Journal of Marine Research, 69, 501–522, 2011

Submesoscale generation by boundaries

by W. K. Dewar[1], Pavel Berloff[2] and Andrew McC. Hogg[3]

ABSTRACT

An important dynamical question involves how oceanic balanced flows lose energy. Recent numerical and analytical studies suggest topography catalyzes energy exchanges between balanced flows and a variety of unbalanced phenomena, which presumably leads to dissipation. We here develop a general theory of inviscid balanced flow interactions with walls that predicts submesoscale and unbalanced flow generation. Comparison with primitive equation-based numerical experiments supports the basic tenets of the theory.

1. Introduction

The principal aim of modern global ocean modeling is the simulation of meso- and larger-scale motions, i.e. roughly anything from lateral scales of 10 km and larger and vertical scales of 100 m and more. There are many reasons for focusing on this subset of oceanic phenomena, with a major one being that it is these scales that connect most cleanly to climate. It is also recognized, however, that the complete modeling of such phenomena is impossible due to the computational weight of the problem. The dynamics of the unresolved, smaller scales are parameterized in terms of resolved dynamics in all current models. Amongst other things, these parameterizations control the dissipation of the mesoscale, although they lack a basis in dynamics. This provides the motivation behind the present work; namely, we hypothesize that boundaries play an important role in the dissipation of the mesoscale and develop here a relatively general analytical theory of such inviscid interactions. The theory predicts excitation of the submesoscale through topographic catalyzation and argues for routine development of internal modes of convection. Comparisons with numerical experiments support the basic tenets of the theory.

a. Background

Energy is introduced into the oceans by means of wind, the tides and buoyancy forcing (Ferrari and Wunsch, 2008; Hughes *et al.*, 2009). High frequency, inertial winds and tides

1. Earth, Ocean and Atmospheric Science, Florida State University, Tallahassee, Florida, 32306, U.S.A. ***email: dewar@ocean.fsu.edu***

2. Department of Mathematics and Grantham Institute for Climate Change, Imperial College London, London SW7 2AZ, United Kingdom.

3. Research School of Earth Sciences, Australian National University, Canberra, Australia.

can directly excite high frequency oceanic motions and, it is thought, eventually effect oceanic mixing. Lower frequency winds and surface buoyancy forcing energize the large-scale balanced flows of the major oceanic gyres and the Antarctic Circumpolar Current. The rates at which balanced flows receive energy have recently been estimated. While the error bars are large, a value of roughly 1 TeraWatt ($1\,\mathrm{TW} = 10^{12}\,\mathrm{W}$; Oort *et al.*, 1994; Ferrari and Wunsch, 2008 and references therein) is generally accepted as the wind power input to the ocean. Input from surface buoyancy forcing is less well constrained but Oort *et al.* (1994) estimate that $\sim$1 TW is injected into available potential energy. The energy so received manifests mainly at the largest of the oceanic scales. These currents in turn energize the smaller oceanic mesoscale field through geostrophic instabilities. Add to this a small amount of direct forcing of the mesoscale by higher frequency variable wind and buoyant forcing, and approximately 1-2 TW enters the eddy field. The state of current ocean modeling is that the eddy field is partially included or well represented in most global scale ocean simulations.

A reasonable assumption is that the mesoscale field has not accelerated appreciably over the last several decades of ocean observation, implying the mesoscale loses energy at roughly the rate it receives it. Interestingly, how the mesoscale loses energy is not clear and has been the subject of much recent work. What is clear is that direct viscous dissipation of the eddies is unimportant, due largely to the well-known upscale energy transfers associated with the nonlinear evolution of the so-called 'balanced'[4] flows (Charney, 1971). This inverse cascade is a result of the simultaneous conservation of energy and potential vorticity by the balanced flows, a class of flows to which mesoscale eddies belong. Boundary dissipation, essentially the direct rubbing of the balanced flows on the boundary, although apparently significant, also appears unable to provide all the needed dissipation (Ferrari and Wunsch, 2008; Sen *et al.*, 2008).

In contrast, the other class of flows, the 'unbalanced' flows, are not similarly constrained by potential vorticity and are thus free to develop the classical downscale energy transfers ultimately leading to molecular dissipation. Examples of unbalanced flows are internal gravity waves and the recently discovered submesoscale (Capet *et al.*, 2008 a,b). Almost by elimination, it appears energy transfers to unbalanced motions are required to effect a balanced energy budget for balanced flow.

The near surface zone appears to be one area where the mesoscale can excite smaller scale and higher frequency flows, leading to balanced energy loss (Boccaletti *et al.*, 2007; Capet *et al.*, 2008a,b). D'Asaro *et al.* (2011) argued for the presence of surface unbalanced instabilities catalyzed by the wind in the separated Kuroshio jet. The inferred, apparently associated, dissipation rates were elevated by at least an order of magnitude beyond background rates. Another possibility is that balanced flows directly excite unbalanced flows through instabilities.

4. Balanced flows, to be defined more precisely later, are flows with diagnostic momentum equations, geostrophy being the best known example.

A somewhat unexplored, complementary possibility is that bottom boundaries catalyze transfers of energy from the balanced flow to unbalanced and/or submesoscale dynamics. Indeed, there have been some encouraging results here. Dewar and Hogg (2010; DH hereafter) have shown how eddies can resonantly pass energy to unbalanced motion at locations where the eddy flow arrests topographic waves. The resulting finite amplitude development drives mixing and overturning and generates considerable inertia gravity wave activity. Nikurashin and Ferrari (2010; NF hereafter), motivated by observations of Naviera-Garabato *et al.* (2004), computed the internal gravity wave generation forced by balanced flow over rough topography. They showed the dependence of the energy flux on the high wavenumber, O(1 km – 100 m), topographic spectrum and generated energy flux rates comparable to those observed in the ACC. Molemaker *et al.* (2012) in a downscaling study of the Monterey Bay region found extreme vorticity generation, with values of roughly $-10f$ (with f the local Coriolis parameter), driven by the flow of the California Undercurrent interacting with the bottom boundary layer. Flow separation at headlands was a typical event, ejecting the extreme vorticities into the interior, where they appeared to drive diapycnal mixing and momentum dissipation through centrifugal instability. In summary, topography provides rich opportunities for mesoscale flows to energize unbalanced and submesoscale dynamics, with estimated energy fluxes comparable to observed levels.

A common characteristic of these transfers is that they are small scale. The DH mechanism involved higher mode boundary waves with lateral scales of a few kilometers and NF emphasized the importance of extremely fine topographic roughness. Modern oceanic general circulation models operate at resolutions far in excess of these and their interactions with lateral boundaries and topographies are mediated by viscous parameterizations with tuned coefficients. Even so, the form of the subgridscale parameterizations, often a spatial smoothing, are motivated more by numerical requirements than by physical insights, underscoring their inadequacy. The problem is further complicated by uncertainty in what boundary conditions to apply: no-slip is required by molecular processes, but at coarse resolutions, free-slip or partial slip might be relevant.

Clearly there is a need to build physics into boundary parameterizations. The approach this paper takes is to exploit recent numerical results to propose a more mechanistically based set of boundary interactions leading to balanced flow dissipation. Specifically, we analyze the DH scenario using a set of surprisingly general proxy boundary equations. Their solutions offer explanations for and predictions of the dominant behavior in the primitive equation simulations.

b. A review of DH

The basic DH problem was the interaction of a subsurface coherent vortex with topography and the solutions were obtained through an implementation of the MITgcm (see DH for details). The parameters describing the vortex were taken from the well-known geophysical example of 'Meddies', i.e. eddies of Mediterranean origin that move at speeds of

O(.05 m/s) (Richardson *et al.*, 2000). The upper left panel of Figure 1 shows a Meddy-like dipole approaching a wall. This comes 24 days after initialization of the model, a time at which the vortex was very balanced in structure. The upper right panel shows the temperature at the wall at the same time, and indicates the vortex has not yet disturbed the wall structure. The early wall evolution of the vortex (see middle panels) was classical, i.e. the dipole splits with the anticyclone headed north and the cyclone south, as expected from the image effect. On the other hand, the wall temperature profile shows considerable distortion associated with the vortex, a result not expected from balanced dynamics alone. The Day 33 bottom left panel shows the creation of a strip of very high relative vorticity emanating from the north side of the anticyclone. The associated temperature structure along the wall also shows some dramatic structure and the development of small-scale structures presumably resulting from overturning. The effect of this on the vortex is clearly dissipative, both in the inherent dissipation involved in the overturning and in the creation of opposite signed relative vorticity compared to that of the vortex. This vorticity anomaly is subsequently injected into the interior and is entrained into the eddy. The calculations in DH used free-slip boundaries. However, very similar results have been found in the presence of 'drag-slip' boundary conditions, i.e. literally no-slip boundaries modeled using the Law of the Wall, (Deremble *et al.*, 2011).

These results were explained in DH using a quasigeostrophic approach as the result of the arresting of high mode Kelvin waves by the vortex flow field. Experiments with more realistic seamount topographies exhibited similar results, arguing the dynamics were not peculiar to walls. In summary, these numerical experiments suggest a potentially important role for unbalanced, boundary waves in balanced energetics. The goal of this paper is to clarify the dynamics of the wall interaction and it is hoped this understanding will eventually lead to a means for parameterizing it.

2. A theory of balanced flow-wall interaction

We seek here to develop as general a theory as possible describing the interactions of a balanced flow with a wall. Motivated by the numerical experiments, we begin with the f-plane hydrostatic equations of motion written in density coordinates

$$\begin{aligned} &u_t + uu_x + vu_y + Hu_\rho - fv = -M_x + X \\ &v_t + uv_x + vv_y + Hv_\rho + fu = -M_y + Y \\ &M_\rho = gz \\ &z_{\rho t} + (uz_\rho)_x + (vz_\rho)_y + (Hz_\rho)_\rho = 0 \end{aligned} \tag{1}$$

where ρ denotes density, $M + P + \rho gz$ is the Montgomery potential, subscripts denote differentiation and notation is standard. The quantity H represents the net diabatic effects

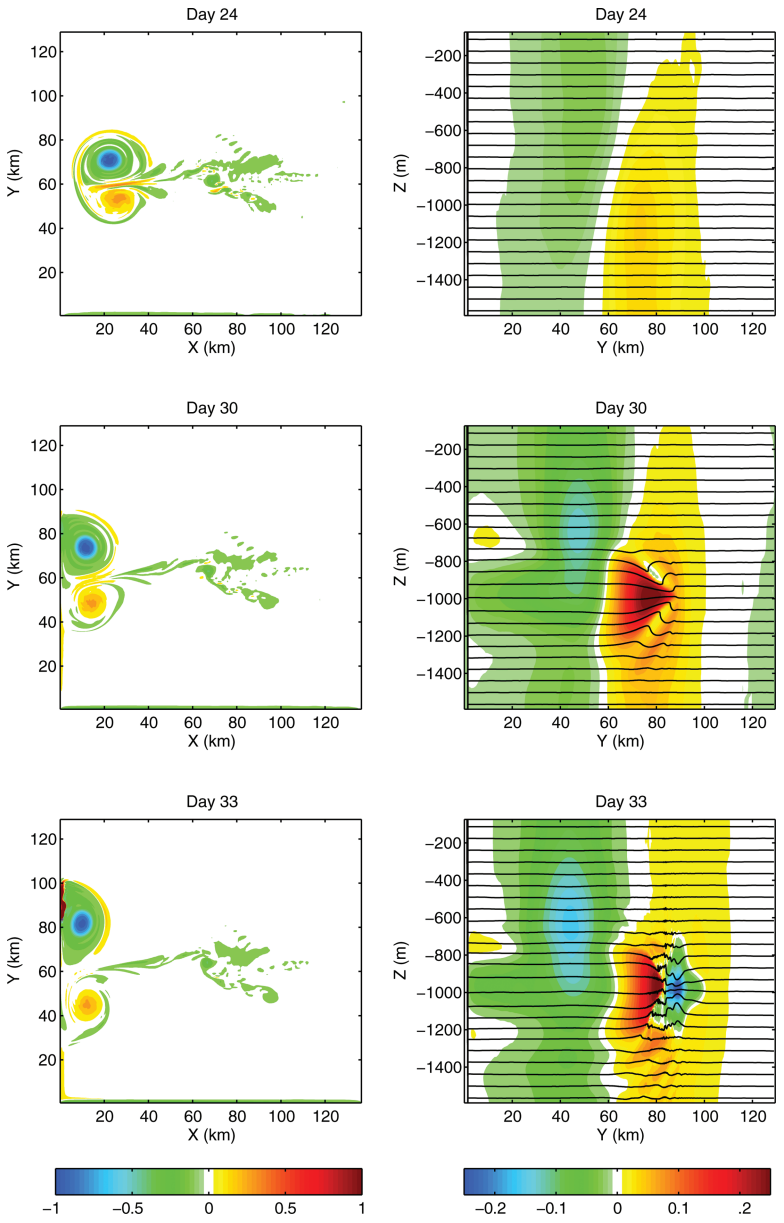

Figure 1. The left-hand panels are of normalized relative vorticity at a depth of roughly 970 m from the experiment discussed in Dewar and Hogg (2010). The right-hand panels are of temperature along the western wall. The top panels are prior to the vortex hitting the wall. The middle panels, showing evidence of front formation are early in the interaction and the bottom panels show a fully developed temperature front. Note the high relative vorticities streaming from the wall and the asymmetric development of the southward moving cyclone and the northward moving anticyclone.

operating in the fluid while X and Y denote net viscous effects. Manipulating these equations in the standard way eventually yields the potential vorticity equation

$$q_t + uq_x + vq_y - \frac{q(Hz_\rho)_\rho}{z_\rho} = \frac{Y_x - X_y - (v_\rho H)_x + (u_\rho H)_y}{z_\rho}; q = \frac{(f + v_x - u_y)}{z_\rho}. \quad (2)$$

These equations contain all fluid physics up to the hydrostatic approximation, inherent in the use of a density coordinate system and explicit in (1c).

For now, we neglect viscous and diabatic effects, but will eventually return to them. A north-south oriented wall imposes the kinematic condition of no-normal flow, here taken as $u = 0$. The momentum and potential vorticity equations then reduce to $fv = M_x$

$$\begin{aligned} &fv = M_x \\ &v_t + vv_y = -M_y \\ &q_t + vq_y = 0; q = (f + v_x)/z_\rho \end{aligned} \quad (3)$$

where it is recognized that the simplifications come from the wall constraint, making the equations accurate up to the hydrostatic assumption. An important result comes from the potential vorticity equation: it is seen that the wall q field is preserved. It is only the potential vorticity on the wall that can ever be there; q in the far field cannot move onto the wall. Considering an initial value problem in which a distant balanced flow moves into a wall interaction, the wall potential vorticity is taken to be that of a stratified fluid at rest

$$q = \frac{fg}{\overline{M}_{\rho\rho}} \quad (4)$$

where $\overline{M}_{\rho\rho}$ represents the background stratification.

In contrast, the impinging balanced flow is characterized by nontrivial potential vorticity anomalies relative to (4). As is generally the case for balanced flows, the potential vorticity field can be inverted, given boundary conditions, to compute the dynamical fields associated with it. The simplest and best known example of this is quasigeostrophy, where potential vorticity anomalies q' are related to pressure via the linear elliptic operator

$$q' = \nabla^2 p + \frac{\partial}{\partial z}\frac{f^2}{N^2}\frac{\partial}{\partial z}p. \quad (5)$$

Solving (5) yields the pressure, from which density anomalies and velocities can be readily obtained. The solution depends of course on the boundary conditions, which for a wall, reduce to the requirement of constant pressure on the wall. More elaborate balanced theories are available that employ different boundary statements, but they are all consistent with the kinematic requirement of no-normal flow.

The process of inverting (5) is nonlocal, i.e. nontrivial potential vorticity anomalies generate nontrivial dynamical fields in their far fields. An important point however is that those

far-field dynamical variables are all characterized locally by zero potential vorticity anomalies. In the present problem, given (3c), the dynamical fields imprinted on the wall by the impinging balanced disturbance all possess a vanishing potential vorticity anomaly and are described by (4). We can thus decompose the along-wall velocity in (3b) into a balanced component, coming from the inversion on the interior potential vorticity and represented by v_g, and a wall response, v'

$$\begin{aligned} &f(v_g + v') = M_{gx} + M'_x \\ &v'_t + \left(\frac{v'^2}{2}\right)_y + (v_g + v')_y + M'_y = -v_{gt} - v_g v_{gy} - M_{gy} \end{aligned} \tag{6}$$

A comparable decomposition appears in Reznik and Sutyrin (2005) and Hogg *et al.* (2011) for a single-layer reduced gravity case. The right-hand side of (6b) can be claimed as known from a balanced theory, while (6a) demonstrates that the along-wall response v' is described by geostrophy.

The equations (6) apply on the boundary; we now argue that they also describe the flow dynamics in the interior near to the boundary, subject to one relatively unrestrictive approximation. The total potential vorticity in the fluid is in general composed of two nontrivial elements: the background $\overline{q} = \frac{fg}{\overline{M}_{\rho\rho}}$[5] and the balanced field, whose potential vorticity will for convenience be denoted gq_g, thus

$$\frac{v_x - u_y + f}{M_{\rho\rho}} = \frac{f}{\bar{M}_{\rho\rho}} + q_g. \tag{7}$$

Recognizing that

$$\frac{v_{gx} - u_{gy} + f}{\bar{M}_{\rho\rho} + M_{g\rho\rho}} = \frac{f}{\bar{M}_{\rho\rho}} + q_g \tag{8}$$

it is possible to write the exact solution of (2) for the wall response, i.e.

$$v'_x - u'_y = \left(\frac{f}{\overline{M}_{\rho\rho}} + q_g\right) M'_{\rho\rho}. \tag{9}$$

This generalizes (6) to the wall-driven response occurring near the wall. Note, (9) reduces without approximation to

$$v'_x = \frac{f}{\overline{M}_{\rho\rho}} M'_{\rho\rho} \tag{10}$$

5. This definition of potential vorticity renders it negative in the Northern Hemisphere as the measure of the mean state stratification, $\bar{M}_{\rho\rho}$, is negative.

on the wall, because normal flow vanishes and the balanced potential vorticity can never reach the wall.

To this point, the discussion has been accurate up to the hydrostatic balance. We now introduce the only substantive approximation made in this paper, i.e. we neglect quantities of $O(U/V)$ where U is the magnitude of the flow normal to the wall and V is the along-wall flow magnitude. This is sensible given the kinematic constraint of the wall and that there is nothing obvious in the equations, like higher order derivatives, capable of introducing fine scales into the velocities. Such an approximation is frequently employed in frontal studies (Hoskins and Bretherton, 1972), but is more immediate here given that the condition is enforced by a solid boundary. We have checked this in our numerical solutions and found it to accurately characterize them. Neglecting terms of this order allows us to approximate the interior momentum equations by (6) to order $\frac{U}{V} \ll 1$.

Dropping primes for notational convenience, the quantity v_x (formerly v_x') can be eliminated from the pair (9) and (6a), yielding

$$M_{xx} = f\left(\frac{f}{\bar{M}_{\rho\rho}} + q_g\right) M_{\rho\rho} + O\left(\frac{U}{V}\right). \tag{11}$$

The quantity $\overline{M}_{\rho\rho}$ is negative for stable stratifications while $\frac{f}{\bar{M}_{\rho\rho}} + q_g$ is of the same sign as $\frac{f}{\bar{M}_{\rho\rho}}$ for a symmetrically stable balanced flow[6], thus (11) is recognized as a linear elliptic equation for the Montgomery. The appearance of y is parametric at leading order, so (11) is fundamentally two-dimensional. It is interesting that this linear equation applies in spite of the amplitude of the wall response. Unique solutions of (11) require either Neumann or Dirichlet boundary conditions. The wall response at large x is required to decay while at the top and bottom the perturbation to the isopycnal depth is required to vanish. The problem is closed in that the set (6) is poised to deliver Neumann boundary data at the wall.

Interestingly, the wall model is now complete. Given information about v and M on the wall at some time, (6b) can be used to predict a new value for v, which via (6a) is the normal derivative of the response Montgomery. Eq. (11) can then be solved to yield the response Montgomery over the domain, and in particular on the wall. This allows the along-wall gradient of the Montgomery to be computed, thus leading to a new prediction for v. The sequence can be iterated indefinitely, although it will turn out that criteria exist on the solutions that limit the utility of the model in its present form. The approximation of anisotropic response along the wall is strongly enforced by the solid boundaries; we are tempted to think of these equations as a proxy for the hydrostatic equations near the wall.

This is a very simple proxy theory and yet is remarkably unrestricted parametrically. Given a statement about the projection of the balanced flow on a wall, the wall response can be computed without appealing to a full three-dimensional primitive equation model. We have tested the validity of the proxy theory by comparing its predictions against the results

6. which will be assumed here.

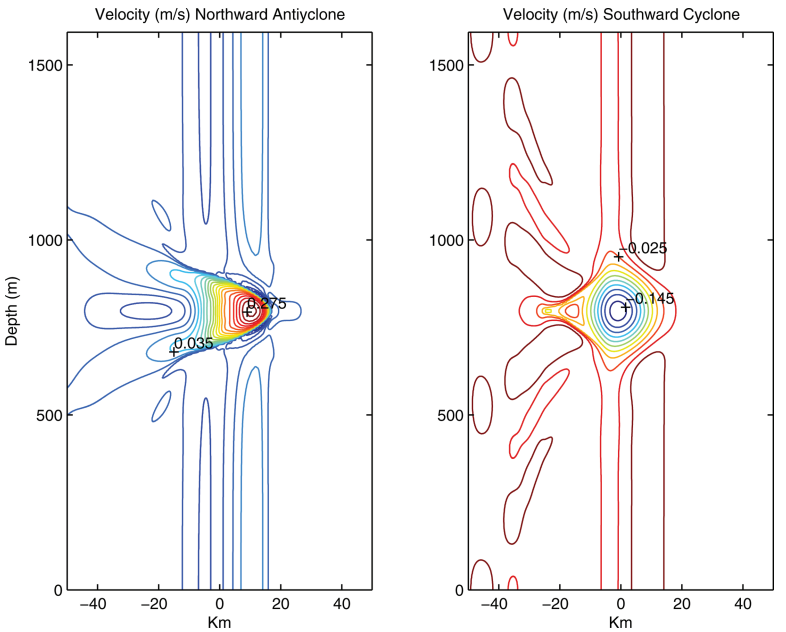

Figure 2. Asymmetric development of the wall response to anticyclones (left) and cyclones (right) computed with the proxy equations. Along-wall velocity is shown and both come at three days after initiation. The initial conditions were north-south symmetric and identical in both cases aside from a change in the sign of the perturbing velocity. Considerable distortion of the initial symmetry is seen for the anticyclone, with less occurring for the cyclone. The velocity anomaly has also amplified on the left relative to the right. The left maximum is.28 m/s and right minimum is $-.15$ m/s, the contour interval is 0.015 m/s in both. The asymmetry and amplification in velocities is characteristic of the MITgcm results in Figure 1.

of the MITgcm; an example coming from the DH experiment is given in Figure 2. The vortex potential vorticity, q_g, in those solutions. In effect, this is looking in the quasigeostrophic limit, which is appropriate for the vortex considered in DH. The most apparent aspect of the MITgcm solutions in Figure 1 is the asymmetry between the wall responses to cyclones and anticyclones. This is well captured by the simple model, as seen in Figure 2. It is also seen that the wall velocities of the anticyclone intensify in both the proxy and primitive equation numerical solutions. The predicted density surfaces (not shown) also steepen as in the MITgcm, even to the point of apparent inversions. The rates at which the wall anomalies develop are well predicted. It thus appears that the simple model qualitatively captures all aspects of the wall response from the MITgcm, and exhibits quantitative skill as well.

3. The dynamics of the wall response to a balanced flow

We now analyze the equation set (6) and (11) as a model of the full primitive equation dynamics of the wall response to a balanced flow perturbation. Of particular interest are

the questions of the source of cyclonic/anticyclonic asymmetry, the roots of the velocity intensification and the finite amplitude structure of the response.

a. Asymmetry

Consider first the linearized form of the equations in the absence of a balanced flow

$$\frac{M_{xt}}{f} + M_y = 0$$

$$M_{xx} - f^2 \frac{M_{\rho\rho}}{\overline{M}_{\rho\rho}} = 0. \tag{12}$$

Eq. (12 b) admits separable solutions of the form

$$M = X(x)R(\rho) \tag{13}$$

which in view of the boundary conditions

$$M_\rho = 0\rho = \rho_s, \rho_b \tag{14}$$

where ρ_z, ρ_b are the upper and lower boundary densities, yields the Sturm-Liouville eigenvalue problem

$$R_{n\rho\rho} - \frac{\lambda^2}{f^2}\overline{M}_{\rho\rho}R_n = 0 \tag{15}$$

where (R, λ_n) are an eigenfunction/eigenvalue pair. The offshore equation

$$X_{nxx} - \lambda_n^2 X_n = 0 \tag{16}$$

admits exponentially growing and decaying solutions from which the decaying solution is chosen to meet the far field boundary condition. Substitution in (12a) yields

$$-\frac{\lambda_n}{f}M_{nt} + M_y = 0 \tag{17}$$

which admits solutions of the form

$$M_n = F\left(y + \frac{f}{\lambda_n}t\right). \tag{18}$$

These are classic Kelvin waves moving south along the wall at speed

$$c_n = -\frac{f}{\lambda_n} \tag{19}$$

The theory of eigenvalue equations assures us that the wave speeds become arbitrarily small as the mode number grows. If now the balanced flow is readmitted into the linearized form

of (6b), the source of the asymmetry is clear. The linear Kelvin wave dynamics impart a southward propagation bias into the system, but the balanced flow can be of either sign. The asymmetry thus appears in solutions of the linearized form of (6) (not shown here).

Yavneh *et al.* (2001) noticed a similar interaction between a shear flow in a channel and Kelvin waves. In the right regimes, the counter-propagating Kelvin waves on opposing sides of the channel could be arrested by the shear flow, allowing a resonant instability between them.

b. Baroclinic balanced flow

Of more interest is the source of the amplitude growth in the wall response. The balanced velocity in general depends upon both ρ and y, but allowing for this dependence simultaneously has not yet lent itself to analytical progress. So, we consider balanced flows of each type independently. First, suppose $v_g = v_g(\rho)$ only, i.e. the flow is strongly baroclinic. Fourier transforming the linearized form of (6b) then yields

$$(v_g - c)\phi + \mu = 0 \tag{20}$$

where ϕ is the Fourier transform of v, μ is the Fourier transform of M and $c = \frac{\omega}{l}$, with ω being the temporal frequency and l the meridional wavenumber. Similar meridional transforms of (6a) and (11) imply

$$\begin{aligned} \mu_x &= f\phi \\ \mu_{xx} &= \frac{f^2}{\overline{M}_{\rho\rho}}\mu_{\rho\rho}. \end{aligned} \tag{21}$$

Taking an x derivative of (20) and using (21) yields

$$(v_g - c)^2\mu_{\rho\rho} - \overline{M}_{\rho\rho}\left(1 + \frac{v_{gx}}{f}\right)\mu = 0 \tag{22}$$

which, with the Fourier transformed boundary conditions

$$\mu_\rho = 0 \quad at \quad \rho = \rho_b, \rho_s \tag{23}$$

constitutes an eigenfunction/eigenvalue problem for μ, c.

We can now follow a fairly classical analysis to find an integral constraint on instabilities. Assume the existence of an unstable mode, so $(v_g - c)^2$ does not vanish on the real axis. Dividing (22) by $(v_g - c)^2$, multiplying by the complex conjugate of $\mu(\mu^*)$, and integrating over density eventually yields

$$\int |\mu_\rho|^2 d\rho + \int \frac{\overline{M}_{\rho\rho}}{(v_g - c)^2}\left(1 + \frac{v_{gx}}{f}\right)|\mu|^2 d\rho = 0 \tag{24}$$

where (23) have been used. Writing

$$\frac{1}{(v_g - c)^2} = \frac{(v_g - c)^{2*}}{|(v_g - c)^2|^2} \tag{25}$$

and requiring both the real and imaginary parts of the resulting integral to independently vanish returns the constraint on unstable modes that

$$\int \frac{(v_g - c_r)\overline{M}_{\rho\rho}\left(1 + \frac{v_{gx}}{f}\right)}{|(v_g - c)^2|^2} |\mu|^2 d\rho = 0 \tag{26}$$

where $c = c_r + ic_i$. Assuming the balanced flow is inertially stable, the phase speed of an unstable mode must thus lie in the range of the velocity field.

On the other hand, combining (20) and (21a) returns

$$(v_g - c)\frac{\mu_x}{f} + \mu = 0 \tag{27}$$

Eqs. (11) and (21b) are both elliptic equations that possess their extreme values on the boundaries. In view of the far field constraint that the wall response must decay, (27) implies $(v_g - c_r)$ must be positive definite for the solutions to be physically admissible. This contradicts the requirement for the existence of unstable modes in (26), thus showing the wall response to baroclinic flows is linearly stable. We have also computed the normal mode solutions to (22) for two representative velocity profiles and seen the lack of complex phase speeds (not shown here).

c. Barotropic balanced flow

We now consider the case of a balanced velocity depending only on y, i.e. a barotropic balanced flow. The linearized form of (6b) is

$$v_t + ((v_g - c)v)_y + M_y = -v_{gy}v + cv_{gy} - v_g v_{gy} - M_{gy} \tag{28}$$

where we have assumed the balanced flow is propagating steadily at speed c, such as would happen if a barotropic point vortex were interacting with the wall. The eigensolutions found from (15) are complete, so projecting (28) onto the R_n eigenfunctions eventually yields

$$M_{nt} + (v_g - c + c_n)M_{ny} = -v_{gy}M_n \tag{29}$$

where M_n is the projection of the Montgomery on the nth eigenfunction. At locations where $(v_g - c + c_n) = 0$, the solution of (29) is either exponential growth or decay and depends on the details of the balanced flow. Regions of convergence, where $v_{gy} < 0$, will promote unbounded growth, at least in this linearized analysis. This result is similar to one appearing in DH, although that argument depended upon quasigeostrophic scalings and the present analysis is much more general.

However, what was not noticed in DH is that cyclones can also sustain a comparable instability. Southward moving cyclones will also possess a range in which $(v_g - c + c_n) = 0$. The south side of a cyclone is convergent, and thus will host exponential growth at such locations. The range of velocities for which this will occur will be smaller than for the anticyclone and biased towards smaller c_n. This implies the previously noted asymmetry in behavior will also manifest in this aspect of unbounded growth. The Kelvin waves involved in the amplification will be of much higher mode. The numerical experiment in DH showed no unstable tendencies on the cyclonic side of the problem. The present analysis explains that result, in that higher modes are more readily damped by the model viscosity. Experimentation with the simple model supports both the existence of exponential growth for cyclones and anticyclones, as well as the much higher modes participant in the cyclonic side growth.

We note also that the unbounded growth seen here differs dynamically from the instability seen by Yavneh *et al.* (2001). Theirs depended on a resonance between Kelvin waves on opposing walls, and ours depends on along-wall flow structure. In contrast, they solved a formal eigenvalue problem to quantify their mechanism; we here cannot appeal to such an approach.

d. Finite amplitude behavior

It remains to clarify the role of nonlinearity in the wall response, especially in view of the exponential growth predicted by the linear analyses. A complication here is that the quadratic terms in (6b) have the effect of coupling all the vertical modes, so a general analysis is apparently not analytically tractable. However, if the setting is restricted to that of a reduced gravity model, progress is possible. This is equivalent mathematically to a one vertical mode truncation, but does correspond to the physical problem of a layered fluid. One relatively subtle aspect is that the density step associated with the interface must be chosen unusually small so as to correspond to a small deformation radius. This restriction differs from most reduced gravity models where the density steps are chosen so as to capture the gravest baroclinic modes, but allows us to examine the range where the advective effects of the balanced flow on the Kelvin wave are of leading order importance. The equations from here to (39) follow an analysis for single-layer reduced gravity flow appearing in Hogg *et al.* (2011).

Eq. (11) reduces to

$$M_{xx} - R^{-2}M = 0 \tag{30}$$

where the Montgomery is given by the more familiar form $M = g'h$, with h the layer thickness and R is the deformation radius. Choosing the decaying exponential to (30) allows the velocity to be related to the Montgomery via

$$v = \frac{-M}{fR}. \tag{31}$$

The momentum equation (6b) becomes

$$M_t + ((v_g - c)M)_y - MM_y/fR - fRM_y = -fR(cv_g - vg^2/2)_y \tag{32}$$

where we have again transformed to a frame fixed in an eddy moving at speed c. Eq. (32) is readily soluble by the method of characteristics, which transforms it to the set of coupled ordinary differential equations

$$\begin{aligned} t_s &= 1 \\ y_s &= \left(v_g - c - fR - \frac{M}{fR}\right) \\ M_s &= -v_{gy}M - fR\left(cv_g - \frac{v_g^2}{2}\right)_y . \end{aligned} \tag{33}$$

We will use the initial data $s = 0$, $r = 0$, $y = r$, $M = 0$, corresponding to a state of rest.

Analytical progress is aided using the linear balanced profile

$$\begin{aligned} v_g &= v_o\left(1 - \frac{|y|}{y_o}\right); |y| < y_o \\ v_g &= 0; |y| > y_o \end{aligned} \tag{34}$$

Staying within $|y| < y_o$, some algebra now reveals

$$\left(fR\frac{v_o}{y_o}|y| + M\right)_s = \mathrm{sgn}(y)(fR)^2\frac{v_o}{y_o} \tag{35}$$

so that

$$\frac{v_o}{y_o}(y - r) = -\mathrm{sgn}(y)\frac{M}{fR} + fR\frac{v_o}{y_o}s.$$

Substitution into (33b) then shows

$$y - r = -\mathrm{sgn}(y)\frac{v_o}{y_o}\frac{fR}{2}t^2 + \left[v_o - c - \mathrm{sgn}(y)\frac{v_o}{y_o}r - fR\right]t \tag{36}$$

Elimination of r permits the full solution for M to be written:

$$\begin{aligned} M = \mathrm{sgn}(y)\frac{v_o^2}{y_o^2}\frac{fR}{2}t^2 + \frac{v_o^2}{y_o^2}rt - \mathrm{sgn}(y)\frac{v_o^2}{y_o}t + \mathrm{sgn}(y)c\frac{v_o}{y_o}t \\ + \frac{(1 + \mathrm{sgn}(y))fRv_o}{y_o}t. \end{aligned} \tag{37}$$

This is not a global solution, of course, limited in space by the trajectories of the characteristics emanating from $y = |y_o|$. It also has the physical limitations that the characteristics cannot cross and the total layer thickness must remain non-negative.

Points of characteristic intersection denote a breakdown in the limited physics of the original model; these locations must receive special treatment. We can compute such trouble spots by asking at what time the characteristic locations lose sensitivity to their initial location r, i.e.:

$$y_r = 0 = 1 - \frac{v_o t}{y_o}\mathrm{sgn}(y). \tag{38}$$

Note that physically acceptable solutions only occur for positive y, where $t = \frac{y_o}{v_o}$. Substitution into (36) shows the characteristics for positive y have the unexpected property that they all arrive at the same meridional location

$$y = y_{crit} = y_o\left(1 + \frac{c}{v_o} - \frac{fR}{2v_o}\right) \tag{39}$$

at the above time. This is not fundamental, but rather reflects the special choice of velocity profile. The coarse structure of the characteristics up to the critical time is indicated in Figure 3.

A curious feature of the solution is that if M is evaluated at the critical time, its value for $o \le r \le y_{crit}$ is less than $-(fR)^2$. Since M is a perturbation to the mean state $\bar{M} = (fR)^2$, such values imply negative layer thicknesses. This is not permissible and the resolution is that the characteristics solution applies only up to the point of vanishing layer thickness. Some analysis shows the first characteristic for which this occurs is the one emanating from $r = 0$, and the layer thickness vanishes quickly after the initiation of the problem. In fact, all characteristics within the range $0 \le r \le r_{crit}$ yield vanishing layer thickness prior to the time of shock formation. Those in the range $y_{crit} < r < y_o$ retain finite layer thicknesses up to the time of shock formation and computing their subsequent evolution requires an extension of the analysis to include a shock joining theory.

The prediction of vanishing layer thickness is dramatic but there are reasons to believe analogs exist in the continuously stratified problem. The evidence for this comes from computing the Ertel potential vorticity from the MITgcm solution. An example located one half grid cell off the boundary appears in Figure 4. The structure of the Ertel potential vorticity consists of a largely uniform background state and a wedge of large negative potential vorticity centered on the leading edge of the anticyclone.

For the most part, the potential vorticity field is dominated by the uniform field associated initially with the undisturbed wall. The pronounced regions of extreme negative potential vorticity are associated with the locations where the fluid densities are overturning and, we argue, are reflections in this system of the tendency for the layer thicknesses to vanish in the reduced gravity system. Note that this boundary-generated potential vorticity is dissipative;

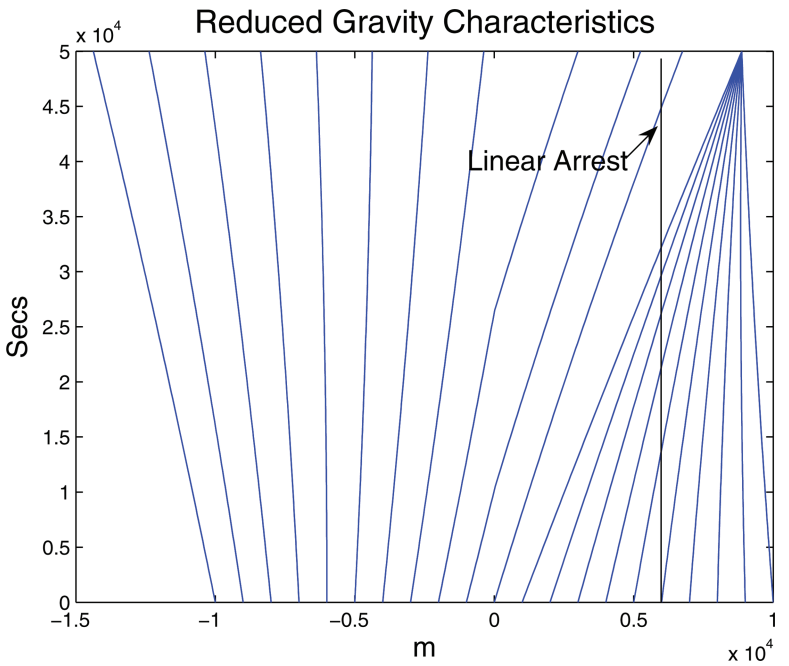

Figure 3. The characteristics structure of the proxy equations for the linearly varying anticyclone velocity field. Time increases up the y axis and location along the wall is on the x axis. Note the convergence of the characteristics on the northern side of the vortex, at a point to the north of the point of linear Kelvin wave arrest.

the anticyclone in this case is characterized by a positive potential vorticity anomaly[7], while the generated potential vorticity has a negative anomaly.

The reduced gravity model of course cannot overturn, given its reliance on hydrostatic dynamics and stable stratification, so we interpret the appearance of vanishing layer thickness as a prediction of overturning. While details of when, where and for what range the thickness vanishes in the reduced gravity setting almost certainly reflect the specific choice of velocity profile, we speculate the tendency for thickness to vanish as seen in the analytical case, or overturn as seen in the continuously stratified case, is a general characteristic of the behavior of this system.

Continuing the characteristics computations for later times shows information from both ahead of and behind the critical location continues to enter the region. This is reminiscent of both classical shock behavior (Whitham, 1974) and the behavior observed in the numerical solutions. The analysis mentioned above argues the structure of the shock consists of a vanishing thickness south of the shock and a finite thickness north of the shock. Clearly,

7. Recall that by our definition of potential vorticity, anticyclones have positive potential vorticity anomalies.

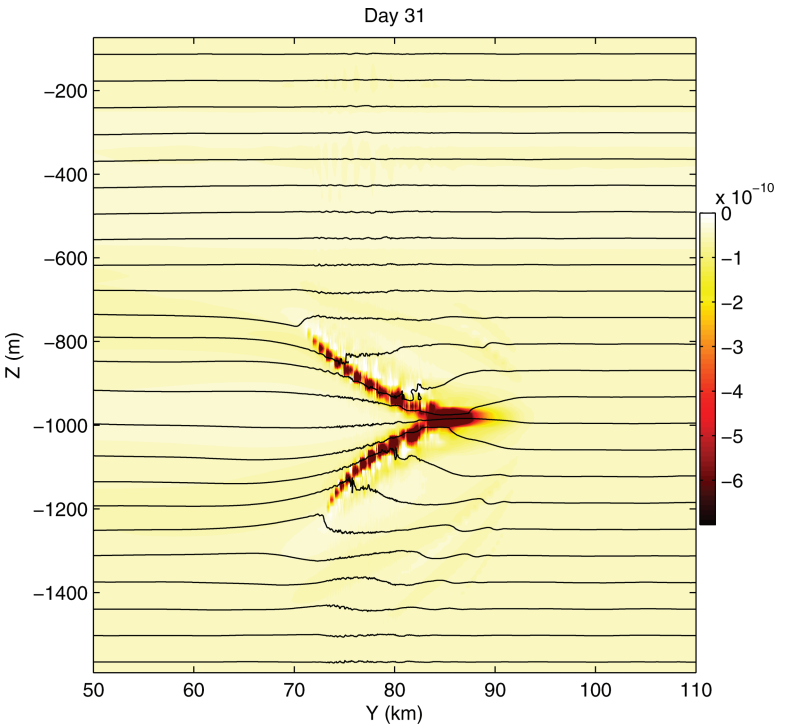

Figure 4. Full Ertel's potential vorticity one half-grid point away from the wall computed from the MITgcm at a point of strong wall-vortex interaction. The vortex is centered at a depth of 950 m and the leading edge of the shock occurs at roughly 90 km along the wall. A strip of anomalously negative potential vorticity appears from the depths of 700 m to 1200 m at 80 to 90 km along the wall. This is associated with the density overturning occurring at the wall. The details of the subsequent shock evolution are central to the net potential vorticity production caused by this event.

Kelvin dynamics north of the shock drive the finite thickness to the south and under denser waters. The gravitationally unstable setting that ensues results in convective overturning, evidences of which appear in Figures 1 and 4. The form of (3) argues there are no dispersive mechanisms available to halt the momentum advection- driven steepening behind the shock formation. In view of this, it appears convective overturning is what balances the steepening and drives shock propagation.

e. Explicit shock dynamics and potential vorticity creation

Given the evidences of convective overturning in the numerical solutions and their implicit presence in the analytical analysis, it is interesting to revisit the full potential vorticity equation, which in layered coordinates is

$$(z_\rho q)_t + (u z_\rho q)_x + ((v-c) z_\rho q)_y = -(H v_\rho)_x + (H u_\rho)_y + Y_x - X_y. \qquad (40)$$

where all adiabatic and viscous quantities have been retained. We do not decompose the variables in (40) into their geostrophic and wall components, as was done earlier. This is for convenience only.

We now consider a surface integral in the density framework of (40) over a region including the shock and bounded on the west by the wall. We take the remaining bounding contours in (x, y) of the domain sufficiently far from the shock that adiabatic and viscous effects can be sensibly neglected. The result is

$$\left[\iint_A z_\rho q \, dx \, dy\right]_t + \int_0^{x_1} [(v-c) z_\rho q]_{y_1}^{y_2} dx + \int_{y_1}^{y_2} [u z_\rho q]^{x_1} dy = \int_{y_1}^{y_2} [H v_\rho - Y]^0 dy \tag{41}$$

The advective potential vorticity fluxes through the sides of the domain are clearly augmented by a nonadvective flux of potential vorticity through the wall. Given that advection and nonadvective effects inside the fluid act only to redistribute on potential density surfaces the potential vorticity already present in the fluid, the creation of a potential vorticity anomaly is clearly the result of the wall processes. This is to be expected given the Impermeability Theorem (Haynes and MacIntyre, 1990).

A simple rearrangement of the full meridional velocity equation leads to

$$v_t + v v_y + M_y = -H v_\rho + Y \tag{42}$$

where again the variables are not decomposed. The right-hand side of (42) is recognized as the (opposite of the) zonal component of the nonadvective potential vorticity flux. The numerical solutions argue once the shock is formed, subsequent evolution is dominated by a nearly steady propagation of the shock structure. We exploit this in (42) by replacing the time derivative by steady meridional propagation. Integrating (42) then along the wall over the region of the convecting front argues the net nonadvective potential vorticity flux from the wall is constrained by the characteristics solutions on either side of the front, i.e.

$$c_s [v]_{y_s}^{y_n} + \left[\frac{v^2}{2}\right]_{y_s}^{y_n} + [M]_{y_s}^{y_n} = \int_{y_s}^{y_n} (-H v_\rho + Y) dy \tag{43}$$

where the brackets $[\bullet]_{y_s}^{y_n}$ denote the jump of the enclosed quantity across the front and the quantity c_s is the shock propagation speed. The implication of (43) is that the net generation of potential vorticity associated with this process is dependent upon the details of the propagating shock. We at the moment lack an adequate shock joining theory which permits us to close the explicit calculation of the potential vorticity generation. Our in-hand numerical solutions are also limited in what they can say about this because of their finite resolution and the apparent role of convective overturning.

However, before exiting this point, it is interesting to return to the reduced gravity setting, where the form of the right-hand side, $\mathbb{R}$, of (43) becomes

$$\mathbb{R} = \int_{y_s}^{y_n} \left(-\frac{evg'}{M} + Y \right) dy \tag{44}$$

where $e = h_t + uh_x + vh_y + w(x, y, -h)$ represents the entrainment rate across the interface of the active layer. Mass conservation in this system is

$$M_t + (uM)_x + (vM)_y = g'e. \tag{45}$$

Supposing e to represent the effects of convective overturning, an area integral of (45) is

$$\left[\iint_A M \right]_t + \oint \vec{u} M \cdot \vec{n} dl = g' \iint_A e dA \tag{46}$$

showing that the irreversible growth of mass in the layer is controlled by the net entrainment. Heat must flow from warm to cold temperatures as a result of mixing, thus convective overturning in our reduced gravity model of a warm layer on top of an inert colder layer should decrease the volume of warm fluid. Regardless of detail, the sign of the integrated entrainment is negative. Now making the potentially dangerous assumptions that the along-wall flow throughout the interval is well described by geostrophy and vanishing potential vorticity anomaly and to the neglect of viscous effects, (44) becomes

$$\mathbb{R} = \int_{y_s}^{y_n} \frac{eMg'}{MfR} dy = \int_{y_s}^{y_n} \frac{eg'}{fR} dy < 0. \tag{47}$$

The form of the integrated potential vorticity equation in this system is

$$\frac{d}{dt} \left[\iint hq \, dA \right] + \oint \vec{u} hq \cdot \vec{n} dl = \int_{y_s}^{y_n} \frac{ve}{h}(0, y) dy \tag{48}$$

thus, the aggregate effect of the wall is to increase the potential vorticity of the system. This is recognized as a deceleration of the balanced flow, as the net potential vorticity anomaly of the anticyclone is negative.

There are many questionable assumptions built into the above estimate. Foremost amongst these is the idea that the velocity in the front is well described by vanishing potential vorticity anomaly. This relates the along-wall flow linearly to the Montgomery which simplifies (47) considerably. Still it is interesting that simple constraints, like the down gradient flow of heat, on the net effects of something as complex as convective overturning lead to a prediction of the proper sign of net potential vorticity generation in this system. This gives us some hope that a parameterization of these processes might be forthcoming. We continue to investigate the details of the shock formation with a view to clarifying these issues.

4. Summary

The problem of the wall response to an interior balanced flow has been considered. It is argued that the problem can be broken into two pieces, an interior 'balanced' part and a wall response. The balanced part is presumed to be governed by an interior distribution of non-trivial potential vorticity, which through the inversion of a diagnostic operator determines the balanced flow on the wall. The wall response, in contrast, is governed by zero potential vorticity anomaly which strongly contrains the structure of the relative vorticity field. Under the relatively loose approximation that along-wall flow is much stronger than that normal to the wall, as a result of the kinematic constraint of the no-normal flow boundary condition, a very simple theory of the wall response can be developed. It consists essentially of a one-dimensional prediction equation for the along- wall velocity, coupled to the solution of a linear elliptic equation that admits separable solutions. Comparisons of the predictions of the simple model with solutions computed using a hydrostatic model argue the model qualitatively and quantitatively captures wall behavior.

Analyzing the simple model leads to several important conclusions about the wall response to balanced flow. In general, we expect a strong asymmetry in the response of the flow to pro-grade and retro-grade flows (in the sense of Kelvin wave propagation). The development of finite amplitude effects and nonhydrostatic responses are to be expected and are the result of advective interactions between the balanced flow and high mode Kelvin wave dynamics. This result was argued previously using a quasigeostrophic approach; here it is generalized well beyond the quasigeostrophic parameter regime and shown to hold almost up to the limits of hydrostatic dynamics. The frontal formation leads to shock wave formation of a Kelvin wave type, the characteristics of which have been computed in a reduced gravity setting.

The dynamics at work in the front to arrest it appear to involve nonhydrostatic effects, as suggested by MITgcm numerical experiments. Incorporating full physics into our potential vorticity statement relates the net anomalous potential vorticity flux to the wall effects in the shock. A crude analysis of these effects argues from simple principles that the convection produces potential vorticity of the proper sense to spin the flow down.

The next steps in this work are to examine the detailed dynamics of the developed shock. Our present intuition is based on the results of a reduced gravity model of a propagating Kelvin shock. Extensions of these ideas to continuously stratified fluids would provide insight into the relevance of the reduced gravity model results; this appears promising and is under current study. Second, our guiding paradigm has been the collision of an open ocean vortex with a wall and this is quite artificial relative to the topography that dominates the ocean interior. On the other hand, relevance of the dynamics has been strongly supported by numerical simulations using more realistic topographies. The study of these dynamics for general topography is still in its early days, however, and more thorough investigations are called for. Of particular interest is the question of whether the present analytical ideas can be generalized to irregular boundaries. We also do not think these ideas are constrained to vortices; they have merely been a convenient and expedient way to introduce a balanced

flow into our model. We anticipate the relevance of the ideas presented here to general balanced flow-boundary interaction.

To the extent that these experiments capture potential vorticity production by balanced flows on boundaries, this paper addresses the classic problem of the closure of the general circulation. All current ocean circulation models rely on frictional parameterizations to perform this closure, an action that is taken more out of numerical necessity than from dynamical considerations. If our ideas are valid, the present analysis is a step in the direction of addressing this major modeling issue. If we are wrong, we hypothesize we are wrong mostly in detail, and that topography will emerge as a major influence controlling the bulk energetics and potential vorticity budgets of the global oceans.

Acknowledgments. This work profited greatly by a joint visit of PB and WKD to the Australian National University, funded by ARC Discovery Project DP0986244. WKD is supported via NSF grants 0961485 and 1049131, the latter in support of the multi-agency Earth Systems Modeling effort. Numerical computations were conducted using the National Facility of the Australian National Computational Infrastructure. This is a submission to the *Journal of Marine Research* volume in honor of Melvin Stern, a long time colleague and friend. WKD would here like to record his immense respect for Professor Stern, having shared an office complex with him for many years. It was always a great learning experience and pleasure to seek him out to pick his brain on any of a number of topics. His door was always open and he was always an inspiration. His loss is shared generally with the field that has been so enriched by his presence.

REFERENCES

Boccaletti, G., R. Ferrari and B. Fox-Kemper. 2007. Mixed layer instabilities and restratification. J. Phys. Oceanogr., *37*, 2228–2250.

Capet, X., J. C. McWilliams, M. J. Molemaker and A. Shchepetkin. 2008a. Mesoscale to submesoscale transition in the California Current System: Flow structure, eddy flux, and observational tests. J. Phys. Oceanogr., *38*, 29–43.

—— 2008b. Mesoscale to submesoscale transition in the California Current System: Frontal processes. J. Phys. Oceanogr., *38*, 44–64.

Charney, J. 1971. Geostrophic turbulence. J. Atmos. Sci, *28*, 1087–1095.

D'Asaro, E., C. Lee, L. Rainville, R. Harcourt and L. Thomas. 2011. Enhanced turbulence and energy dissipation at ocean fronts. Science, *332*, 318–322.

Deremble, B., A. M. Hogg, P. Berloff and W.K. Dewar. 2011. On the application of no-slip lateral boundary conditions to 'coarsely' resolved ocean models. Ocean Model., *39*, 411–415.

Dewar, W. K. and A. M. Hogg. 2010. Topographic inviscid dissipation of balanced flow. Ocean Model., *32*, 1–13.

Ferrari, R. and C. Wunsch. 2008. Ocean circulation kinetic energy—Reservoirs, sources and sinks. Ann. Rev. Fluid Mech., *41*, 253–282.

Haynes, P. H. and M. E. McIntyre. 1990. On the conservation and impermeability theorems for potential vorticity. J. Atmos. Sci., *47*, 2021–2031.

Hogg, A. M., W. K. Dewar, P. Berloff and M. Ward, 2011. Kelvin wave hydraulic control induced by interactions between vortices and topography. J. Fluid Mech., *687*, 194–208.

Hoskins, B. J. and F. P. Bretherton. 1972. Atmospheric frontogenesis models: Mathematical formulation and solution. J. Atmos. Sci. *29*, 11–37.

Hughes, G. O., A. M. Hogg and R. W. Griffiths. 2009. Available potential energy and irreversible mixing in the meridional overturning circulation. J. Phys. Oceanogr., *39*, 3130–3146.

Molemaker, M. J., J. C. McWilliams and W. K. Dewar. 2012. Submesoscale interactions catalyzed by topography in the California Undercurrent System. (in preparation).

Naviera-Garabato, A., K. Polzin, B. King, K. Heywood and M. Visbeck. 2004. Widespread intense turbulent mixing in the Southern Ocean. Science, *303*, 210–213.

Nikurashin, M. and R. Ferrari. 2010. Radiation and dissipation of internal waves generated by geostrophic motions impinging on small-scale topography: Application to the Southern Ocean. J. Phys. Oceanogr., *40*, 2025–2042.

Oort, A. H., L. A. Anderson and J. P. Peixoto. 1994. Estimates of the energy cycle of the oceans. J. Geophys. Res., *99*, 7665–7688.

Reznik, G. M. and G. G. Sutyrin. 2005. Non-conservation of "geostrophic mass" in the presence of a long boundary and related Kelvin wave. J. Fluid Mech., *527*, 235–264.

Richardson, P. L., A. S. Bower and W. Zenk. 2000. A census of Meddies tracked by floats. Prog. Oceanogr., *45*, 209–250.

Sen, A., R. B. Scott and B. K. Arbic. 2008. Global energy dissipation rate of deep-ocean low-frequency flows by quadratic bottom boundary layer drag: Computations from current meter data. Geophys. Res. Lett., *35*, L09606, doi:10.1029/2008GL033407.

Whitham, G. 1974. Linear and Nonlinear Waves, John Wiley & Sons, New York, 636 pp.

Yavneh, I., J. C. McWilliams and M. J. Molemaker. 2001. Non-axisymmetric instability of centrifugally stable, stratified Taylor-Couette flow. J. Fluid. Mech. *448*, 1–21.

Received: *9 May, 2011*; revised: *30 November, 2011*.