 and 2.8 s. For these runs $\kappa = 0$ and $V = -1$ cm s^{-1}. All other parameters were as before. The three curves collapsed to one (except for the inertial oscillations), indicating that τ was the correct parameter to non-dimensionalize the time. Also plotted $(+)$ is a fit to the curves based on our transport prediction (2.27). The constant C we used was 0.8072, which came from fitting (2.27) to the numerical data at $t/\tau = 30$. The fit looks good, even as early as $t/\tau = 2$, indicating that, at least for $\kappa = 0$, the transport resembles our prediction.

When $\alpha = 15°$, $\tau = 2.8$ s, and so we would expect buoyancy effects to become important even before the Ekman layer is fully set up, possibly violating the scaling requirement $\tau \gg (f\cos\alpha)^{-1}$. The only noticeable effect of this short τ, however, was to excite somewhat larger inertial oscillations than in the cases with longer τ.

3.2. *Numerical integrations with large diffusivity*

Subsequent numerical runs were carried out with $\kappa = \nu = 0.01\ \mathrm{cm^2\,s^{-1}}$ $(\sigma = 1)$ to explore the effects of large density diffusion upon the boundary layer. Again, $f = 1\ \mathrm{s^{-1}}$, and $N = 2\ \mathrm{s^{-1}}$. For all runs the slope was 15°. The only parameter that was varied was the sign of the initial interior velocity, V, being either plus or minus $1\ \mathrm{cm\,s^{-1}}$.

Figure 3(*a*, *c*) shows $\hat{u}$- and $\hat{v}$-velocity profiles versus $\hat{z}$ at $t/\tau = 20$ $(\tau = 12.6$ s) for (*a*) $V = -1\ \mathrm{cm\,s^{-1}}$, and (*c*) $V = +1\ \mathrm{cm\,s^{-1}}$. Plotted as dashed lines are solutions to the slow diffusion equation at the same time. For comparison the $\hat{u}$- and $\hat{v}$-velocities for Thorpe's steady solution with the same parameters (except V, which we are not free to specify) are plotted as dotted lines. In both cases the $\hat{v}$-velocity profile had diffused away from the boundary approximately as much as the slow-diffusion solution predicted, but the magnitude of the $\hat{v}$-velocity did not match the slow-diffusion solution, especially near the boundary. The $\hat{u}$-velocity had become positive in both cases by this time, and was concentrated near the boundary in a Thorpe-like profile. In these cases the assumption of thermal wind balance in $\hat{v}$ was probably flawed close to the boundary, owing to the density boundary condition. Specifically, the Ekman number, E_{S}, had probably become large there owing to the thinness of D_u.

The stratification parameter, $\bar{\rho}+\rho'$, is shown at the same time, for these two cases in figure 3(*b*, *d*). In both cases the stratification remained statically stable because of smoothing by the density diffusion. The perturbation to the mean stratification diffused upwards approximately as far as the $\hat{v}$-velocity profiles. In contrast to the $\kappa = 0$ cases, here diffusion was the primary means of altering the density, except near the boundary where $\hat{u}$ remained large.

Figure 4(*a*) shows two families of $\hat{v}$-velocity profiles (with parameters as in figure 3) over a long period of time. The profiles suggest strongly that the time-dependent solution is moving toward Thorpe's steady solution. The $\hat{u}$-velocity may achieve this goal fairly easily since it is concentrated near the boundary, but the $\hat{v}$-velocity must change everywhere to attain the steady solution. This then appears to be the role of slow diffusion: the gradual transformation of the interior along-slope flow to that of Thorpe's steady solution, V_{T}. In addition, the numerical solutions imply that V_T is eventually the correct bottom boundary condition for the slow-diffusion equation, not $\hat{v} = 0$ as we used. When $\kappa = 0$, V_{T} is also zero, which explains why the slow-diffusion solution was so accurate in that case (figure 1).

The normalized cross-slope transport for the two runs of figure 3 is plotted versus t/τ in figure 4(*b*). Also plotted $(+)$ are fits to the runs. The fits are from

$$\frac{M}{|M_0|} = C\frac{(M_0-(\kappa\cot\alpha)\,\delta_{\mathrm{T}}/\delta_{\mathrm{S}})}{|M_0|}\left(\frac{\tau}{t}\right)^{\frac{1}{2}}+\frac{\kappa\cot\alpha}{|M_0|}, \tag{3.1}$$

which was derived in exactly the same manner as (2.27), but using the idea,

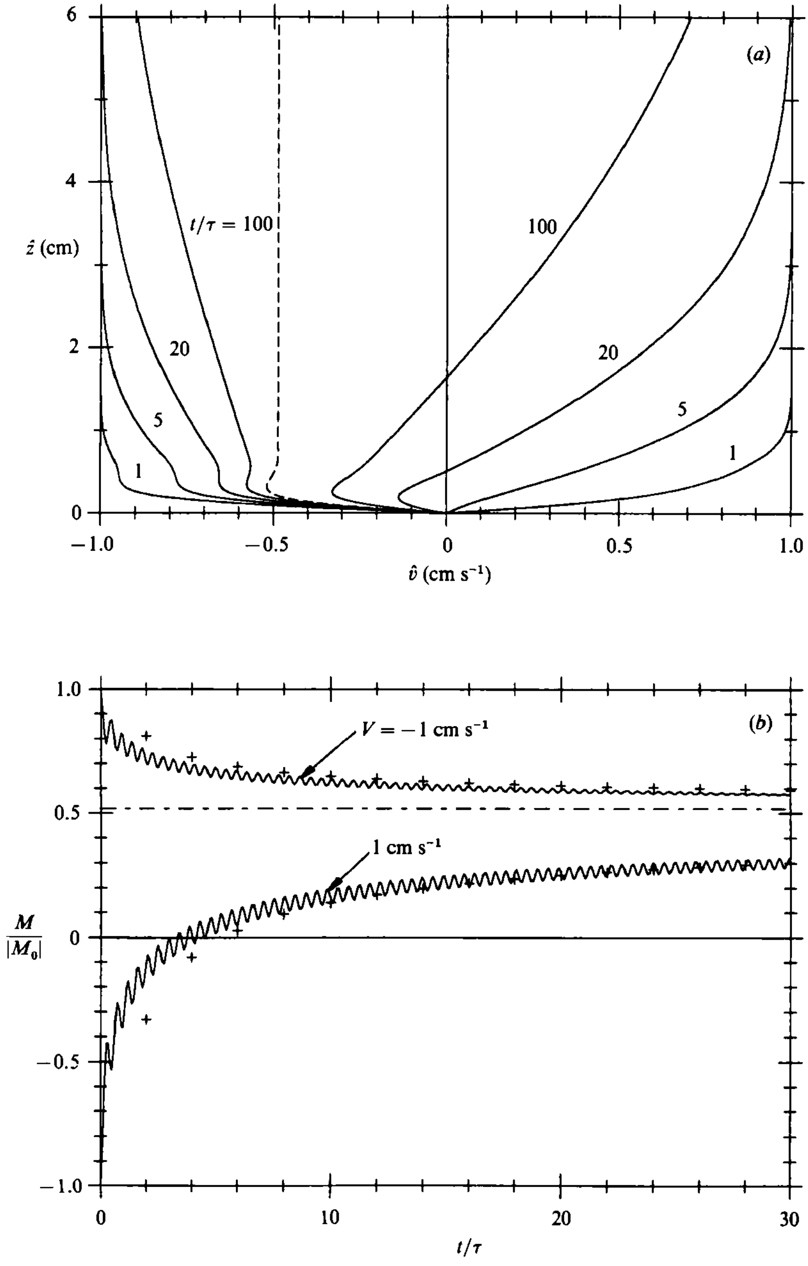


FIGURE 4. (*a*) Numerical solutions for $\hat{v}$ at different times (times shown are non-dimensionalized by τ), with parameters as in figure 3. Thorpe's steady solution for $\hat{v}$ is also shown (dashed line). Profiles to the left of the dashed line had $V = -1$ cm s^{-1}, and those to the right had $V = 1$ cm s^{-1}. Normalized cross-slope transport versus t/τ is plotted in (*b*) for two numerical solutions to (2.7). All parameters were as in figure 3. Also shown (+) are fits to the data based on (3.1). The limiting value of the (normalized) transport, $\kappa \cot\alpha/|M_0|$, is also shown (dashed line).

suggested by the numerical solutions, that V_T is eventually the proper bottom boundary condition for the slow diffusion equation. Hence instead of scaling $\partial\hat{v}/\partial t$ as $-V/t$, we used $\partial\hat{v}/\partial t \sim -(V-V_T)/t$. This should somewhat overestimate the magnitude of $\partial\hat{v}/\partial t$, since the assumed bottom boundary condition was actually only gradually achieved, and this was the sense of the error in the plotted fits. Note that the analysis in §2 is unable to determine the rate at which this new bottom boundary condition is established. The fits used the same empirical constant, $C = 0.8072$, as was determined from the $\kappa = 0$ case, and are still fairly good, especially at larger t/τ.

The theoretical ideas of §2 and the above numerical simulations support the conclusions that: (i) the $\hat{v}$ boundary layer diffuses into the interior at a rate predicted by the slow-diffusion equation, eventually bringing $\hat{v}$ in the interior to V_T, and (ii) the cross-slope transport changes from M_0 to $\kappa \cot \alpha$ over time as $(t/\tau)^{-\frac{1}{2}}$.

4. Experimental set-up

Laboratory experiments were carried out on a rotating table at the CSIRO Marine Laboratories in Tasmania, Australia. Rotation speed was controlled to within ± 0.001 rad s^{-1}. The main tank (figure 5) used in these experiments was a Plexiglas section of a sphere. It varied from axial symmetry by no more than ± 0.25 cm. This tank will be referred to as 'the bowl'. For comparison we also did experiments in a right-circular cylindrical tank, 45.5 cm in radius and filled to the same depth as the bowl.

The tank was stratified with salt, hence density diffusion was almost non-existent over the course of an experiment. Typically the tank was stratified in five layers of increasingly salty water, from 0 to 140 parts per thousand, giving a nominal 10% density difference from top to bottom. The kinematic viscosity also varied by about 10% owing to the salinity, and account was taken of this variation in the analysis, §6. The layers were allowed to diffuse to a smooth profile overnight. Density was measured with a profiler that recorded conductivity, temperature and depth. The profiler had a spatial resolution of about 0.2 cm vertically. Typical profiles of density, ρ, and buoyancy frequency, N, are shown in figure 6. N goes to zero near the top and bottom owing to density diffusion. The profile shown is from the centre of the bowl. Profiles of N taken away from the centre do not go to zero near the lower floor except in a very small diffusive boundary layer. Previous experimenters have gone to great lengths (see for example Buzyna & Veronis 1971) to achieve a constant N-profile so that they could compare their experimental results with theory. For our experiments we shall make local comparisons of the flow with the theory developed in §2, so a local knowledge of N and α is sufficient. The centre of the bowl almost unavoidably had a small pool of low-N water at the bottom, and the dynamics there were generally very different from those at greater radius where N and α conspired rapidly to make buoyancy important in the boundary layer. Our comparisons are based on the assumption that the two regions did not interact significantly. This will be discussed further in §6.

For a typical experiment the fluid was stratified and allowed to spin-up overnight, with a lid on top to avoid air stress at the surface. The lid was 3 cm above the surface of the water. At $t = 0$ the rotation rate of the container was increased by an amount $\Delta\Omega$ (approximately 10%), over about 5 s. Initially the fluid was in solid-body rotation relative to the container, which was the new frame of reference, at an angular velocity $-\Delta\Omega$ (for a spin-up). The Rossby number, ϵ, of the flow is given by $\epsilon = \Delta\Omega/\Omega$. For small ϵ forcing along the boundary at lengthscale L will penetrate a depth H_P into the fluid, where $H_P = fL/N$. For our experiments f/N was around 1/3 in the body of the fluid, whereas the aspect ratio, depth/radius, of our containers was around 1/4. So we expected that the effects of boundary forcing would extend to the surface of the fluid, although showing some noticeable attentuation by then.

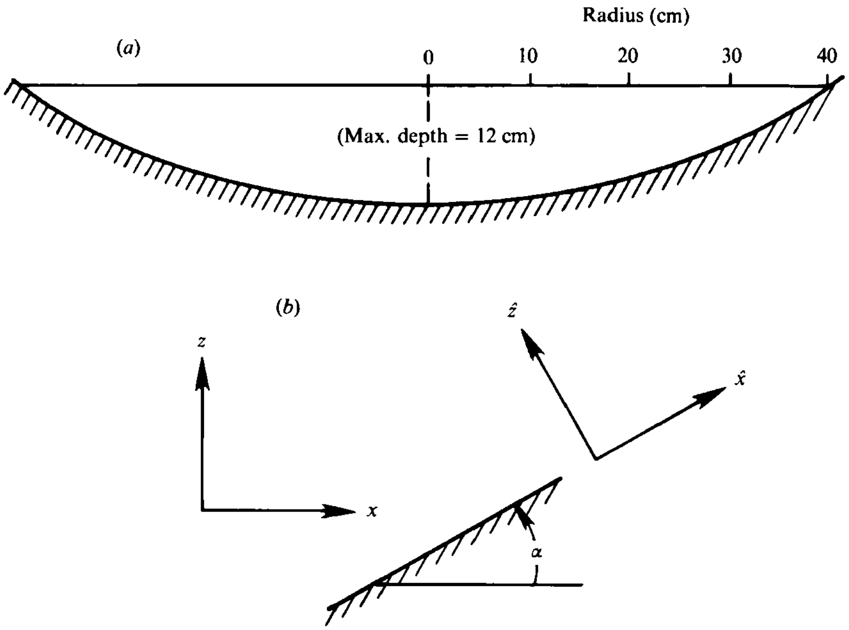


FIGURE 5. (a) Cross-section of the 'bowl' used in the laboratory experiments, and (b) a definition sketch of the slope coordinate frame.

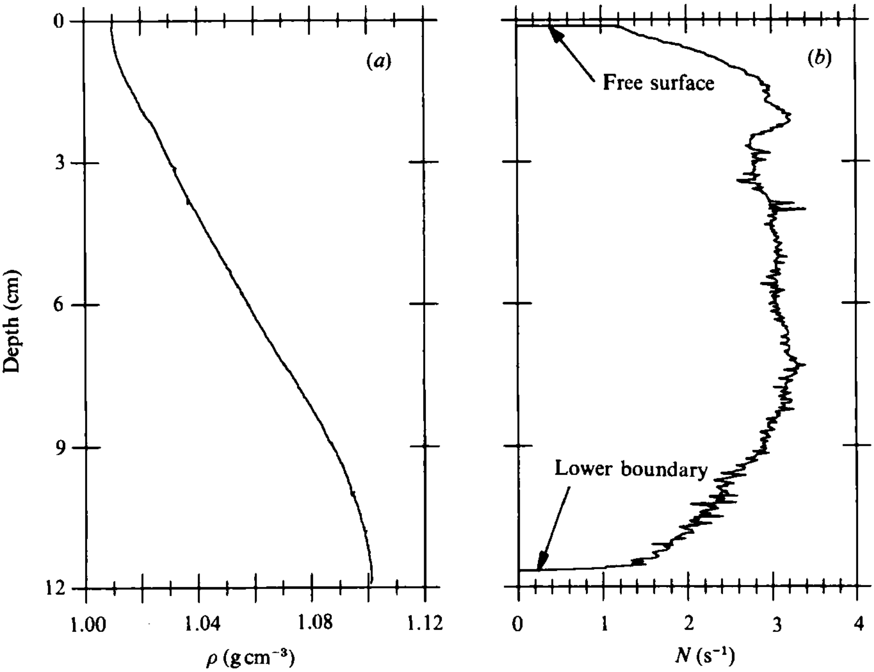


FIGURE 6. Measured profiles of (a) density and (b) buoyancy frequency versus depth, at the centre of the bowl, for a typical laboratory experiment.

Various flow visualization techniques, described in §5, allowed us to track the fluid as it eventually spun-up to the new rotation rate of the container. Density profiles were taken before and after experiments, and in general were nearly indistinguishable.

Although we present the laboratory experiments as support for the theoretical ideas of §2, in fact the laboratory work was done first, and served as a guide for the theory, which was developed later.

5. Experimental results

We used three types of flow visualization to explore spin-up in the stratified bowl and cylinder: (i) placing dye in the boundary layer to show the direction of stress at the wall, (ii) following beads floating on an isopycnal to determine zonal velocities (zonal is defined as along a circumference), and (iii) placing dye in the body of the fluid to show the shear history of the flow.

5.1. *Boundary-layer visualization*

To visualize flow in the boundary layer, just before a spin-up experiment a number of potassium permanganate crystals were dropped onto the bottom of the tank, along a radius. These exude a thick purple dye showing the direction of flow just above the bottom, and hence the direction of stress at the boundary. Figure 7(*a*, *b*) shows these dye streaks during an experiment in the cylinder, and figure 7(*c*, *d*) shows dye streaks for a similar experiment in the bowl. Both containers had similar stratifications.

Flow in the boundary layer of the cylinder remained very much like an Ekman layer, moving out and downstream at 45° to the local radius. This behaviour continued even as the interior flow was decreased by the fast spin-up. The bowl exhibited very different behaviour. Flow in the outer half of the tank, where α is greater and the shut-down time was very short, soon became mostly zonal, suggesting that the Ekman boundary layer was shut-down by the buoyancy forces on the sloping bowl wall. For this experiment τ was less than 5 s everywhere outside of $r = 15$ cm.

5.2. *Flow velocity measurements*

In order to measure zonal velocity we made long-exposure photos of particles ('beads') moving along density surfaces, similar to the technique used by Holton (1965). The beads were made of candle-wax and crayon, which could be combined in any ratio to conform even to the rather high densities near the bottom of our stratifications. A heated mixture of the two was sucked into a large syringe and then dotted out onto a flat surface to solidify in drops about 2 mm in diameter. The advantage of these beads is that they may be used for larger velocities than the thymol-blue technique (Buzyna & Veronis 1971). Also they give information on an entire density surface, instead of at just one point as in laser-Doppler velocimetry. The disadvantages are that it is difficult to make beads of perfectly consistent density, and the data reduction is somewhat time consuming. In a given experiment we had up to 60 beads floating along one density surface, with a vertical scatter in their positions of up to ± 0.5 cm. This led to some scatter in the velocity profiles in regions of strong vertical shear, and was the largest source of error in the data. Degassed water was used to avoid bubble formation on the bead surfaces.

With a group of light-coloured beads floating along an isopycnal, the room was

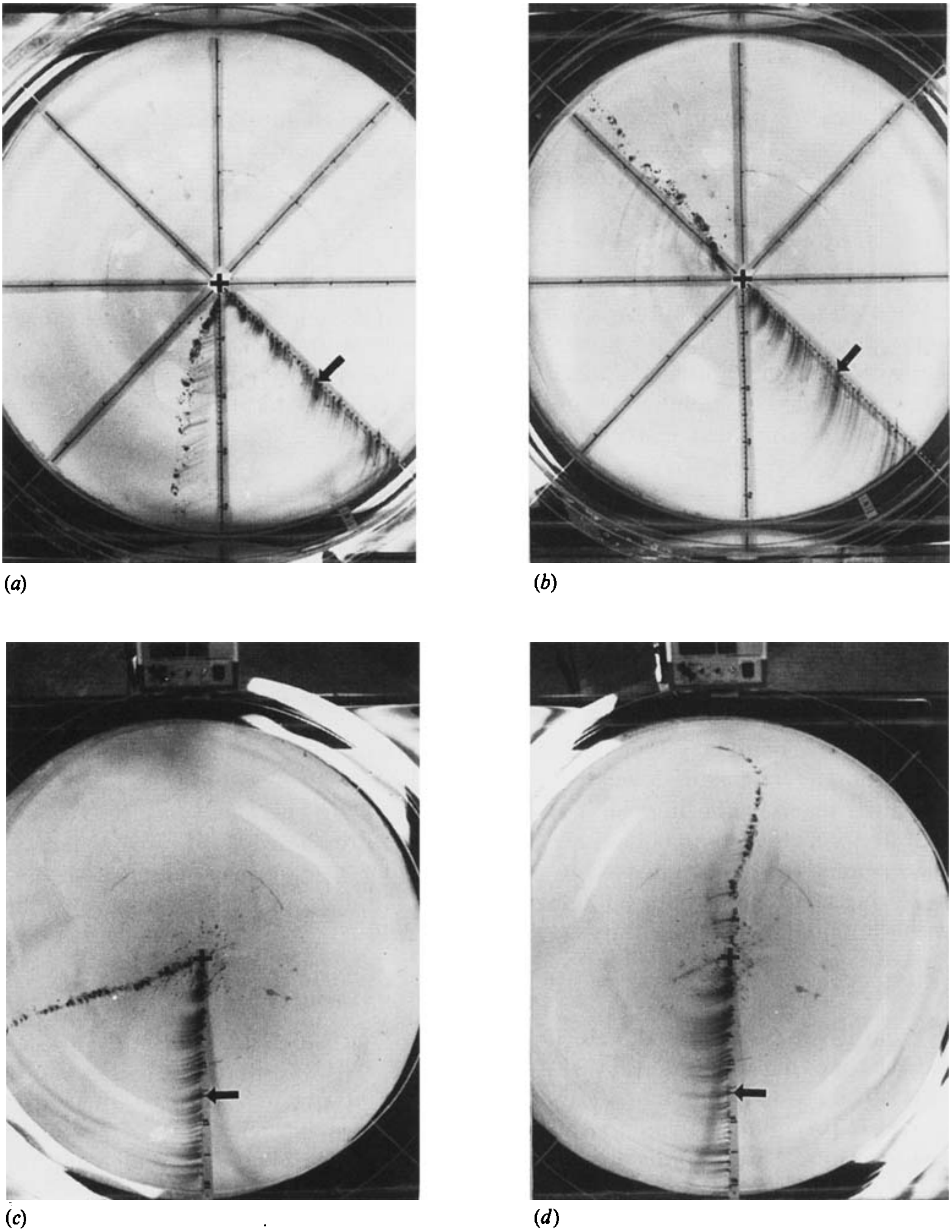


FIGURE 7. Photos of potassium permanganate in the boundary layer of two stratified spin-up experiments. The upper photos are looking down on the cylinder at (*a*) $t = 10$ s, and (*b*) $t = 40$ s. The lower photos show the bowl at (*c*) $t = 10$ s, and (*d*) $t = 40$ s. Crosses mark the centres, and the arrows point at $r = 25$ cm along the radius where the dye crystals lay. Both experiments were spun-up 15 % from $f = 1\ \mathrm{s}^{-1}$. Radially outward flow persisted in the cylinder, yet it quickly disappeared in the bowl, especially at larger radius. Note the shape of the surface velocity profiles, indicated by the line of dye from very small crystals which dissolved immediately as they entered the water.

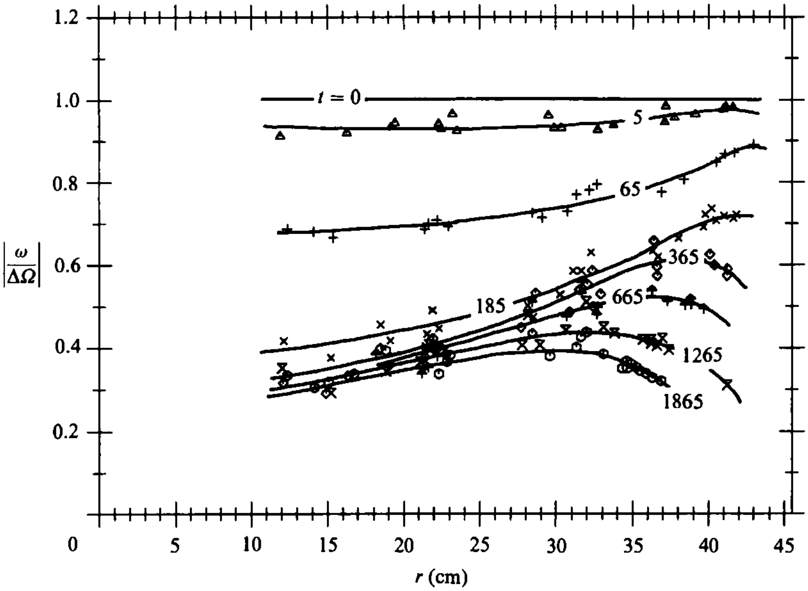


FIGURE 8. Normalized angular velocity versus radius, r, for a stratified spin-up experiment in the cylinder at a depth of 7 cm from the top (total depth was 11.6 cm), at eight different times (in seconds), as listed. The container was spun-up 15 % from $f = 1.1\ \mathrm{s}^{-1}$. The solid lines are hand-drawn fits to the data. The fluid rapidly spun-up in the central region, which could be reached by the meridional circulation initially emanating from the outer corner.

darkened and a spin-up experiment initiated. The beads were lit by a strobe light, and were photographed from above in the rotating frame of reference. Each exposure spanned 10–20 flashes of the strobe, hence each bead would appear in a photo as a series of dots along an arc about the centre of the tank. Each photo gives a profile of angular velocity as a function of radius, at a given depth and time (the time was approximated as being midway through a photograph).

Figure 8 shows the time-history of the angular velocity, ω, for a 15 % spin-up experiment in the stratified cylinder. The $t = 0$ line was drawn using the known $\Delta\Omega$ of the experiment. Within three minutes most of the fluid had been spun-up to about 55 % of the new rotation rate. This was the signature of 'fast' spin-up in the cylinder, caused by vortex stretching driven by Ekman transport in the bottom boundary layer. Only over much longer times did the viscous boundary layer from the side and bottom walls begin to complete the spin-up to the new rotation rate. This region of 'fast' spin-up is qualitatively in agreement with previous theoretical and experimental work such as Walin (1969), and Buzyna & Veronis (1971). Notice the jet-like region near the outer boundary, which occurs in a region 'missed' by the meridional circulation.

Although there is some vertical displacement of density surfaces during 'fast' spin-up owing to vortex stretching, the density surfaces slump back almost to their initial positions over the longer viscous timescale. There is a tiny shift in isopycnals in the long term to fit the geopotential paraboloid of the new rotation rate, but this requires no vortex stretching. Ultimately, fluid particles in the interior achieve the new rotation rate because their potential vorticity has been altered by viscous stress.

Figure 9 shows a similar angular velocity history for a spin-up at two levels in the stratified bowl, starting from $f = 0.66\ \mathrm{s}^{-1}$. In this experiment there was little, if any,

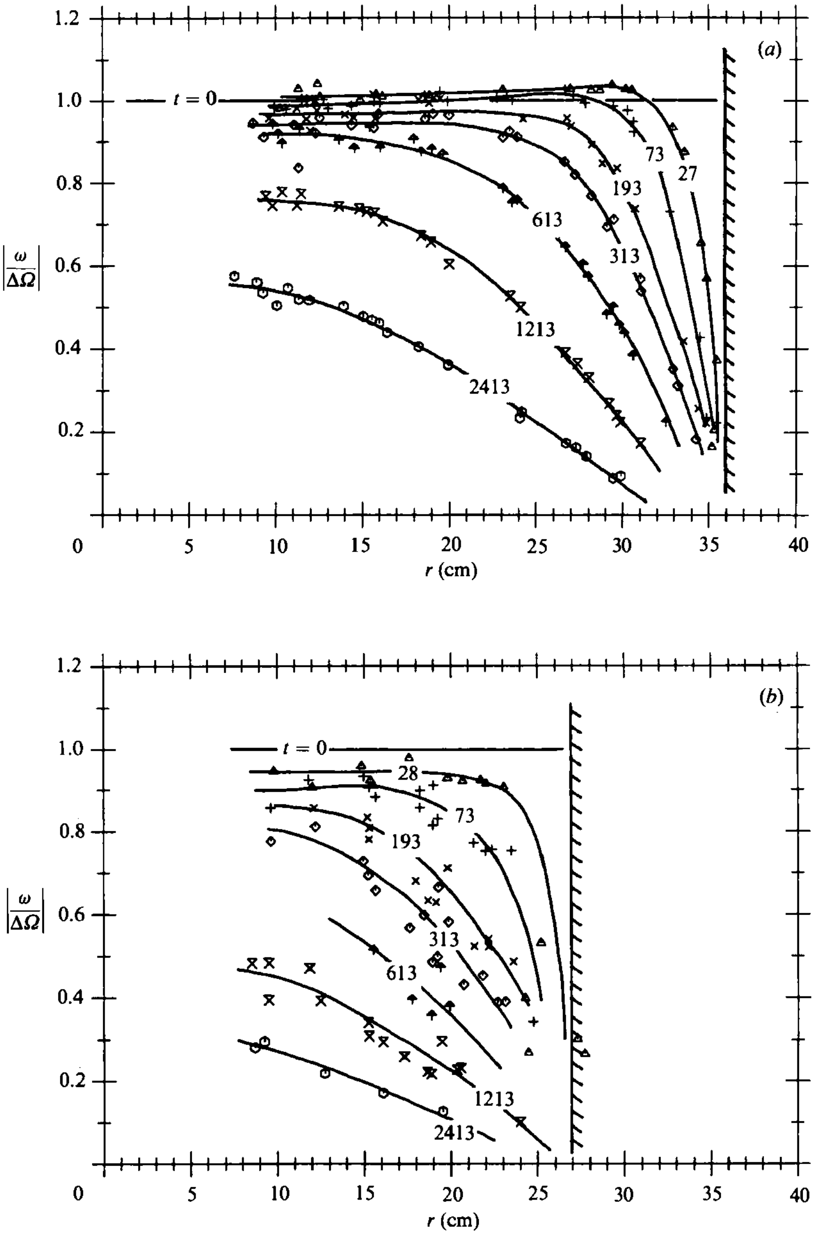


FIGURE 9. Normalized angular velocity versus radius for two, essentially identical, stratified spin-up experiments in the bowl at depths of (*a*) 3.4 cm from the top, and (*b*) 7.45 cm from the top (total depth at the centre was 11.7 cm), at eight different times (in seconds), as listed. The container was spun-up 10% from $f = 0.66\ \mathrm{s}^{-1}$. The edge of the tank at the given depth is marked with cross-hatching. The fluid adjusted slowly to the new rotation rate, and did so mainly by diffusion inward from the bottom. The early profiles in (*a*) indicate that a slight zonal circulation was present at the start of the experiment.

evidence of a 'fast' spin-up such as in the stratified cylinder. Presumably much of the Ekman layer had been shut-down by buoyancy forces, so the meridional circulation which causes 'fast' spin-up never developed.

Figure 10 is for a case similar to figure 9, except that $f = 1.5\ \mathrm{s}^{-1}$. In this experiment

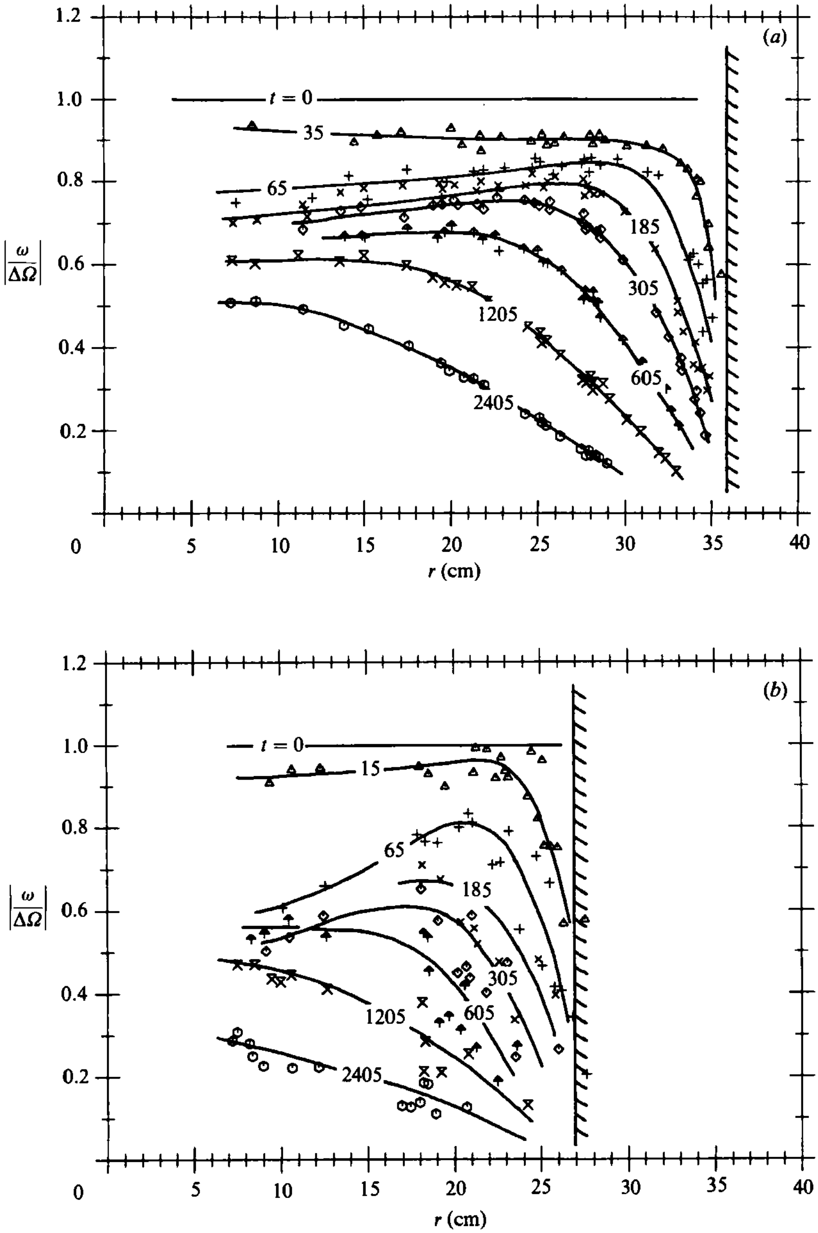


FIGURE 10. Same as figure 9 except the bowl was spun-up from $f = 1.5\ \mathrm{s}^{-1}$. There was clear evidence of some 'fast' spin-up in the acceleration of the inner region over the first minute. As in the cylinder, a zonal differential jet occurred in the corner where the meridional circulation had not reached.

there was some evidence of 'fast' spin-up. Recall that the shut-down time (2.28) increases with increasing f. In this case it appeared that the Ekman layer was operative long enough to affect the inner 15 cm of the flow, over the first minute. This issue will be taken up more quantitatively in §6.

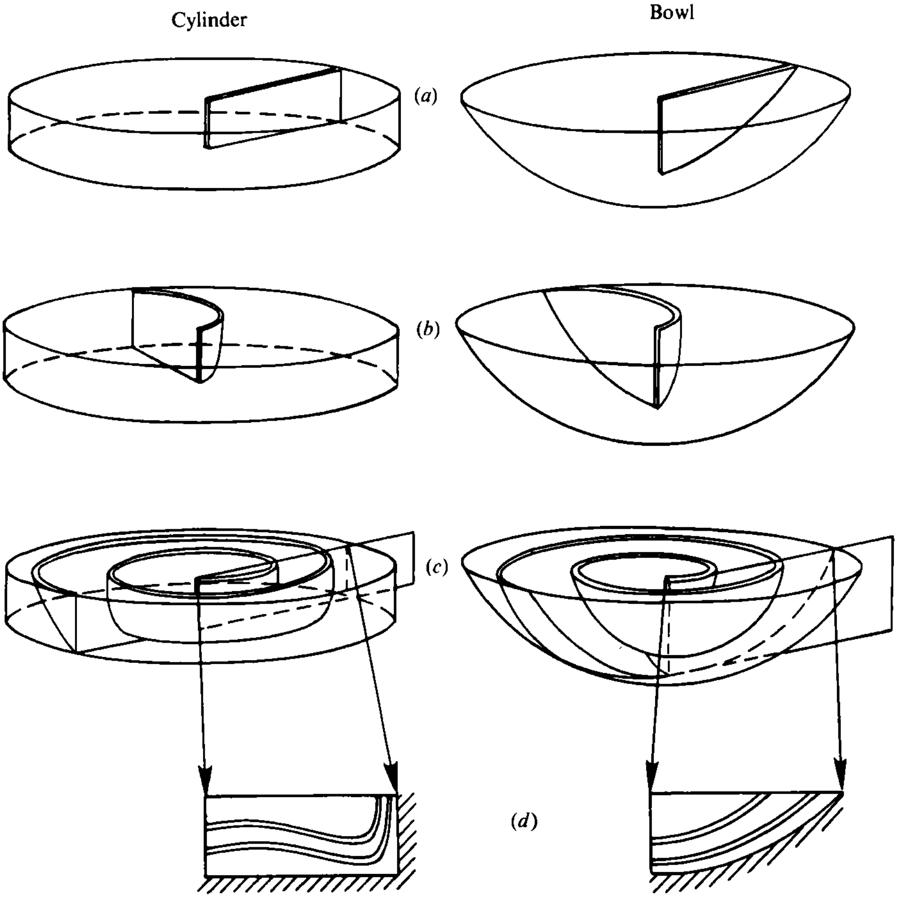
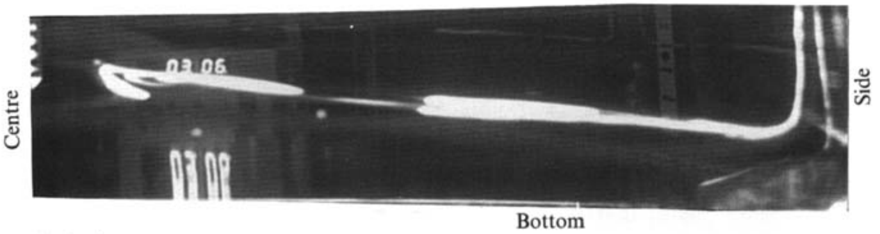


FIGURE 11. Onion-slice flow visualization in the stratified cylinder or bowl. Just before a spin-up experiment a sheet of dye (*a*) was injected into the fluid from top to bottom along a radius. As the spin-up started, (*b*), the dye sheet was stretched around the tank. As the spin-up proceeded, (*c*), the dye sheet was wrapped up into a spiral. A slice (*d*) through the dye sheet revealed lines of constant angular displacement. Fluid particles on adjacent dye lines in the slice had gone around the tank a complete revolution relative to each other.

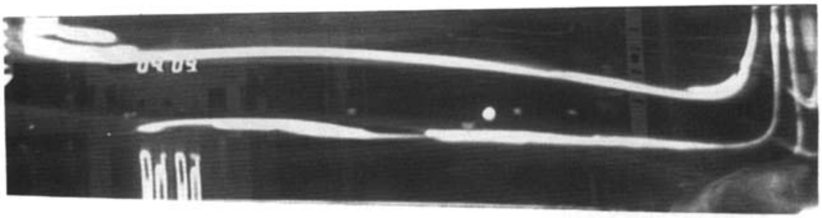
5.3. *Onion-slice flow visualization*

One final method was used to see the 'shear history' of an experiment. A vertical plane of fluorescein dye was introduced with a 'rake' of injection tubes. Following figure 11, one can see how this dye plane is deformed by the sheared zonal velocity field. Slicing through this structure we find the dye lying along lines of constant angular displacement, with fluid particles on adjacent lines having travelled around the tank one revolution relative to each other. The slice was made by shining a sheet of light vertically down through the tank, so the dye lines appeared white on a dark background. We call this flow visualization technique an 'onion-slice' because of its appearance, especially in the stratified bowl. In a photo of an onion-slice experiment, regions where there has been a radial or vertical gradient of the zonal velocity show up with dye lines normal to that gradient. The more shear there has been, the more closely spaced these lines are.

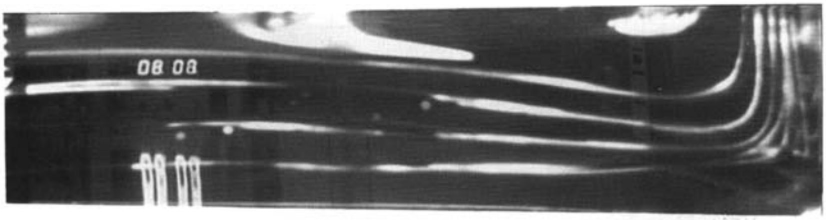
Figure 12 shows onion-slice photos from a spin-up experiment in the stratified



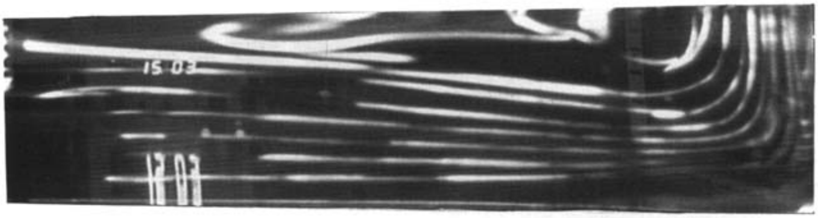
(*a*) 3 min



(*b*) 4 min

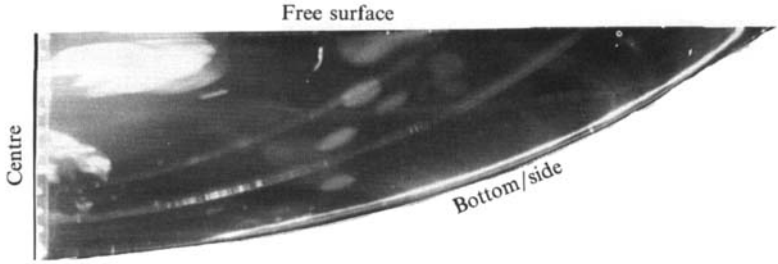


(*c*) 8 min

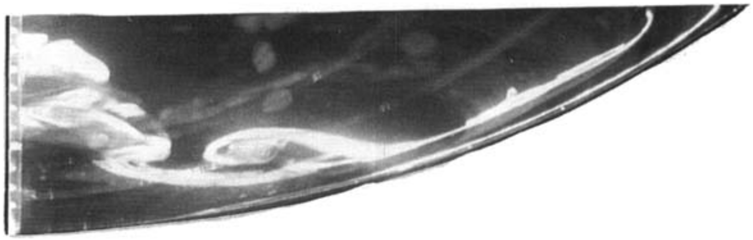


(*d*) 15 min

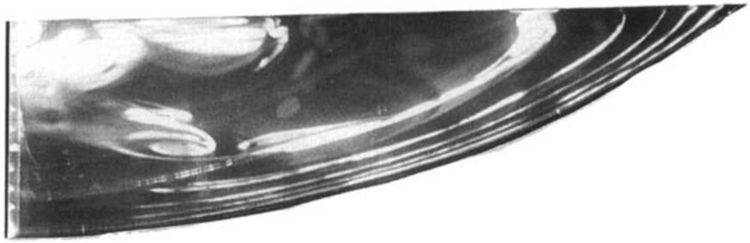
FIGURE 12. Onion-slice photos of a stratified spin-up experiment in the cylinder at four times, as listed. The container was spun-up 15% from $f = 1\ \mathrm{s}^{-1}$, and had a stratification similar to that shown in figure 6. The horizontal dye sheets bulged upward in the centre owing to the broadly distributed vertical shear of the zonal flow during 'fast' spin-up, while the sidewall boundary layer formed vertical sheets.



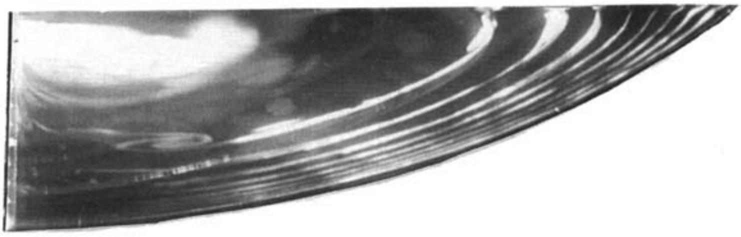
(*a*) 1.5 min



(*b*) 2.5 min

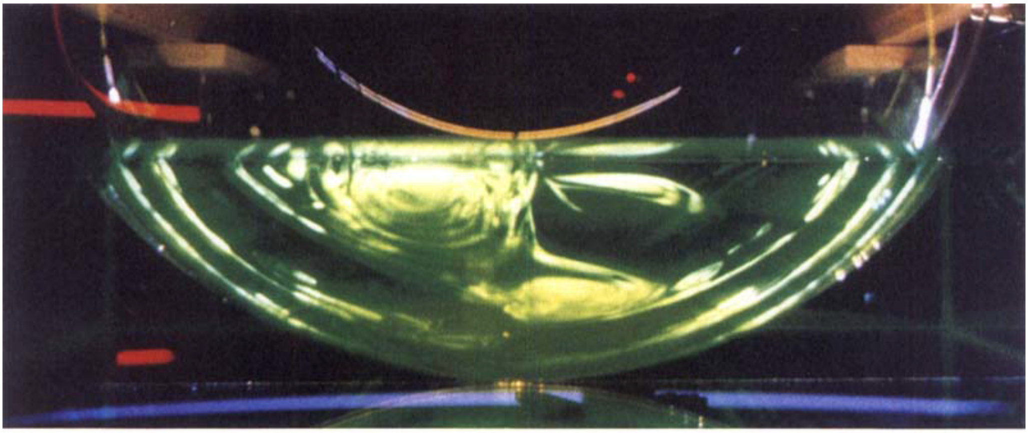


(*c*) 7 min



(*d*) 13 min

FIGURE 13. Same as figure 12, except in the bowl. Note that in (*a*) there was only one actual dye line, which lay close to the edge of bowl; the fainter lines were reflections. The adjusting flow was dominated by diffusive penetration normal to the boundary.


(*a*)

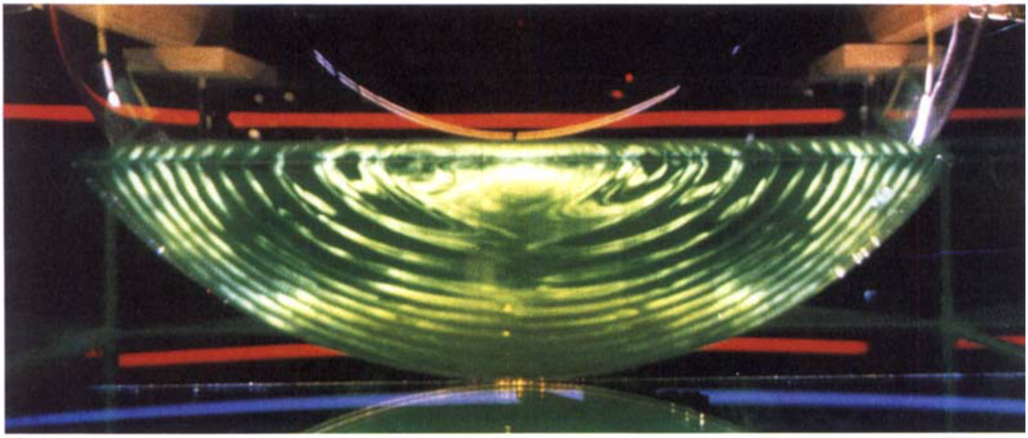

(*b*)

FIGURE 14. Colour photos of a stratified spin-up experiment in a hemispherical container, showing the complete fluorescein signal, (*a*) early in the experiment, and (*b*) at a later time. In general, the dye injected was initially rather messy, but the shear in the zonal velocity field sharpened the gradients of the dye, leading eventually to well-defined lines in the light sheet.

cylinder, and figure 13 shows photos from an experiment in the stratified bowl, graphically demonstrating the different nature of the flow in the two containers. In the first two photos of the cylinder (figure 12*a*, *b*) the widely spaced dye lines nearly filled the body of the fluid, expressing the penetration of the 'fast' spin-up throughout most of the tank. Later (figure 12*d*) the shear had continued fairly evenly through the depth of the fluid. The dye surfaces bowed upwards from the bottom centre where 'fast' spin-up was most intense, while a zonal differential jet near the outer wall represented fluid that was only slowly spun-up. The diffusive inward penetration near the outer wall was reflected in the vertical strike of the dye lines there. Fluid which had been transported radially outward in the bottom Ekman layer could be seen as a wedge in the corner of the cylinder. This wedge was the finite-Rossby-number expression of the circumferential forcing of the interior by the boundary layer flow.

Experiments in the bowl, figure 13 and figure 14 (plate 1), showed remarkably smooth quasi-diffusion of momentum inward from the sloping bottom. Early in an experiment, figure 13(*a*), just a single dye sheet was visible near the boundary. The inner core of fluid had made about one revolution relative to the boundary. At this rotation rate ($f = 1\ \mathrm{s}^{-1}$), intermediate between the two cases shown in figures 9 and 10, there was little sign of significant 'fast' classical spin-up, which would cause an upward bowing of the deeper dye sheets. Regions free of dye lines represent unsheared fluid, rotating at nearly their original angular velocity. In this core there was a slight 'fast' spin-up even in the cases with rapidly shut-down boundary layers (e.g. the inner core of figure 9(*b*) experiences weak spin-up without waiting for diffusion from the boundary).

Looking at spin-up in a variety of containers we found a consistent tendency for the dye sheets to mimic the shape of the lower boundary. Figure 12 represents perhaps the most extreme violation of this result, in the case of an extensive horizontal bottom above which the Ekman layer can continue to flow.

6. Comparison of theory with laboratory experiments

The theoretical ideas of shut-down and slow diffusion developed in §2 may be useful in understanding the experimental results if we assume that the experimental boundary layer developed in accordance with its *local* flow parameters. This approximation is commonly made for geophysical boundary layers when the scale of variation of the relevant parameters, such as N and V, is much greater than the boundary-layer thickness. Since N and V typically vary in our experiments only over the depth and radius of the bowl, we expect our local theory to be valid until the boundary layer has 'slow diffused' a good portion of the depth into the fluid. If significant fast spin-up has occurred then we would have to take account of the space–time structure of V in applying the theory.

Figure 15 shows the shut-down time τ versus radius, for the two bowl experiments. In the small-f experiment there was almost no sign of fast spin-up, particularly for the shallow beads. The brevity of the shut-down time implied that there would be essentially no radial boundary-layer transport, and hence no fast spin-up outside $r = 10$ cm. Fast spin-up inside that radius would penetrate less than one-third of the full depth at the centre, barely affecting the deeper beads. Thus the small-f swirl velocity data appear consistent with the interpretation that the boundary-layer transport was strongly limited by the shut-down mechanism.

In the large-f experiment there was evidence of some 'fast' spin-up, although not

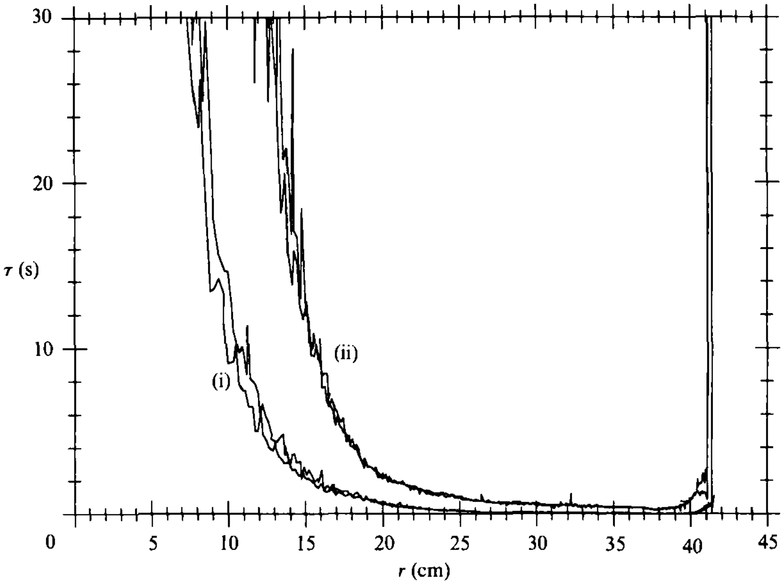


FIGURE 15. Shut-down time at the lower boundary versus radius for two bowl spin-up experiments: (i) refers to the experiment whose velocity data are shown in figure 9 (small-f), and (ii) refers to the experiment whose velocity data are shown in figure 10 (large-f). There are two profiles for each experiment since an experiment had to be performed twice in order to measure velocity at the two different depths.

nearly as much as in the cylinder, nor for as long. In the cylinder 'fast' spin-up occurred for about 3 min, whereas in the bowl with large-f it only lasted about 1 min. The shut-down time implied little transport outside of $r = 15$ cm. Owing to the increased rotation rate, the e-folding height of the region of fast spin-up would extend through about two-thirds of the depth of the fluid, affecting both deep and shallow beads. The large-f data are thus also qualitatively consistent with the shut-down hypothesis.

Assuming that the small-f experiment was, over most of its radius, unaffected by 'fast' spin-up, we may compare the angular velocity with that predicted by the slow-diffusion equation. Figure 16 shows angular velocity versus time at four locations in the small-f experiment. Also shown are predicted slow-diffusion solutions based on the value of $\nu S/(1+S)$ at the boundary nearest to the position in question. The three comparisons at larger radius are fairly close, indicating that slow diffusion was probably a good description of the situation. An alternative hypothesis is that the time rate-of-change of the zonal velocity was due to meridional circulation driven by whatever boundary-layer transport was present, and not slow diffusion. Without knowing the actual boundary-layer transport it is difficult to prove or disprove this idea, and we may say only that the data are consistent with slow diffusion. The innermost comparison (figure 16*d*) diverged strongly from the slow-diffusion prediction. This was probably due to the effects of 'fast' spin-up near the centre. Similar data for the large-f experiment (not shown here) also diverged strongly from the slow-diffusion prediction, again presumably due to the effects of 'fast' spin-up.

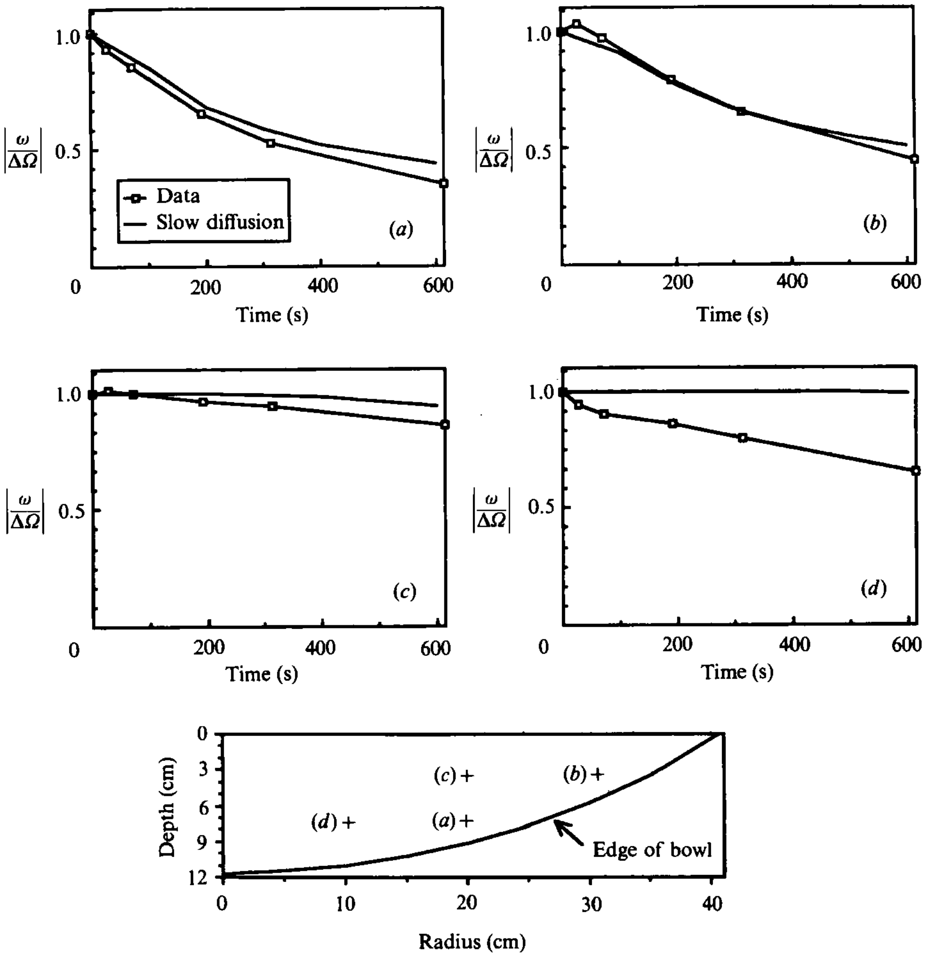


FIGURE 16. Normalized angular velocity versus time, at four different locations during the small-f (refer to figure 9) stratified spin-up experiment in the bowl, compared with velocities predicted by local solutions to the slow diffusion equation (2.20). The four locations are shown schematically at the bottom.

7. Conclusion

Our analysis of the boundary layer for rotating, stratified flow along a slope led to a circulation markedly different from that of an Ekman layer. It was neither steady, nor was it simply related to the interior velocity. Instead, the velocity profiles, and the cross-slope transport, were in general highly time-dependent. Buoyancy forces due to advection and diffusion of the stratification modified the force balance, tending to resist up- or down-slope velocities.

Very soon after its initiation the flow began to resemble an Ekman layer, but at large t/τ the flow looked like the steady solution (2.13) of Thorpe (1987), which allows only one interior along-slope velocity, and only one cross-slope transport. These properties make Thorpe's solution difficult to apply to geophysical situations. Thorpe (1987) and Garrett (1990) addressed this difficulty by modifying the steady theory to include depth-varying profiles of viscosity and diffusivity, which could be interpreted as saying that a given internal geostrophic flow exerts control over the magnitude of turbulent diffusion near the boundary. We have taken a different approach, trying to make the boundary-layer representation more realistic by allowing time dependence. In doing so we found that there was a whole continuum of solutions between those of Ekman and Thorpe.

The 'slow diffusion' equation (2.20) predicted that the boundary condition on the along-slope velocity would diffuse into the interior, and the numerical simulations of §3 implied that the eventual boundary condition for this equation was the interior velocity of Thorpe's solution (which vanishes as $\kappa \rightarrow 0$). What is of interest about slow diffusion is that it actually penetrates the interior much like a non-rotating viscous boundary layer, in contrast to both Ekman and Thorpe's solutions which remain confined close to the boundary.

In §2 we also predicted the temporal evolution of the cross-slope transport. With the aid of numerical solutions to determine an unknown constant, we found that the transport changed smoothly from the initial Ekman-layer value to the final value of Thorpe's solution over a timescale τ (2.28), the 'shut-down time'.

The shut-down time gives a sensible way to evaluate what sort of boundary-layer theory we should be using for a given situation. If the flow is varying much more rapidly than the shut-down time, say owing to the effects of 'fast' spin-up, then standard Ekman theory is a good approximation. If we are only interested in the flow long after the shut-down time, and only over a region where slow diffusion will have accomplished its work without being countermanded by other circulations or buoyancy sources, then Thorpe's steady solution is appropriate. Yet anywhere between these two limits we must necessarily be aware of the unsteady nature of the boundary layer.

The stratified spin-up experiments in a bowl with a sloping bottom boundary described in §4 and 5 occupy a regime where slow diffusion and the shut-down time were very important. Over much of the bowl with small-f the shut-down time was so short that almost no up-slope transport was allowed, and hence almost no 'fast' spin-up occurred. In these cases the main decrease in zonal velocity was apparently due to slow diffusion. In the bowl spin-up experiments the large-scale dynamics were completely altered by the buoyancy modification of the boundary layer. Although the bowl was everywhere shallower than a cylinder wherein comparison experiments were done, the fluid in the bowl spun-up much more slowly than that in the cylinder. Thus, by varying the geometry of the experiment we ended up with a boundary that effectively had significantly lower drag than that of the cylinder.

If the interior circulation is naturally oscillatory (as, for example, with a Rossby wave), these boundary-layer dynamics imply a frequency-dependent 'bottom friction' which is strong only at high frequencies, where the flow changes direction before shut-down can occur.

We may make a simple estimate of the along-slope stress that the boundary exerts upon the fluid using the equations developed in §2. Taking $\hat{z}$-integrals of the rotated equations of motion (2.7) with boundary conditions (2.8) and solving for the boundary stress in terms of the integrated cross-slope transport M (again with vanishingly small R_t and E_s) we find

$$-\nu\rho_0\frac{\partial\hat{v}}{\partial\hat{z}} = \rho_0 f\cos\alpha\,[M(1+S)-S\kappa\cot\alpha] \quad \text{at} \quad \hat{z}=0, \tag{7.1}$$

or

$$-\nu\rho_0\frac{\partial\hat{v}}{\partial\hat{z}} \approx M\rho_0 f\cos\alpha \quad \text{at} \quad \hat{z}=0 \quad \text{for} \quad S\ll 1. \tag{7.2}$$

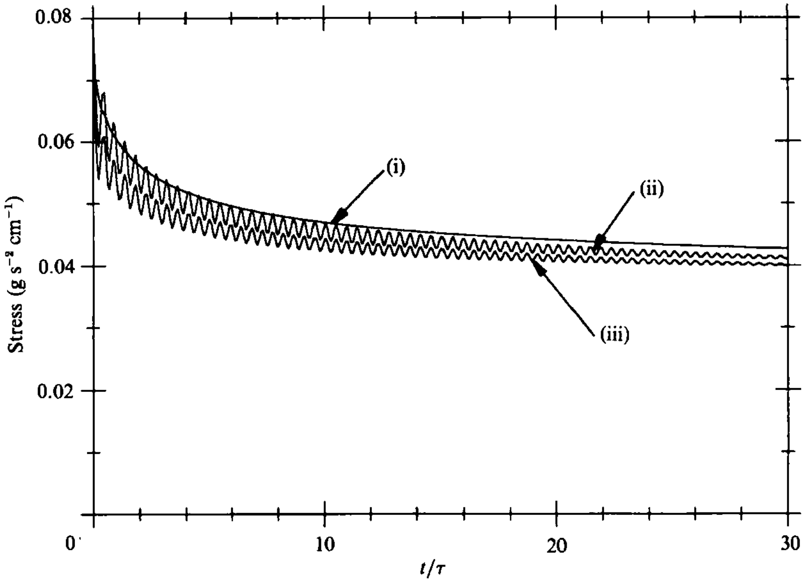


FIGURE 17. Along-slope stress at $\hat{z} = 0$ from a numerical solution to (2.7) with all parameters as in figure 3(*a*). The actual stress, $-\nu\rho_0(\partial\hat{v}/\partial\hat{z})$, is shown by curve (i), while curve (ii) shows the stress estimated from the right-hand side of (7.1), and curve (iii) shows the stress estimated from the right-hand side of (7.2).

The last expression (7.2) is exact for both the Ekman layer and Thorpe's solution, simply expressing that the boundary stress is balanced by the Coriolis force on the cross-slope transport. For the time-dependent boundary layer the boundary stress is balanced both by the Coriolis force and the acceleration of the along-slope flow. Equation (7.2) tells us that, for S small (which we expect for small slopes), the acceleration term is negligible and the stress is tied to the transport, exactly as in the steady theories. One may question the validity of (7.2) on the grounds that it is an estimate of stress in a region where E_s is likely to be large, violating our assumptions. Yet since we are using $\hat{z}$-integrals of the equations it turns out that the scale estimate of the $\hat{z}$-integrated viscous term in (2.7*a*), divided by the $\hat{z}$-integrated Coriolis term, is $(U/V)(E_s^{\frac{1}{2}})(\delta_s/D_v)$. While the first two parts of this scale estimate may be $O(1)$ when $\sigma = 1$, the third term becomes small as the along-slope boundary layer 'slow-diffuses' into the interior. Figure 17 shows the along-slope stress at $\hat{z} = 0$ versus t/τ for a numerical run with $\sigma = 1$ and $S = 0.29$ (all parameters as in figure 3*a*). We also plot the right-hand sides of (7.1) and (7.2). The figure demonstrates that, even in a case with strong density diffusion and S not far from 1, the boundary stress is approximately equal to the Coriolis force on the cross-slope transport. This balance is ironic when we consider that the slow-diffusion equation implies that boundary stress balances along-slope acceleration when $\sigma = 1$. But from the numerical simulations (§3) we know that the slow-diffusion equation gives a poor representation of boundary stress when $\sigma = 1$. As $\sigma \rightarrow \infty$ the slow-diffusion solution was nearly perfect at the boundary and indeed, in this case its $\hat{z}$-integral is exactly equal to (7.1). For the stratified spin-up experiments (7.1) shows that the torque exerted by the bowl upon the fluid dropped off rapidly as the up-slope transport was shut-down by buoyancy forces.

The boundary-layer theory developed in §2 is only applicable if the assumptions that went into its development are reasonably satisfied. We assumed constant viscosity, diffusivity, and stratification, we ignored variations along or across the slope, and we ignored the overturning of statically unstable stratifications. To make the theory more geophysically relevant the most important direction to pursue would be the inclusion of variable viscosity and diffusivity, as Thorpe (1987) and Garrett (1990) have done for the steady theory. We hope to pursue this avenue in future work.

In the atmosphere, the low thermal mass of air, combined with radiative means of changing temperature, make a conducting boundary condition on the density realistic. This boundary condition allows relatively simple Ekman-like boundary-layer solutions (Holton 1967 and Hsueh 1969). In contrast, the ocean floor is best modelled as an insulating boundary, except in the few areas of geothermal heating. Hence it is in the ocean that we expect our theory could be used. There are practical difficulties, though, with applying any theory of the oceanic bottom boundary layer. Data are very scarce, and determining turbulent coefficients of viscosity and diffusivity is inexact. It is generally accepted that in a turbulent mixed layer near the boundary the Prandtl number may be close to one. Yet stratified turbulence may easily send its energy into internal wave motion, which can transport momentum far from the boundary, leading to either a very high Prandtl number or the abandonment of Fickian diffusion entirely.

As a lowest-level approximation, let us assume that the ocean has a constant vertical eddy diffusivity, K, given by Munk's (1966) canonical value of 1 $\text{cm}^2\,\text{s}^{-1}$. Thorpe's steady transport, $K \cot \alpha$, is then simply a function of the bottom slope. Note that the transport becomes infinite as the slope goes to zero, and so is certainly unphysical over flat areas (the interior along-slope velocity also goes to infinity in this case). Consider two regions in the ocean: a continental slope with relatively strong stratification, and an abyssal region with less slope and a small stratification. Over the sloping region, with $\tan \alpha = 10^{-2}$ and $N = 3.5 \times 10^{-3}\,\text{s}^{-1}$, Thorpe's transport becomes very small, having up-slope velocities of only 0.1 cm s^{-1} if distributed over a 10 m thickness. The shut-down time indicates that the transport begins to approach this small value in about a week. Over an abyssal plane, with $\tan \alpha = 10^{-3}$ and $N = 10^{-3}\,\text{s}^{-1}$, Thorpe's transport is large enough to be important, 1 cm s^{-1} if distributed over a 10 m thickness. Yet this value is only achieved over the shut-down time, which is 3×10^4 years. These are only rough estimates, which do not account for variable viscosity and diffusivity. They do indicate, however, that shut-down may be very important on continental slopes, while standard Ekman theory is useful over an abyssal plane. A yardstick to judge the possible importance of shut-down is that the e-folding timescale of ocean currents due to 'fast' spin-up is typically taken as one year in numerical models. Any shut-down time shorter than this is likely to influence the evolution of flow along the ocean's bottom boundary.

In general, the shut-down time and its supporting theory presented in §2 should help in determining when and how buoyancy becomes important to the boundary-layer force balance in a given situation.

This work was supported by National Science Foundation grant no. 8613725, and Office of Naval Research grant no. N00014-86-K-0690. The authors wish to thank Chris Garrett for his continued interest in this work, and in particular for his helpful insight about the slow-diffusion equation. We also wish to thank Trevor McDougall, Angus McEwan, the laboratory group at CSIRO, Hobart, Tasmania, as well as Eric Semtner. Portions of this work were previously reported at the 'Aha Huliko'a Hawaii Winter Workshop, 1989, and at the American Geophysical Union Fall 1989 Meeting.

REFERENCES

Benton, E. R. & Clark, A. 1974 Spin-up. *Ann. Rev. Fluid Mech.* **6**, 257–280.

Buzyna, G. & Veronis, G. 1971 Spin-up of a stratified fluid: theory and experiment. *J. Fluid Mech.* **50**, 579–608.

Garrett, C. 1982 On spindown in the ocean interior. *J. Phys. Oceanogr.* **12**, 989–993.

Garrett, C. 1990 The role of secondary circulation in boundary mixing. *J. Geophys. Res.* **95**, 3181–3188.

Gill, A. E. 1981 Homogeneous intrusions in a rotating stratified fluid. *J. Fluid Mech.* **103**, 275–295.

Greenspan, H. P. & Howard, L. N. 1963 On a time-dependent motion of a rotating fluid. *J. Fluid Mech.* **17**, 385–404.

Holton, J. R. 1965 The influence of viscous boundary layers on transient motions in a stratified rotating fluid: Part I. *J. Atmos. Sci.* **22**, 402–411.

Holton, J. R. 1967 The diurnal boundary layer wind oscillation above sloping terrain. *Tellus* **19**, 199–205.

Hsueh, Y. 1969 Buoyant Ekman layer. *Phys. Fluids.* **12**, 1757–1762.

Munk, W. H. 1966 Abyssal recipes. *Deep-Sea Res.* **13**, 707–730.

Pedlosky, J. 1987 *Geophysical Fluid Dynamics.* Springer.

Phillips, O. M. 1970 On flows induced by diffusion in a stably stratified fluid. *Deep-Sea Res.* **17**, 435–443.

Phillips, O. M., Shyu, J.-H. & Salmun, H. 1986 An experiment on boundary mixing: mean circulation and transport rates. *J. Fluid Mech.* **173**, 473–499.

Siegmann, W. L. 1971 The spin-down of rotating stratified fluids. *J. Fluid Mech.* **47**, 689–711.

St-Maurice, J.-P. & Veronis, G. 1975 A multi-scaling analysis of the spin-up problem. *J. Fluid Mech.* **68**, 417–445.

Thorpe, S. A. 1987 Current and temperature variability on the continental slope. *Phil. Trans. R. Soc. Lond.* A**323**, 471–517.

Walin, G. 1969 Some aspects of time-dependent motion of a stratified rotating fluid. *J. Fluid Mech.* **36**, 289–307.

Weatherly, G. L. & Martin, P. J. 1978 On the structure and dynamics of the oceanic bottom boundary layer. *J. Phys. Oceanogr.* **8**, 557–570.

J. Fluid Mech. (1991), *vol.* 223, *pp.* 662–663
Printed in Great Britain

REVIEWS

Chaotic Dynamics: an Introduction. By G. L. BAKER and J. P. GOLLUB. Cambridge University Press, 1990. 182 pp. £25 (hardback) or £9.95 (paperback).

This excellent book develops the foundations of chaotic dynamics from the viewpoint of a physicist at a level that is accessible to the undergraduate student. It is comparable in intellectual demand with F. C. Moon's *Chaotic Vibrations* (Wiley, 1987) but requires much less background than that graduate-level text. The authors manage to cover almost all of the basic concepts and tools of chaotic dynamics – logistic maps, phase spaces and trajectories therein, Poincaré sections, time series and power spectra, basins of attraction, bifurcations (rather lightly), period-doubling cascades, fractal dimensions, and Lyapunov exponents – by focusing on the periodically excited, damped pendulum as a canonical example. They do include, in their final chapter, applications to fluid dynamics, chemical reactions, lasers, quantum mechanics, and statistical mechanics, but these require more antecedent knowledge than the preceding chapters and presumably would require selection and expansion by the lecturer. They also include exercises for the student, an appendix on Runge–Kutta numerical integration of differential equations, and program listings (in True Basic language) for the exercises. The illustrations are frequent, simple and clear. A diskette for IBM-PC-compatible computers (a Macintosh version may now be available) with programs that can duplicate or vary the simulations discussed in the text is available from the authors for $12.

Baker & Gollub have, I believe, succeeded admirably in achieving their declared goal of making this important and fascinating subject accessible to undergraduates, and I have but minor complaints to offer. They allege that the only prerequisites for their text are linear differential equations, introductory physics (basic mechanics), and PC-level computing, but it appears to me that either the student requires some prior knowledge of phase-plane analysis and elementary bifurcation theory or the material in chapter two requires amplification and supplementation by the lecturer. The resonance curve (response versus drive) of the periodically driven pendulum, which I have usually found to be the most revealing of all plane portraits, is neither mentioned nor displayed. And, although it is difficult to overestimate the contributions of Poincaré to this subject, I believe that the authors' attributions (or absence thereof) may have succeeded in doing just that.

That *Chaotic Dynamics* is aimed primarily at undergraduates should not preclude its use at more advanced levels. The teacher of a graduate fluid-mechanics course who finds Guckenheimer & Holmes (*Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Springer, 1983) too intimidating or Moon (*l.c.a.*) too wide-ranging might well look to Baker & Gollub for the fundamentals of chaotic dynamics and proceed from there to discuss its implications for such problems as turbulence. Or, perhaps with even greater profit, the fluid-mechanics specialist might consider the challenges and rewards of offering an undergraduate course on one of the most exciting developments in classical physics in this century.

JOHN MILES

Theories of Fluid Flows Through Natural Rocks. By G. I. Barenblatt, V. M. Entov and V. M. Ryzhik. Kluwer Academic Publishers, 1990. 395 pp. £95 or $147.

This is an expanded and fully revised edition of the original Russian edition. Although the purist might wish to see a few grammatical and stylistic faults corrected, no reader will have any difficulty whatever in reading this text, and the authors are to be congratulated on mastering what to them is a foreign language.

Its contents are built around work done over a period of 30 years by the three authors and largely reported in the Soviet literature. In many cases they anticipated similar work done much later in the 'West', and in some their work is essentially unknown to the world at large. For this reason alone, it is extremely valuable to have it collected together in one volume. Most of the work refers to applications in the oil and gas production industry, though for good historical reasons some problems of groundwater flow are introduced (there is an exact parallel between certain gently sloping unconfined water flows and confined compressible gas flows).

The text stands well on its own and should be read and studied as a whole and not in parts. The further I got into the book, the better I appreciated it. It is not a compendium of isolated results obtained using different approaches, approximations and notations (as is sometimes the case with prolific authors), but a beautifully crafted and balanced work setting out in a rational, consistent and scholarly way both the continuum-mechanical models that have provided a successful description of testing and production flows in the oilfield, and a fairly complete set of basic analytic solutions to the main important initial and boundary-value problems that result from these models.

Those familiar with Barenblatt's other texts will expect to see powerful use made of dimensionless, asymptotic and self-similar techniques: they will not be disappointed. To an older generation such methods provide unique insight into the mechanics of the real processes involved, early recourse to computed solutions being no substitute for dimensional analysis, for perturbation and matching methods in deriving general results that can be used in analysing field problems. The level of detailed mathematics included is perfectly judged to make the book both readable and convincing. Those intending to carry out detailed investigations are given an excellent start, and younger workers may learn some new techniques.

There are six chapters, covering basic physical concepts, classical linear models of homogeneous fluid flow, classical nonlinear models of homogeneous fluid flow, non-classical flows (fractured porous or layered reservoirs, non-Newtonian homogeneous fluids, elasto-plastic porous media), two-phase flows and finally physico-chemical flows (those involved in enhanced oil recovery, involving heat or mass diffusion, and where relevant chemical reaction). Reservoir engineers will notice few if any important omissions, and should benefit not only from the references given to Russian papers but also from the informed selection made from the wider literature. They and chemical engineers will welcome the firm grasp displayed of practical aspects of the processes involved. Some will meet hodograph methods for the first time; others will be intrigued by the use of a Bingham model for certain crudes; most will enjoy the crisp treatment of fingering instabilities, flow in inhomogeneous media, end effects and lag effects in reaching capillary equilibrium.

J. R. A. Pearson

INDEX