*J. Fluid Mech.* (2000), *vol.* 407, *pp.* 235–263. Printed in the United Kingdom



# On the vorticity transport due to dissipating or breaking waves in shallow-water flow

By OLIVER BÜHLER

School of Mathematics, University of St Andrews, St Andrews KY16 9SS, UK

(Received 20 July 1999 and in revised form 8 November 1999)

Theoretical and numerical results are presented on the transport of vorticity (or potential vorticity) due to dissipating gravity waves in a shallow-water system with background rotation and bottom topography. The results are obtained under the assumption that the flow can be decomposed into small-scale gravity waves and a large-scale mean flow. The particle-following formalism of 'generalized Lagrangian-mean' theory is then used to derive an 'effective mean force' that captures the vorticity transport due to the dissipating waves. This can be achieved without neglecting other, non-dissipative, effects which is an important practical consideration. It is then shown that the effective mean force obeys the so-called 'pseudomomentum rule', i.e. the force is approximately equal to minus the local dissipation rate of the wave's pseudomomentum. However, it is also shown that this holds only if the underlying dissipation mechanism is momentum-conserving. This requirement has important implications for numerical simulations, and these are discussed.

The novelty of the results presented here is that they have been derived within a uniform theoretical framework, that they are not restricted to small wave amplitude, ray-tracing or JWKB-type approximations, and that they also include wave dissipation by breaking, or shock formation. The theory is tested carefully against shock-capturing nonlinear numerical simulations, which includes the detailed study of a wavetrain subject to slowly varying bottom topography. The theory is also cross-checked in the appropriate asymptotic limit against recently formulated weakly nonlinear theories. In addition to the general finite-amplitude theory, detailed small-amplitude expressions for the main results are provided in which the explicit appearance of Lagrangian fields can be avoided. The motivation for this work stems partly from an on-going study of high-altitude breaking of internal gravity waves in the atmosphere, and some preliminary remarks on atmospheric applications and on three-dimensional stratified versions of these results are given.

## 1. Introduction

The creation of vorticity due to dissipating or breaking waves is a basic fluid-dynamical phenomenon that is central to many geophysical and engineering applications. Classical examples include acoustic streaming and the so-called 'quartz wind' (e.g. Lighthill 1978), the vorticity creation due to the breaking of surface waves in the ocean (especially near shorelines), and the vorticity created by the passage of an irrotational supersonic flow through a curved shock, or through a shock of varying strength. An important geophysical application is the vorticity creation due to the breaking of internal gravity waves at high altitudes in the atmosphere (e.g. Holton *et*

*al.* 1995), which is recognized as being significant for the global-scale circulation of the atmosphere.

Theoretical descriptions of wave dissipation can often establish a useful quantitative link between vorticity creation and the dissipative decay of local wave properties such as wave energy or wave pseudomomentum. Indeed, numerical models for the atmosphere rely completely on such a theoretical description of gravity-wave breaking, because most atmospheric gravity waves are much too small in vertical scale to be directly resolved in such models. This process is known as gravity-wave parametrization, and in it the effect of dissipating gravity waves is represented by a suitable effective mean force $\tilde{\boldsymbol{F}}$ directed along stratification surfaces. The magnitude of $\tilde{\boldsymbol{F}}$ is then calculated using assumptions about the prevailing gravity-wave spectrum together with approximate results of wave–mean interaction theory.

However, it can be argued that the relevant wave–mean interaction theory is simple and well understood only for calculating the zonal acceleration of a zonally symmetric mean flow (i.e. a mean flow defined by averaging around latitude circles; 'zonal' stands for east–west). This is because in this case there can be no mean pressure gradient in the zonal direction, and this greatly simplifies the theory. The theory is significantly more complicated and less well understood in the case of a local average, say, a local average over small-scale gravity waves. This is partly because of the more complicated physical situation and partly because of subtle technical difficulties that arise when mean-and-disturbance formalisms are applied to compressible or stratified fluids. These difficulties can easily obscure essential physical features of the problem, such as the way circulation around material contours evolves when waves are dissipating or breaking. Studies that investigate these physical features directly are hence of great value; a recent example is given by the calculations of Peregrine (1998) for breaking ocean waves.

A central reference for studies of this kind is McIntyre & Norton (1990, hereafter MN). Their paper presented a conceptual and mathematical *tour de force*, drawing together many different theoretical threads in order to embrace a large number of geophysical and engineering applications. By focusing on the basic properties of vorticity and potential vorticity (PV), MN avoided the technical difficulties of any particular averaging formalism. Their aim was to link $\tilde{\boldsymbol{F}}$ directly to the dissipation rate of the wave's pseudomomentum $\boldsymbol{p}$, an approximate connection that has been called the 'pseudomomentum rule' in the literature. Such a rule is practically useful because the pseudomomentum $\boldsymbol{p}$ plays an important part in linear wave theory. For instance, atmospheric gravity-wave spectra are routinely described in terms of the pseudomomentum spectra of these waves. In order to validate the pseudomomentum rule MN presented several brief examples of their methods, but they did not present a full mathematical description (or justification) of them. Indeed, applying their methods naïvely to more complicated examples has sometimes failed (R. Mo & M. E. McIntyre, personal communication; see also § 2 below), and this has partly motivated the work presented here.

Here, a detailed study is undertaken of the relevant wave–mean interaction theory in the simplest possible flow system in which these interactions are non-trivial, namely the two-dimensional shallow-water system. This provides a test bed and stepping stone towards more complicated three-dimensional stratified models, for which some preliminary results are noted in § 7. The mean flow is defined by a local space average, assuming a suitable scale separation between mean flow and small-scale gravity waves. The averaging method itself is the particle-following generalized Lagrangian-mean (GLM) formalism of Andrews & McIntyre (1978*a*, *b*, hereafter AM78*a*, *b*) (see

also McIntyre 1981), which gives access to averaged versions of Kelvin's circulation theorem and to other conservation laws. In this way, the simplifying power of averaging can be combined with a focus on the essential vorticity dynamics. Specifically, the wave–mean interaction theory is developed in this paper in two distinct steps:

(*a*) a precise definition is given of an effective mean force $\tilde{\boldsymbol{F}}$ that captures the vorticity transport due to dissipating gravity waves, and

(*b*) the conditions are found under which the pseudomomentum rule holds, meaning that $\tilde{\boldsymbol{F}}$ can be approximately equated to minus the dissipative decay rate of the wave's pseudomomentum $\boldsymbol{p}$.

The first step can be achieved at finite amplitude and without any approximations beyond the already assumed spatial scale separation needed to define the mean flow. Also, non-dissipative mean-flow corrections that are unrelated to the effective mean force $\tilde{\boldsymbol{F}}$ are naturally incorporated at this step. Such corrections can be significant in some circumstances, as discussed in detail in Bühler & McIntyre (1998, hereafter BM). Remarkably, the second, approximate step can be achieved by making just one further assumption: the wave dissipation process must be momentum-conserving, meaning that the underlying dissipative force must arise from a stress-tensor divergence in the usual way.

The plan for this paper is as follows. The equations of motion for a shallow-water system with background rotation and bottom topography are introduced in § 2, and the vorticity transport in that system is discussed broadly in the terms of MN, who did not consider shallow-water flow explicitly. This sets the scene for the GLM theory developed later, and it also highlights some of the difficulties with the original method of MN and certain *ad hoc* variations of it. The first step of the wave–mean programme noted above is then implemented in § 3, resulting in an exact formula for the effective mean force $\tilde{\boldsymbol{F}}$ derived using GLM theory. The second step follows in § 4, where momentum-conserving forces and the validity of the pseudomomentum rule are discussed. This section also contains useful small-amplitude relations for the various results and a discussion of the implications for direct numerical simulations. Shock-capturing nonlinear numerical simulations of initial-value problems and forced–dissipative problems with bottom topography are presented in § 5. These simulations are in very good agreement with the theory except in one case with very strong background rotation, where the diagnostic signal-to-noise ratio was low and the comparison was less clear. The theory is also compared and checked against the aforementioned asymptotic theories in § 6, which clarifies certain aspects of both theories. Some concluding remarks on atmospheric applications are given in § 7.

## 2. Vorticity transport in shallow-water flow

The equations of motion for the two-dimensional shallow-water system in a rotating frame and including bottom topography as well as momentum forcing are given by the continuity equation

$$\frac{\mathrm{D}h}{\mathrm{D}t} + h\nabla\cdot\boldsymbol{u} = 0 \tag{2.1}$$

and the momentum equation

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \boldsymbol{f}\times\boldsymbol{u} + c_0^2\,\nabla(h+h_B) = \boldsymbol{F}. \tag{2.2}$$

Here $h$ is the non-dimensional fluid layer depth such that $h = 1$ corresponds to a nominal reference background layer depth, $\boldsymbol{u} = (u, v)$ is the two-dimensional velocity

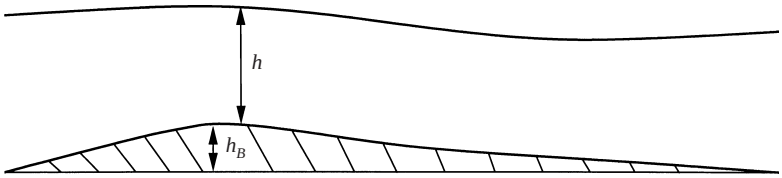


FIGURE 1. Schematic cross-section of the shallow-water system showing the fluid layer depth $h$ and the bottom elevation $h_B$, which add to give the fluid surface height $h + h_B$.

in the $(x, y)$ directions, $\boldsymbol{f} = f\hat{\boldsymbol{z}}$ is the constant Coriolis parameter $f$ times the unit vector normal to the $(x, y)$-plane $\hat{\boldsymbol{z}}$, and the speed $c_0$ is a measure of the gravitational force acting on undulations of the fluid surface. The bottom topography is described through the bottom elevation $h_B$ as illustrated in figure 1, and $h_B$ can also be thought of as including the centrifugal acceleration due to the rotating frame. Note that an undisturbed fluid surface corresponds to $h + h_B = \text{const}$. Finally, $\boldsymbol{F}$ is an arbitrary body force per unit mass, of dissipative origin or otherwise. The PV is denoted by $q$, and it can be seen to satisfy

$$q \equiv \frac{\nabla \times \boldsymbol{u} + f}{h} \quad \text{and} \quad \frac{\mathrm{D}q}{\mathrm{D}t} = \frac{\nabla \times \boldsymbol{F}}{h}, \tag{2.3}$$

where the curl of a two-dimensional vector field is treated as a (pseudo-)scalar.

The PV evolution equation can be re-written in conservation or flux form as

$$\frac{\partial(hq)}{\partial t} + \nabla \cdot [hq\boldsymbol{u} + \boldsymbol{J}] = 0, \quad \text{where } \boldsymbol{J} \equiv \hat{\boldsymbol{z}} \times \boldsymbol{F}. \tag{2.4}$$

Here, the body force $\boldsymbol{F}$ has been replaced by an equivalent vorticity flux $\boldsymbol{J}$, where equivalence means that $\nabla \times \boldsymbol{F} = -\nabla \cdot \boldsymbol{J}$ by construction. As noted in MN, irrotational contributions to $\boldsymbol{F}$ correspond to non-divergent contributions to $\boldsymbol{J}$. The conservation form (2.4) makes obvious that changes in absolute vorticity $hq = \nabla \times \boldsymbol{u} + f$ can always be thought of as arising through vorticity transport. It can be noted in passing that the flux vector $\boldsymbol{J}$ can be made the basis of a number of conceptual interpretations of PV evolution in shallow-water flow and in other flow systems, as has been discussed in detail in MN and elsewhere. However, in this paper it is sufficient and simpler to work directly with the force $\boldsymbol{F}$, and hence the flux vector $\boldsymbol{J}$ will be used only in this section, where it facilitates discussion and comparison with MN.

Now, the basic approach discussed in MN is to consider wave-induced contributions to the flux vector $\boldsymbol{J}$ in (2.4) for the case where these fluxes arise through wave dissipation. Note that in this section, as in MN, only a broad outline of this approach is given, leaving many mathematical details unspecified. To this end, consider a wavetrain of small-amplitude linearized gravity waves, whose amplitude is $O(a)$ where $a \ll 1$ is a suitable non-dimensional wave amplitude. The linear, $O(a)$ contribution to the dissipative flux $\boldsymbol{J}$ averages to zero over one wave period, but at $O(a^2)$ there can be a systematic contribution, which will be denoted by $\overline{\boldsymbol{J}}$. This $O(a^2)$ contribution to $\boldsymbol{J}$ can often be calculated from a knowledge of the linearized, $O(a)$ wave solution alone.†

If (2.4) is integrated over a fixed, Eulerian control volume and then averaged over a

† In general, quantities for which this statement holds were called 'wave properties' in AM78*a*, *b*, a useful terminology that is adopted in this paper. Other examples of wave properties are wave energy, pseudomomentum, and Stokes corrections. On the other hand, $O(a^2)$ corrections to the layer depth $h$ and to the velocity $\boldsymbol{u}$ are not wave properties.

wave period, then $\overline{\boldsymbol{J}}$ captures the wave-induced vorticity flux in and out of this control volume, *provided* that the advective vorticity flux $hq\boldsymbol{u}$ makes no contribution at $O(a^2)$. This can be shown to hold for the examples studied in MN and here for irrotational gravity waves with $f = 0$, but fails when $f \neq 0$. In the latter case the advective term must be considered explicitly to obtain the flux at $O(a^2)$. This is inconvenient, as there are now more terms to consider, despite there being no explicit presence of dissipative terms in $hq\boldsymbol{u}$. The simplicity of dealing only with $\boldsymbol{J}$ can be regained if (2.4) is integrated over a moving, Lagrangian control volume, for which (2.4) takes the relevant form

$$h\frac{\mathrm{D}q}{\mathrm{D}t} + \nabla\cdot\boldsymbol{J} = 0. \tag{2.5}$$

The relevant flux is now a flux through undulating material control surfaces, which involves working out the dot product between $\boldsymbol{J}$ and undulating surface normals (R. Mo & M. E. McIntyre, personal communication). This *ad hoc* Lagrangian approach was applied to stratified flows using zonal averaging in Mo (1994) and using JWKB approximations in Bühler (1996).

However, the use of Lagrangian control volumes makes the obtained results hard to interpret in a conventional, Eulerian framework. Furthermore, Lagrangian control volumes themselves are subject to $O(a^2)$ irreversible displacements due to mean-flow effects first described by Bretherton (1969). For instance, in the presence of $O(1)$ PV gradients these displacements produce non-dissipative $O(a^2)$ changes in the local PV that can be comparable to dissipative PV changes in some circumstances, as discussed in detail in BM.

The overall conclusion is that the dissipation-related vorticity transport is simplest in a Lagrangian, particle-following description, but that a clear interpretation of this vorticity transport requires a solid mathematical framework in which all relevant Eulerian and Lagrangian effects can be properly formulated. The GLM theory of AM78*a*, *b* provides such a framework, and its use for the present purpose has been further prompted by the discovery of certain results concerning the GLM PV described in BM. The next two sections will pursue the application of GLM theory to this problem.

## 3. Finite-amplitude GLM theory and the pseudomomentum rule

This section develops the finite-amplitude GLM theory for the wave-induced vorticity transport in the shallow-water system. A short self-contained introduction to GLM theory and to the relevance of GLM theorems governing vorticity evolution can be found in §§ 5.1–5.2 of BM, while a full account of the theory is given in the original publications (AM78*a*, *b*). Only a brief summary of the special GLM notation is given here, together with some useful details not stressed in these other references.

### 3.1. *GLM theory*

It is assumed from the outset that the flow is such that a spatial averaging operation $\overline{(\ldots)}$ can be defined that decomposes any flow variable $\phi$ uniquely into a slowly varying mean part $\overline{\phi}$ and a rapidly varying, small-scale disturbance part $\phi' \equiv \phi - \overline{\phi}$. No particular assumption is made about the time-dependence of the mean and the disturbance parts. It is assumed, in the usual way, that the spatial scale separation is sufficiently large such that the mean fields are uniquely defined (i.e. that $\overline{\phi'} \equiv 0$).

This assumption represents the only approximation in the finite-amplitude theory developed here.†

GLM theory differs from Eulerian-mean theories by explicitly introducing an additional field, namely a disturbance-associated particle displacement field $\boldsymbol{\xi}(\boldsymbol{x},t)$ such that $\boldsymbol{x}+\boldsymbol{\xi}(\boldsymbol{x},t)$ is the *actual* position of the particle whose *mean* position is $\boldsymbol{x}$ (cf. figure 3 in BM for an illustration and further discussion of the nature of $\boldsymbol{\xi}$). By construction, $\overline{\boldsymbol{\xi}}\equiv 0$ holds. The additional field $\boldsymbol{\xi}$ allows the 'lifting' of a function from mean to actual particle positions, as described by the useful notation

$$\phi^{\xi}(\boldsymbol{x},t)\equiv\phi(\boldsymbol{x}+\boldsymbol{\xi}(\boldsymbol{x},t),t)\quad\text{such that }\overline{\phi}^{L}\equiv\overline{\phi^{\xi}} \tag{3.1}$$

is the Lagrangian (i.e. particle-following) mean of $\phi$. Here $\phi^{\xi}(\boldsymbol{x},t)$ is the value of $\phi$ at the location of the particle whose mean position is $\boldsymbol{x}$ at time $t$, and $\overline{(\ldots)}$ is the averaging operator introduced above. Thus, $\overline{\phi}^{L}(\boldsymbol{x},t)$ is the average of $\phi$ as seen by the particle whose mean position is $\boldsymbol{x}$ at time $t$.

By analogy with $\phi'$, the Lagrangian disturbance of $\phi$ is defined as $\phi^{\ell}\equiv\phi^{\xi}-\overline{\phi}^{L}$ and $\overline{\phi^{\ell}}\equiv 0$ holds by construction. The Lagrangian-mean velocity $\overline{\boldsymbol{u}}^{L}$ defines the mean material time derivative as

$$\overline{\mathrm{D}}^{L}\equiv\frac{\partial}{\partial t}+\overline{\boldsymbol{u}}^{L}\cdot\nabla, \tag{3.2}$$

and integral curves of $\overline{\boldsymbol{u}}^{L}$ are mean particle trajectories. Correspondingly, $\boldsymbol{x}+\boldsymbol{\xi}(\boldsymbol{x},t)$ evaluated along a mean particle trajectory traces out the actual, rapidly varying particle trajectory. This can be shown to lead to the remarkably simple GLM formulae for averaging material derivatives:

$$\overline{\mathrm{D}}^{L}\phi^{\xi}=\left(\frac{\mathrm{D}\phi}{\mathrm{D}t}\right)^{\xi}\quad\text{and}\quad\overline{\mathrm{D}}^{L}\overline{\phi}^{L}=\overline{\left(\frac{\mathrm{D}\phi}{\mathrm{D}t}\right)}^{L}. \tag{3.3}$$

The price for this simplicity is that Lagrangian averaging and partial differentiation do not commute. Specifically, the chain rule gives

$$(\phi^{\xi})_{,t}=(\phi_{,t})^{\xi}+(\phi_{,j})^{\xi}\xi_{j,t}\quad\text{and}\quad(\phi^{\xi})_{,i}=(\phi_{,j})^{\xi}\,(\delta_{ji}+\xi_{j,i}), \tag{3.4}$$

where $(\,)_{,t}$ denotes time-differentiation, $(\,)_{,i}$ denotes differentiation with respect to the Cartesian coordinate $x_i$, the $\xi_j$ are the components of the two-dimensional vector $\boldsymbol{\xi}=(\xi,\eta)$, and summation over repeated indices is understood. The spatial derivative in (3.4) has the inverse formula

$$(\phi_{,i})^{\xi}=\frac{1}{J}K_{ij}(\phi^{\xi})_{,j}, \tag{3.5}$$

where the scalar $J$ is the Jacobian of the lifting map $\boldsymbol{x}\to\boldsymbol{x}+\boldsymbol{\xi}$, i.e.

$$J\equiv\frac{\partial(x+\xi,y+\eta)}{\partial(x,y)}=\det\{\delta_{ij}+\xi_{i,j}\}=(1+\xi_{,x}+\eta_{,y}+\xi_{,x}\eta_{,y}-\xi_{,y}\eta_{,x}) \tag{3.6}$$

in two dimensions, and where the matrix $K_{ij}$ contains the cofactors‡ of the matrix

† It can be noted that local averaging alone does not guarantee that the mean fields are slowly varying. The simplest counterexample is that of evanescent waves, e.g. surface waves trapped at an interface, whose envelope remains rapidly varying in the transverse direction even after averaging along the interface (cf. McIntyre 1988). Additional averaging over the transverse waveguide structure is then necessary to make the mean fields slowly varying in all directions.

‡ The cofactors $m_{ij}$ of a matrix $\boldsymbol{M}$ are the sub-determinants (with appropriate sign) obtained by removing the $i$th row and the $j$th column from $\boldsymbol{M}$. It follows that $m_{ij}/\det(\boldsymbol{M})$ is the inverse of $\boldsymbol{M}^{T}$.

$\{\delta_{ij} + \xi_{i,j}\}$, i.e.

$$K_{ij} \equiv \begin{pmatrix} 1+\eta_{,y} & -\eta_{,x} \\ -\xi_{,y} & 1+\xi_{,x} \end{pmatrix}, \quad \text{or} \quad K_{ij} = \frac{\partial J}{\partial(\xi_{i,j})}. \tag{3.7}$$

In the useful second relation the partial derivative acts on the functional dependence of $J$ on the numbers $\xi_{i,j}$. The explicit expressions for $J$ and $K_{ij}$ in (3.6)–(3.7) hold for two space dimensions. In general, $J$ and $K_{ij}$ are polynomials in the numbers $\xi_{i,j}$ of order $n$ and $n-1$ respectively, where $n$ is the number of space dimensions. A general property of $K_{ij}$ is that it provides the lifting map for surface elements. This means that if $\mathrm{d}A_i$ is a vectorial surface element in $n$-dimensional space, then

$$(\mathrm{d}A_i)^{\xi} = K_{ij}\,\mathrm{d}A_j, \quad \text{which implies } K_{ij,j} \equiv 0 \tag{3.8}$$

as shown in AM78*a*.

### 3.2. *Shallow-water GLM theory and mean vorticity budget*

The GLM theory is now applied to the shallow-water system. The functional form of the continuity equation (2.1) can be preserved in GLM theory by defining a mean layer depth $\tilde{h}$ such that

$$\tilde{h}\,\mathrm{d}x\,\mathrm{d}y \equiv h^{\xi}(\mathrm{d}x\,\mathrm{d}y)^{\xi} \quad \Leftrightarrow \quad \tilde{h} = h^{\xi} J \tag{3.9}$$

holds, where $(\mathrm{d}x\,\mathrm{d}y)^{\xi} = J\mathrm{d}x\,\mathrm{d}y$ denotes the actual material 'volume' element corresponding to the mean material element $\mathrm{d}x\,\mathrm{d}y$. If mass is exactly conserved by the flow, then $\tilde{h}$ is a mean quantity at all times provided it was so initially (AM78*a*). The definition (3.9) is then equivalent to

$$\boxed{\overline{\mathrm{D}}^{L}\tilde{h} + \tilde{h}\nabla\cdot\bar{\boldsymbol{u}}^{L} = 0}\,. \tag{3.10}$$

In general $\tilde{h} \neq \bar{h}^{L}$, yet (3.10) exemplifies that $\tilde{h}$ rather than $\bar{h}^{L}$ is the natural GLM mass density. In analogy with $\tilde{h}$, the tilde notation will be used repeatedly in what follows to denote GLM quantities that are defined in a way that leaves important fluid-mechanical relations (such as (2.1)) form-invariant under averaging. See, for another example, the appearance of $\tilde{\boldsymbol{F}}$ in (3.16) below.

There are several ways to derive the GLM vorticity budget. One approach is to manipulate and average the momentum equation (2.2), and then to consider its curl (cf. Theorem I and Corollary III in AM78*a*; and § 5.2 in BM). However, this approach necessitates manipulating several complicated irrotational mean vector fields at intermediate steps, although eventually these fields have no effect on the vorticity budget. This is why a different approach directly based on the PV equation (2.3) is adopted here, which side-steps these complications. This approach also makes obvious the surprisingly simple relation of the Lagrangian-mean PV $\bar{q}^{L}$ to other GLM fields, whose validity in the presence of forcing and dissipation was first demonstrated in BM. It is convenient to formulate first the following useful lemma, from which the other results then follow easily.

LEMMA 1. *Let* $\boldsymbol{B}(\boldsymbol{x},t)$ *be any vector field and let* $\tilde{\boldsymbol{B}}(\boldsymbol{x},t)$ *be a mean vector field such that*

$$\tilde{\boldsymbol{B}}\cdot\mathrm{d}\boldsymbol{x} = \overline{(\boldsymbol{B}\cdot\mathrm{d}\boldsymbol{x})}^{L}, \tag{3.11}$$

*where* $\mathrm{d}\boldsymbol{x}$ *is a material line element. Then*

$$\boxed{\tilde{B}_i = \overline{B}_i^L + \overline{\xi_{j,i} B_j^\ell} \quad \textit{and} \quad \overline{\left(\frac{\nabla \times \boldsymbol{B}}{h}\right)}^L = \frac{\nabla \times \tilde{\boldsymbol{B}}}{\tilde{h}}}\,. \tag{3.12}$$

*Proof.* The first equality follows from substituting

$$(\mathrm{d}x_i)^\xi = \mathrm{d}x_i + \mathrm{d}\xi_i = \mathrm{d}x_i + \xi_{i,j}\,\mathrm{d}x_j \tag{3.13}$$

in the right-hand side of (3.11) and noting that $\mathrm{d}x_i$ is arbitrary and that $\overline{\xi_{j,i} B_j^\xi} = \overline{\xi_{j,i} B_j^\ell}$ because $\overline{\xi_i} = 0$. The second equality in (3.12) follows from a straightforward generalization of BM's proof† for the special case in which $\boldsymbol{B}$ was the absolute velocity; the proof is given here in an Appendix. □

This lemma can now be applied to the Lagrangian-mean of the definition of $q$ and to its evolution equation, as given in (2.3). The first yields

$$\boxed{\overline{q}^L = \overline{\left(\frac{\nabla \times \boldsymbol{u} + f}{h}\right)}^L = \frac{\nabla \times (\overline{\boldsymbol{u}}^L - \boldsymbol{p}) + f}{\tilde{h}}}\,, \tag{3.14}$$

using the absolute velocity $\boldsymbol{B} = \boldsymbol{u} + \frac{1}{2}\boldsymbol{f} \times \boldsymbol{x}$ in (3.12) and the fact that $f$ is constant.‡ This introduces the GLM pseudomomentum vector $\boldsymbol{p}$ defined as

$$p_i \equiv -\overline{\xi_{j,i} [\boldsymbol{u}^\ell + \tfrac{1}{2}\boldsymbol{f} \times \boldsymbol{\xi}]_j}, \tag{3.15}$$

where $\boldsymbol{u}^\ell$ could be replaced by $\boldsymbol{u}^\xi = \boldsymbol{u}^\ell + \overline{\boldsymbol{u}}^L$ without changing $\boldsymbol{p}$, because $\overline{\boldsymbol{\xi}} = 0$. The striking simplification achieved in (3.14) should be noted: the curl and the (inverse) density were both 'pulled out' of the Lagrangian averaging process by introducing the GLM fields $\tilde{h}$ and $\boldsymbol{p}$ (cf. BM). Also, the appearance of $\boldsymbol{p}$ in the functional relation between $\overline{q}^L$ and $\overline{\boldsymbol{u}}^L$ forms the basis for the non-dissipative wave–mean interactions studied in BM.

The lemma (3.12) is now applied with $\boldsymbol{B} = \boldsymbol{F}$ to the Lagrangian-mean of the second equation in (2.3), yielding

$$\boxed{\overline{\mathrm{D}}^L \overline{q}^L = \frac{\nabla \times \tilde{\boldsymbol{F}}}{\tilde{h}} \quad \text{and} \quad \tilde{F}_i = \overline{F}_i^L + \overline{\xi_{j,i} F_j^\ell}}\,. \tag{3.16}$$

This equation¶ has the same functional form as the PV equation before averaging in (2.3), which makes obvious that the global conservation property of $hq$ has been inherited by $\tilde{h}\overline{q}^L$. Equation (3.16) establishes that the GLM vorticity transport can indeed be described by an effective mean force, as discussed in the introduction. This mean force, $\tilde{\boldsymbol{F}}$, is well-defined in the sense that it depends only on the local flow fields, and that its curl is insensitive to the irrotational part of the force $\boldsymbol{F}$. In fact, it is straightforward to show (using (3.4)) that if $\boldsymbol{F} \to \boldsymbol{F} + \nabla\phi$, then $\tilde{\boldsymbol{F}} \to \tilde{\boldsymbol{F}} + \nabla\overline{\phi}^L$,

† There is a typographical error in BM's proof: the symbol $A$ must be deleted in the first two integrals in their (5.22).

‡ The case of non-constant $f$ is relevant in geophysics, and in this case the exact definition of $\boldsymbol{p}$ must be modified. However, the modification is often negligible in typical scaling regimes (BM; Bühler 1996).

¶ A variant of (3.16) that does not recognize the general validity of (3.14) follows directly from taking the curl of Theorem I in AM78*a*; see also Mo (1994).

leaving $\nabla \times \tilde{\boldsymbol{F}}$ unchanged. This completes the first step of the programme laid out in the introduction; the second step of this programme is now to establish clearly the extent to which $\tilde{\boldsymbol{F}}$ can be related to wave dissipation.

### 3.3. *GLM pseudomomentum rule*

This is achieved by deriving the evolution equation for $\boldsymbol{p}$. To do this, the $j$th component of the Lagrangian-disturbance part of the momentum equation (2.2) is multiplied by $-\xi_{j,i}$ and then averaged. The details of this conceptually straightforward derivation are described in AM78*b*. The result is

$$\overline{\mathrm{D}}^L p_i + \bar{u}^L_{k,i} p_k + \tfrac{1}{2}\overline{(u^\ell_k u^\ell_k)}_{,i} + \tfrac{1}{2}\overline{(\boldsymbol{u}^\ell \cdot \boldsymbol{f} \times \boldsymbol{\xi})}_{,i} - c_0^2\,\overline{\xi_{j,i}(h_{,j})^\ell} - c_0^2\,\overline{\xi_{j,i}(h_{B,j})^\ell} = \mathscr{F}_i, \qquad (3.17)$$

where

$$\mathscr{F}_i \equiv -\overline{\xi_{j,i} F^\ell_j}. \qquad (3.18)$$

It can be noted in passing that after multiplication with $\tilde{h}$, the terms on the left-hand side can be further decomposed into conservative flux-divergence parts and non-conservative source/sink parts; cf. AM78*b*. Now, the forcing term $\mathscr{F}_i$ is a wave property (i.e. it can be consistently evaluated at $O(a^2)$ in terms of the linearized, $O(a)$ disturbances, where $a \ll 1$ is the disturbance amplitude; cf. AM78*a*) and it describes the mean generation or destruction of pseudomomentum $p_i$ due to the body force $F_i$. Comparing (3.18) and the second of (3.16) produces the remarkably simple result that

$$\boxed{\tilde{F}_i = \overline{F}^L_i - \mathscr{F}_i}\,. \qquad (3.19)$$

This equation together with (3.14) and (3.16) is the main result of this section.

How should (3.19) be interpreted? The pseudomomentum rule as described in MN states that $\tilde{F}_i$ should be equal at leading order to minus the force-induced dissipation rate of pseudomomentum, which is $-\mathscr{F}_i$. Equation (3.19) now makes obvious that this is in general only an approximation, i.e. the pseudomomentum rule holds only if†

$$\tilde{F}_i = \overline{F}^L_i - \mathscr{F}_i \approx -\mathscr{F}_i \quad \Leftrightarrow \quad |\overline{F}^L_i| \ll |\mathscr{F}_i|. \qquad (3.20)$$

There are no general reasons why (3.20) should hold for arbitrary $F_i$. Indeed, (3.20) could fail even for forces with vanishing Eulerian mean, because $|\overline{F}_i|$ and $|\overline{F}^L_i|$ can differ by Stokes corrections of the same magnitude as $|\mathscr{F}_i|$. A case in point is the generation of small-scale waves by a suitable irrotational force $\boldsymbol{F} = \nabla\phi_w$ with vanishing mean (cf. the numerical simulations in § 5.3 below). The force clearly generates pseudomomentum and hence $\mathscr{F}_i \neq 0$, but because $\nabla \times \boldsymbol{F} = 0$ there can be no material change in PV whatever and hence the pseudomomentum rule does not apply.

In general, (3.20) must be carefully checked to establish the validity of the pseudomomentum rule in any given case, but a much more definite statement can be made in the special case of a momentum-conserving force. This is investigated in the next section.

† Strictly speaking, (3.20) needs only to hold for the respective curls of the vectors fields in (3.19). However, I have not been able to find a case in which this more strict condition leads to interesting differences.

## 4. Momentum-conserving body force $F_i$

A strong argument can be made for the pseudomomentum rule to hold approximately provided that the body force $F_i$ is momentum-conserving and that the mean fields are slowly varying. The argument is not restricted to small-amplitude disturbances (though this is an important theoretical regime), it applies to both smooth and shock-forming disturbances, and it has significant implications for numerical schemes.

### 4.1. *Validity of the pseudomomentum rule*

The density of momentum per unit area in shallow-water flow is $h\boldsymbol{u}$ and hence a momentum-conserving body force $F_i$ in (2.2) must derive from a stress tensor $\sigma_{ij}$ via

$$F_i = \frac{1}{h}\,\sigma_{ij,j}, \tag{4.1}$$

where the inverse density pre-factor $1/h$ is crucial for the conservation property. This conservative form of $F_i$ is clearly satisfied, for instance, by the usual viscous force. It can be shown by using (3.5), (3.8), and (3.9) that this conservative form is inherited by $\overline{F}_i^L$ as

$$\overline{F}_i^L = \frac{1}{\tilde{h}}\,\tilde{\sigma}_{ij,j}, \quad \text{where } \tilde{\sigma}_{ij} = \overline{\sigma_{ik}^{\xi} K_{kj}} \tag{4.2}$$

is the effective mean stress tensor in GLM theory. In comparison, the expression for $\mathscr{F}_i$ is

$$\mathscr{F}_i = -\overline{\xi_{j,i} F_j'} = -\overline{\xi_{j,i} F_j^{\xi}} = -\overline{\xi_{j,i} (\sigma_{jk,k})^{\xi}/h^{\xi}} \tag{4.3}$$

which cannot, for general $\sigma_{ij}$, be written in conservative form. Now, the divergence of $\tilde{\sigma}_{ij}$ in (4.2) involves spatial derivatives of slowly varying mean fields, which are weak. Formally, if the gradients of unaveraged fields are $O(1)$, then the gradients of mean fields are $O(\mu)$, where $\mu \ll 1$ is a small parameter measuring the scale separation. Therefore, while $\overline{F}_i^L = O(\mu)$ arises due to mean-field variability on a slowly varying envelope scale, $\mathscr{F}_i$ arises due to processes in the 'bulk' of the disturbance, i.e. at $O(\mu^0)$. Hence it can be argued that the latter will dominate the former, provided that the scale separation is large enough.

Of course, the above argument is not a rigorous proof of the pseudomomentum rule under the stated conditions; indeed no such proof is possible for general $\sigma_{ij}$. Specifically, there are two clear exceptions to the rule: first, the effective mean stress tensor $\tilde{\sigma}_{ij}$ can contain stresses even in the absence of any disturbance. For instance, if $\sigma_{ij}$ is the usual viscous stress tensor, then $\tilde{\sigma}_{ij}$ contains viscous stresses due to mean-flow strains alone. The associated $\overline{F}_i^L$ is then a genuine part of the mean flow and need not be small compared to $\mathscr{F}_i$, especially if the disturbances are very weak. In such cases, the argument for the pseudomomentum rule holds only for the disturbance-dependent part of $\tilde{\sigma}_{ij}$ (cf. §4.3 below). The second exception arises should there be no $O(\mu^0)$ bulk contribution to $\mathscr{F}_i$. This could happen, for instance, if the disturbance fields in (4.3) are in quadrature. In this case the first contributions to both $\mathscr{F}_i$ and $\overline{F}_i^L$ arise at $O(\mu)$, and no general argument for the pseudomomentum rule can be made.

Apart from these exceptions, the generality of the argument is considerable. Specifically, no assumptions with regard to small amplitude, irrotational flow, near-plane monochromatic wave disturbances, etc. have been made. It would be very hard to verify (or even to formulate) the pseudomomentum rule in such generality using the standard Eulerian averaging as employed, e.g., in MN.

A further point with regard to practical applications can be noted in passing. The scale separation in any given application will only be finite, and hence the question arises whether the pseudomomentum rule holds to sufficient approximation for the application at hand. It is then useful to note that there is also a *qualitative* difference in appearance between the PV change in (3.16) due to the bulk-related $\mathscr{F}_i$ and that due to the envelope-related $\overline{F}_i^L$. The curl of these fields enters (3.16) and this gives the characteristic dipolar signature for the curl of $\mathscr{F}_i$, whereas the signature of $\nabla \times \overline{\boldsymbol{F}}^L$ is at least quadrupolar. This is a valuable diagnostic aid for numerical simulations (cf. figure 8 in § 5.3 below).

### 4.2. *Shock waves*

The argument of the previous section can be extended to apply to dissipative forces related to shock formation. To see this, consider the usual viscous stress tensor in two dimensions

$$\sigma_{ij} = \eta_1(u_{i,j} + u_{j,i} - \delta_{ij}u_{k,k}) + \eta_2\delta_{ij}u_{k,k}, \tag{4.4}$$

where the diffusivities $\eta_1$, $\eta_2 \geqslant 0$ to ensure that kinetic energy is always dissipated. It is now assumed in the usual way (cf. Whitham 1974) that any non-zero diffusivity maintains the differentiability of all flow fields, thus allowing the formal application of GLM theory, and that the standard jump conditions for mass and momentum fluxes at emerging near-discontinuities are independent of $\eta_1$, $\eta_2$ for sufficiently small values of these diffusivities. For any flow configuration it is then possible to imagine a sequence of flows with diminishing diffusivities, the limit of which is the appropriate 'weak' solution† of the ideal, non-dissipative equations. The pseudomomentum rule then holds for each member of the sequence and hence it presumably holds for the limiting weak solution as well. The overall conclusion is that the pseudomomentum rule holds for shocks on slowly varying mean flows, subject only to the general reservations noted in the previous section.

### 4.3. *Small-amplitude and slowly varying approximations*

Detailed mathematical calculations of wave–mean interactions are typically restricted to small disturbance amplitudes, as this allows detailed predictions to be made that can be tested, for instance, against numerical simulations. Furthermore, it is often possible at small amplitude to avoid using $\boldsymbol{\xi}$ explicitly in the diagnostic relationships needed for extracting various Lagrangian-mean quantities from the numerical data. Both reasons will come into play in § 5 below, and hence the necessary small-amplitude expressions are provided here.

All flow fields are expanded in the leading three powers of a suitable small-amplitude parameter $a \ll 1$. This produces an $O(1)$ background mean state, $O(a)$ disturbances on top of that background state, and an $O(a^2)$ mean-flow response to the disturbances. In the cases studied here, the background state is a rest state with depth $h_0$ defined such that the fluid surface height $h_0 + h_B$ is constant. The derivatives of the disturbance fields are assumed to be $O(1)$, such that, for instance, $\xi_{i,j} = O(a)$. Consistent with this, the spatial derivatives of mean fields are $O(\mu)$, where $\mu \ll 1$ is a suitable scale separation parameter. For example, $\overline{\xi_{i,j}u_j'} = O(a^2)$, but $(\overline{\xi_i u_j'})_{,j} = O(\mu a^2)$. No specific restriction is placed on $\mu/a$, and no scaling is specified for time derivatives.

† In the usual way (cf. Whitham 1974), a 'weak' solution satisfies the governing differential equations in smooth flow regions whereas it satisfies appropriate integral conservation laws across flow discontinuities such as shocks.

The Lagrangian disturbance $\phi^\ell$ of any field $\phi$ is approximated by (AM78*a*)

$$\phi^\ell = \phi' + \boldsymbol{\xi}\cdot\nabla\overline{\phi} + O(a^2) = \phi' + O(\mu a, a^2). \tag{4.5}$$

The Lagrangian-mean of $\phi$ can be approximated based on the relation $\overline{\phi}^L = \overline{\phi} + \overline{\phi}^S$, where $\overline{\phi}^S$ is the Stokes correction. At small amplitude $\overline{\phi}^S$ is (AM78*a*)

$$\overline{\phi}^S = \overline{\xi_j'\phi'_{,j}} + \tfrac{1}{2}\overline{\xi_i'\xi_j'}\,\overline{\phi}_{,ij} + O(a^3) = (\overline{\xi_j'\phi'})_{,j} - \overline{\xi'_{j,j}\phi'} + \tfrac{1}{2}\overline{\xi_i'\xi_j'}\,\overline{\phi}_{,ij} + O(a^3), \tag{4.6}$$

where $\xi_i'$ is the conventional notation for the $O(a)$ components of $\boldsymbol{\xi}$. Under the slowly varying assumption the term $-\overline{\xi'_{j,j}\phi'}$ dominates in (4.6). Now, the disturbance depth satisfies (cf. (3.8) in BM, or (5.16) in § 5.3 below)

$$h' + h_0\xi'_{j,j} = -\boldsymbol{\xi}'\cdot\nabla h_0 + O(a^2) = O(\mu a, a^2). \tag{4.7}$$

Substituting in (4.6) then yields

$$\overline{\phi}^S = \frac{\overline{h'\phi'}}{h_0} + O(\mu a^2, a^3), \quad \text{i.e. } \overline{\phi}^L \approx \overline{\phi} + \frac{\overline{h'\phi'}}{h_0} \tag{4.8}$$

for use as a Lagrangian-mean diagnostic of numerical data.

An Appendix shows how leading-order expressions for $\tilde{h} = \overline{h^\xi J}$ and $\boldsymbol{p}$ as defined in (3.15) can be derived in the same way. The result is

$$\tilde{h} = \overline{h} \quad \text{and} \quad \boldsymbol{p} = \overline{h'\boldsymbol{u}'}/h_0 = \overline{\boldsymbol{u}}^S \quad \text{to } O(a^2). \tag{4.9}$$

The surprisingly simple expression for $\boldsymbol{p}$ makes use of $\nabla\times\boldsymbol{u}' - fh'/h_0 = O(a^2)$. As noted in BM, the relation $\boldsymbol{p} = \overline{\boldsymbol{u}}^S$ does not generalize to other flow systems such as the Boussinesq system, for instance.

The use of $h_0$ in these expressions restricts their validity to cases where the mean state has only changed by an $O(a^2)$ amount. This is justified for an unforced initial-value problem, or in the early stages of a forced–dissipative problem.

For small-amplitude disturbances the disturbance-dependent parts of both $F_i$ and $\sigma_{ij}$ are $O(\tau^{-1}a)$, where $\tau$ is a typical time scale associated with the action of the force $F_i$. For instance, in the viscous case this would be the usual dissipation time scale. By comparison with (3.17), this scaling means that $\mathscr{F}_i$ is typically $O(\tau^{-1}a^2)$.

The numerical simulations in § 5 feature $O(a)$ waves that develop shocks on a nonlinear time scale $\propto 1/a$. To describe such $O(a)$ shocks properly requires following the evolution of the $O(a)$ fields over such a long time scale with a suitable singular perturbation theory, whose formidable complexity in the presence of rotation and bottom topography is not worth pursuing here. Instead, the above expressions will be used as a simplistic diagnostic tool for the numerical data even in the presence of shocks. This is not formally justifiable, but it is motivated by the observations that the amplitude of shock-forming fields (such as $h'$, for instance) remains $O(a)$ throughout, and that the scale separation between disturbance and mean fields is only increased by the shock formation. It is noteworthy that this simplistic approach affects only the numerical diagnostics; it does not affect the validity of the theory developed previously.

### 4.4. *Implications for direct numerical simulations*

The local conservation of momentum during small-scale dissipation in direct numerical simulations depends on the functional representation of the dissipative terms and on the numerical discretization technique used in the model. Dissipative terms in a numerical model that are to represent internal physical processes ought to be

momentum-conserving (i.e. they should be of the form (4.1)), which then also implies that the pseudomomentum rule should hold under the assumptions discussed above. (Of course, in some cases forcing terms in numerical models are meant to represent external momentum sources and then the foregoing does not apply, but these cases will not be discussed here.) In practice, this can be a fairly difficult test to pass.

To begin with, it is clear that *ad hoc* dissipative terms such as Rayleigh damping (in which $\boldsymbol{F} \propto -\boldsymbol{u}$) are not momentum-conserving. More subtly, the diffusion (or hyper-diffusion) that is sometimes applied directly to the variables $\{u, v\}$ (say, in a spectral model) is also not momentum-conserving, because it corresponds to (4.1) without the factor $1/h$. This leads to a leading-order disagreement with the pseudomomentum rule even at small disturbance amplitude $a$. This is despite the fact that $\mathscr{F}_i$ can be computed to $O(a^2)$ using only the $O(a)$ part of $F_i'$, which is $\sigma_{ij,j}'/h_0$. In other words, to $O(a^2)$ the neglect of $h'$ in the denominator does not affect the wave property $\mathscr{F}_i$, but it does affect $\overline{F_i}^L$.

Similar comments apply to local mass conservation; specifically any mass 'diffusion' or damping may invalidate the pseudomomentum rule. Finally, the numerical discretization method also affects the conservation properties. For instance, a spectral model might guarantee the global conservation of $h$ and $h\boldsymbol{u}$, but this might entail non-local redistribution of mass and momentum, which again invalidates the local conservation property on which the pseudomomentum rule is based. Of course, such non-local effects can be minimized by increasing the model resolution. Alternatively, finite-volume conservative schemes can be used (as in § 5 below), which are designed to achieve exact local conservation of mass and momentum at any resolution. However, at present the performance of typical finite-volume schemes deteriorates severely when applied to the strongly rotating, strongly dispersive systems of most interest in atmosphere and ocean fluid dynamics, which by itself can lead to substantially increased resolution requirements. In practice, this must be offset against the desirable in-built conservation properties of these schemes.

## 5. Numerical simulations

Nonlinear numerical simulations were performed to illustrate and verify the wave-induced vorticity transport described theoretically in the previous sections. Two basic cases were considered: first a freely decaying zonally symmetric wavetrain, and second a slowly varying wavetrain subject to continuous forcing and dissipation.

### 5.1. *Numerical model description*

The numerical model was a purpose-built shock-capturing finite-volume scheme for the rotating shallow-water system with bottom topography. The scheme uses the mass and momentum densities $\{h, hu, hv\}$ as flow variables in a conservative discretization of the equations of motion (cf. (2.1)–(2.2) for notation)

$$\frac{\partial h}{\partial t} + \nabla \cdot (h\boldsymbol{u}) = 0, \tag{5.1}$$

$$\frac{\partial (h\boldsymbol{u})}{\partial t} + \nabla \cdot \left( h\boldsymbol{u}\boldsymbol{u} + \boldsymbol{I}\frac{c_0^2}{2}h^2 \right) + \boldsymbol{f} \times h\boldsymbol{u} + hc_0^2 \nabla h_B = h\nabla \phi_w. \tag{5.2}$$

Here, provision has been made for including wave forcing by an irrotational body force with potential $\phi_w$. No explicit diffusion mechanism is present and hence the

only physical diffusion and dissipation can occur through shock formation. The computational domain is rectangular with side lengths $L$ and $D$ in the $x$ (zonal) and $y$ (meridional) directions. The domain is sub-divided into computational cells using a uniform Cartesian grid with constant grid spacings $\Delta x$ and $\Delta y$. The above equations contain conservative fluxes as well as momentum source terms due to (constant) rotation, topography, and wave forcing. Fluxes and sources are considered separately in the model. At each time step, the fluxes of mass and momentum between adjacent cells are calculated and the net change in mass and momentum in each cell is computed. The mass and momentum that leaves a cell through a particular cell interface necessarily enters an adjacent cell, which means that mass and momentum are always locally conserved by design. This discrete conservation property is known to be crucial for simulating gas-dynamical flows with shocks, and it also appears to be crucial for capturing wave–mean interaction effects according to the pseudomomentum rule, as was demonstrated in § 4.

The inter-cell fluxes are calculated using Roe's approximate one-dimensional Riemann solver at each cell interface. Nominal second-order accuracy in space is achieved by using a bilinear interpolation scheme (with limiter for stability) to calculate the two-dimensional gradients of each variable in each cell. The interfacial states that enter the Riemann solver are then calculated using these gradients. Fluxes through all interfaces are calculated simultaneously in this model, i.e. there is no operator splitting of the space dimensions. Second-order accuracy in time is achieved by using a Lax–Wendroff two-step explicit time stepping method and the source terms in (5.2) are incorporated at this stage.

The boundary conditions in the model can be periodic, resembling a solid wall, or resembling a radiation condition. The latter two are implemented by using a ring of ghost cells surrounding the computational domain. Each ghost cell contains either the mirror image (at a wall) or simply a copy (at a radiation boundary) of the state in the adjacent computational cell. These simplistic boundary procedures could be made more accurate if necessary, but no such need was found in the simulations reported here. All runs were performed on a workstation.

### 5.2. *Case I: freely decaying wavetrain*

This is a pure initial-value problem without bottom topography or forcing, i.e. $h_B = 0$ and $\phi_w = 0$ in (5.2). The initial conditions are given by a zonally symmetric train of gravity waves with Gaussian envelope shape in the meridional direction. The boundary conditions are periodicity in $x$ and radiation conditions at $y = 0$ and $y = D$. Specifically, the initial fields are

$$h = 1 + a\sin(kx)\left(1 + \frac{1}{k^2d^2}\left(2 - \frac{4}{d^2}y_c^2\right)\right)\exp(-y_c^2/d^2), \tag{5.3}$$

$$u = a\sin(kx)\left(\frac{\hat{\omega}}{k} - \frac{2f}{k^2d^2}y_c\right)\exp(-y_c^2/d^2), \tag{5.4}$$

$$v = -a\cos(kx)\left(\frac{f}{k} - \frac{2\hat{\omega}}{k^2d^2}y_c\right)\exp(-y_c^2/d^2), \tag{5.5}$$

introducing the non-dimensional wave amplitude $a$, wavenumber vector $\boldsymbol{k} = (k, 0)$, envelope scale $d$, centred $y$-coordinate $y_c = y - D/2$, and intrinsic frequency

$$\hat{\omega}(k) = +\sqrt{c_0^2\kappa^2 + f^2} \quad \text{where } \kappa \equiv |\boldsymbol{k}|. \tag{5.6}$$

| Parameter | Symbol | Case Ia non-rotating | Case Ib weakly rotating | Case Ic strongly rotating |
|---|---|---|---|---|
| Coriolis parameter | $f$ | 0 | 1/40 | 1 |
| Rossy deformation length | $L_R \equiv c_0/f$ | $+\infty$ | 40 | 1 |
| Wave amplitude | $a$ | 0.2 | 0.2 | 0.2 |
| Wavenumber magnitude | $\kappa$ | 1 | 1 | 1 |
| Meriodional envelope scale | $d$ | 40 | 40 | 40 |
| Slowly varying parameter | $\mu = (\kappa d)^{-1}$ | 0.025 | 0.025 | 0.025 |
| Domain size in $x$ and $y$ | $(L, D)$ | $(2\pi, 320)$ | $(2\pi, 320)$ | $(2\pi, 320)$ |
| Resolution | $(L/\Delta x, D/\Delta y)$ | (42, 54) | (42, 54) | (126, 162) |
| CFL number | $c_0\Delta t(1/\Delta x + 1/\Delta y)$ | 0.7 | 0.7 | 0.54 |

TABLE 1. Physical and numerical parameters for cases Ia–c. Note that the parameter $c_0 = L_R f$ was the same in all runs and that the numerical resolution was substantially increased in the strongly rotating case Ic.

The terms involving inverse powers of $d$ are necessary to ensure that there is no initial PV disturbance, i.e. that $q = f$ at $t = 0$. The wavenumber $k$ is chosen positive, which together with (5.6) corresponds to linear waves with positive phase and group velocities. No effort has been made to match nonlinear modifications to the wave structure at higher order in $a$.

Three runs with different values of the Coriolis parameter $f$ were performed and the relevant parameters are collected in table 1. A useful measure of the strength of the rotation in each case is the Rossby deformation length $L_R \equiv c_0/f$, above which rotation strongly influences the linear shallow-water dynamics. The length $L_R$ is formally infinite in the non-rotating case Ia, whilst $L_R = d$ in the weakly rotating case Ib, which implies that rotation affects the zonally averaged mean flow but not yet the small-scale gravity waves. Finally, $L_R = \kappa^{-1}$ in the strongly rotating case Ic, and hence the gravity waves themselves are strongly affected by the rotation.

The numerical grid was fine enough to render numerical diffusion unimportant during the integrations. In order to achieve sufficient accuracy of the PV field a finer grid was necessary in the strongly rotating case than in the other two cases. This was due to the numerical stiffness of the problem in this regime, in which source and flux terms must balance in (5.2) despite them being represented in completely different ways in the numerical scheme. The CFL number also had to be reduced in case Ic, as the waves developed substantial transient crests, with correspondingly increased nonlinear wave speeds (cf. figure 2).

As illustrated in figure 2, both the non-rotating case Ia and the weakly rotating case Ib exhibit rapid shock formation followed by sustained dissipative decay. In the strongly rotating case Ic, on the other hand, the wave steepens initially but then adjusts after a transient phase of dissipative decay into a steadily propagating nonlinear equilibrium shape, in which nonlinearity and dispersion balance. This is verified in figure 3, where the time evolution of the domain-integrated energy shows essentially constant amplitude at large times. It has been checked that the nonlinear equilibrium shape closely resembles the detailed structure of exact nonlinear periodic waves in rotating shallow water (e.g. Bühler 1993). Somewhat surprisingly, these nonlinear waves appear to be sufficiently stable against small-scale perturbations to emerge naturally in these simulations.

Let us now turn to the evolution of the mean flow. Mean fields are defined here by

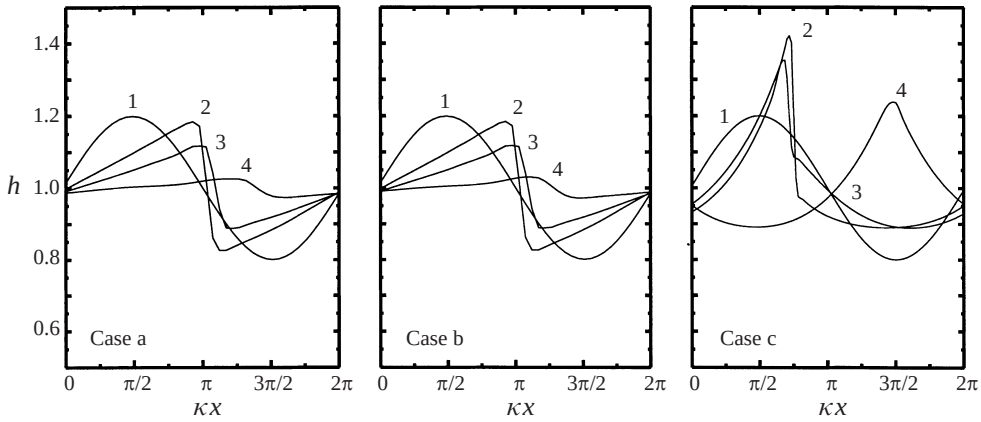


FIGURE 2. Snapshots of $h$ at $y = D/2$ at various times, showing the initial sinusoidal shape, shock formation, decay, and (in case Ic) adjustment to a nonlinear equilibrium shape. The four snapshots are denoted by 1–4, and the corresponding times are $t = (0, 2\pi, 4\pi, 20\pi)$ in Ia and Ib, and $t = (0, 1.4\pi, 2.8\pi, 5.6\pi)$ in Ic. Note that the amplitude of $h$ is steadily decreasing with time in Ia and Ib, whilst in Ic the second snapshot, which is just at the onset of breaking, has maximum amplitude.

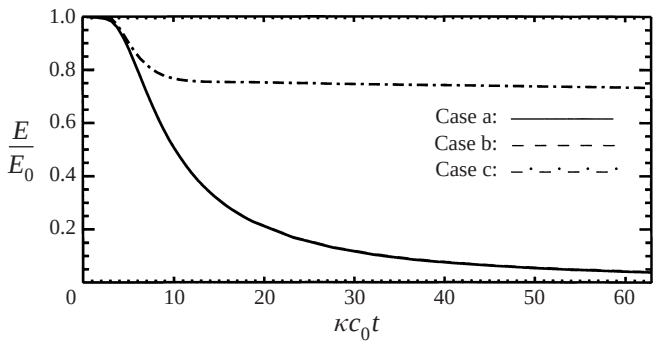


FIGURE 3. Plot of the time evolution of the domain-integrated energy $0.5(u^2 + v^2 + c_0^2(h-1)^2)$. The rapidly decaying curves for Ia–b fall on top of each other, while for Ic the curve shows initial decay followed by adjustment to a nearly steady energy level.

a zonal $x$-average, which implies that $\overline{\phi}_{,x} = 0$ for any field $\phi$. Zonal averaging is a valid limiting case of small-scale averaging, and hence all the previous theory applies here. The simple diagnostic expression (4.8) for Stokes corrections $\overline{\phi}^S \equiv \overline{\phi}^L - \overline{\phi}$ is used to compute $\overline{\phi}^L$, as are the expressions in (4.9) for $\tilde{h}$ and $\boldsymbol{p}$. Initial conditions for the relevant mean fields $\{\tilde{h}, \overline{\boldsymbol{u}}^L, \overline{q}^L, \boldsymbol{p}\}$ can now be diagnosed from (5.3). This gives $\tilde{h} = 1$, $\overline{\boldsymbol{u}}^L = \boldsymbol{p}_0$, $\overline{q}^L = f$, with

$$\boldsymbol{p}_0 = \frac{\boldsymbol{k}}{\hat{\omega}} E_0 = \frac{\boldsymbol{k}}{\hat{\omega}} \tfrac{1}{2} \overline{(\boldsymbol{u}'^2 + c_0^2 h'^2)}\big|_{t=0} = \boldsymbol{k} \tfrac{1}{2} a^2 \frac{\hat{\omega}}{k^2} \exp(-2y_c^2/d^2) \quad \text{to } O(a^2), \tag{5.7}$$

where $\boldsymbol{k} = (k, 0)$ and $E_0$ is the mean disturbance energy density per unit mass at $t = 0$. In summary, the initial mean flow consists of a fluid layer with uniform depth and uniform PV together with a zonal jet with Gaussian envelope scale $d/\sqrt{2}$ and maximum speed $0.5a^2\hat{\omega}/k$. Note that non-uniform $\overline{\boldsymbol{u}}^L$ is indeed compatible with uniform $\overline{q}^L$ in the presence of waves (cf. (3.14) and BM).

The time evolution of $\overline{q}^L$ under the pseudomomentum rule is governed by (3.16)

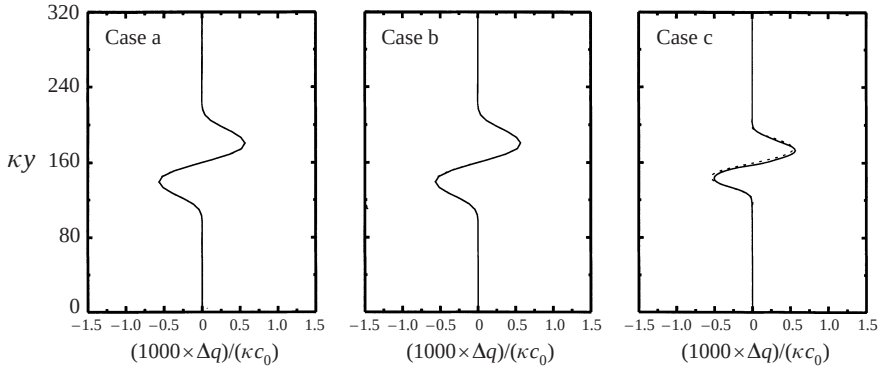


FIGURE 4. Plots of $\Delta q = \overline{q}^L - f$ as a function of $y$ at time $t = 6\pi$. The scaling factor $1000 = (\mu a^2)^{-1}$. The solid line is $\Delta q$ and the dashed line is $\mathsf{p}_{,y} - kE_{0,y}/\hat{\omega}$. Here, $\overline{q}^L$ has been diagnosed directly from $q = (\nabla \times \boldsymbol{u} + f)/h$ as $\overline{q}^L \approx \overline{q} + \overline{h'q'}$. The closeness of the solid and the dashed lines indicates the very good agreement with the pseudomomentum rule.

and (3.20), i.e.

$$\overline{q}^L_{,t} = \mathscr{F}_{,y} + O(a^3) \tag{5.8}$$

where $\mathscr{F}$ is the zonal component of the dissipative pseudomomentum source vector $\mathscr{F}_i$ in (3.17). The dissipative term $\mathscr{F}_{,y}$ arises through the dissipative shock formation and can only be diagnosed indirectly from the numerical simulations. To this end the $x$-component of (3.17) must be considered, i.e.

$$\mathsf{p}_{,t} - \mathscr{F} + O(a^3) = c_0^2\, \overline{\xi'_{j,x} h'_{,j}} = c_0^2\, (\overline{\eta'_{,x} h'})_{,y}, \tag{5.9}$$

where $\mathsf{p}$ is the zonal component of $\boldsymbol{p}$. The term on the right-hand side has been reformulated using $\overline{\phi}_{,x} = 0$ and (4.7); it describes a meridional flux of zonal pseudomomentum that results in the meridional spreading of the wavetrain. The scale of this term is at most $O(\mu a^2)$ and it will be neglected for simplicity.† Substituting (5.9) in (5.8) then yields $\overline{q}^L_{,t} = \mathsf{p}_{,yt}$, or

$$\boxed{\overline{q}^L - f = \mathsf{p}_{,y} - \frac{k}{\hat{\omega}} E_{0,y}}\,, \tag{5.10}$$

where $O(a^3)$ terms have been neglected and the mean-flow initial conditions have been used. This result shows that the shock-induced $\overline{q}^L - f = O(\mu a^2)$. The two sides of (5.10) are plotted in figure 4 and very good agreement between the numerical results and the theoretical predictions based on the pseudomomentum rule is found.

The evolution of $\overline{q}^L$ is essentially similar in all three cases, but this is not true for the evolution of $\overline{\boldsymbol{u}}^L$ due to the strong effect of background rotation when $L_R \leqslant d$. This can be demonstrated most easily by considering the GLM expression for $\overline{q}^L$ in (3.14), which gives

$$\overline{q}^L = -\overline{u}^L_{,y} + \mathsf{p}_{,y} + f - f(\tilde{h} - 1) + O(a^3) \tag{5.11}$$

valid throughout the evolution. Substituting from (5.10) gives

$$\overline{u}^L_{,y} + f(\tilde{h} - 1) = \frac{k}{\hat{\omega}} E_{0,y} \tag{5.12}$$

† Strictly speaking, this requires $\tau\mu \ll 1$, where $\tau$ is the dissipation time scale introduced in §4 such that $\mathscr{F} = O(\tau^{-1} a^2)$. This is a very good approximation for the strongly decaying cases Ia–b, but leads to a small error for the weakly decaying late-time evolution in case Ic.

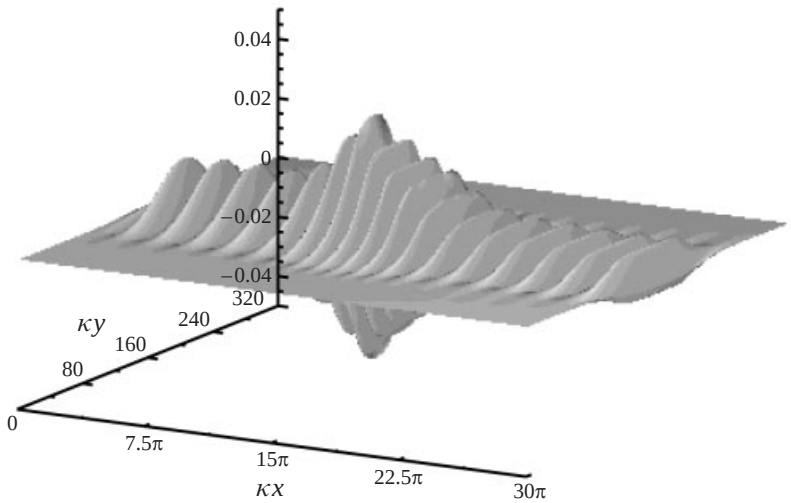


FIGURE 5. Snapshot of $u'(x, y)$ at $t = 60\pi$ in case IIa. Boundary conditions are solid walls at $x = 0$, $y = 0$, $y = D$, and radiation at $x = L$.

neglecting $O(a^3)$ terms as before.† This shows that in the non-rotating case Ia there is no change in $\overline{u}^L$, which is compatible with simplistic ideas about the wave's Stokes drift being transferred dissipatively to the Eulerian mean flow. In the rotating cases, however, $\overline{u}^L$ changes significantly as $f(\tilde{h}-1)$ changes and no simple relationship with the wave's Stokes drift holds. This emerges clearly in the long-time limit, when the $O(a)$ wave field as well as all $O(a^2)$ transients have settled down. Then, as investigated in detail in BM, a geostrophic balance relationship holds approximately between $c_0^2\tilde{h}_{,y}$, $f\overline{u}^L$, and (in case Ic) $E_{,y}$. This relation together with (5.12) allows $\overline{u}^L$ and $\tilde{h}$ to be calculated, which results in a $\overline{u}^L$ structure that is strongly influenced by rotation and quite different from the wave's Stokes drift.

### 5.3. *Case II: forced–dissipative wavetrain*

A wavetrain subject to continuous forcing and dissipation may produce a mean PV change that grows in time. This is demonstrated here by studying a slowly varying wavetrain of the form (5.3) that is continually forced near the left-hand $x$-boundary. The wavetrain then travels to the right until it encounters a strong decrease in background depth due to a hump-shaped topography in the centre region. The concomitant increase in wave amplitude pushes the waves over the breaking threshold, and shock-related dissipation occurs locally. Thereafter the topography reduces to zero again and a wavetrain with reduced amplitude propagates out of the right-hand $x$-boundary (see figure 5 for a wavetrain snapshot). The averaging operation is now a local average over the small-scale waves.

The (irrotational) wave-forcing and topography source terms in the momentum equations (5.2) do not affect the conservation of mass or PV. The topography term is also energy-conserving, as can be seen from the exact energy law

$$h\frac{\mathrm{D}}{\mathrm{D}t}\left\{\frac{\boldsymbol{u}^2}{2}+\frac{c_0^2}{2h}\Delta h^2\right\}+\nabla\cdot\left(c_0^2\,\Delta h\,h\boldsymbol{u}-\frac{c_0^2}{2}\Delta h^2\boldsymbol{u}\right)=0, \tag{5.13}$$

† For diagnostic purposes $\tilde{h}$ can be replaced by $\overline{h}$ to $O(a^2)$, although in case Ic this is not strictly justified because $f(\tilde{h}-\overline{h}) = O(f\mu a^2)$ is then not strictly negligible in (5.12).

where the wave-forcing term has been omitted. Here $\Delta h$ is the deviation of $h$ from a variable background $h_0(\boldsymbol{x})$ that corresponds to constant surface height, i.e.

$$\Delta h \equiv h - h_0 \quad \text{and} \quad h_0 + h_B = 1. \tag{5.14}$$

In the usual way, the energy density in curly brackets in (5.13) measures the energy of the present state relative to an equal-mass rest state with an undisturbed surface.

It turns out to be extremely useful to consider first the unforced linearized equations in one dimension in the usual ray-tracing JWKB approximation. This will fix the appropriate definitions of $O(a)$ wave amplitudes, it will allow setting parameters for the amplitude increases due to $h_B > 0$ in the centre, and it will also permit a simple diagnostic for $\mathscr{F}_i$ to be defined. The one-dimensional linear equations are obtained from (5.1)–(5.2) by setting all $y$-derivatives to zero, letting

$$\boldsymbol{u} = \boldsymbol{u}' + O(a^2) \quad \text{and} \quad \Delta h = h' + O(a^2), \tag{5.15}$$

and collecting terms at $O(a)$. The result is

$$h'_{,t} + (h_0 u')_{,x} = 0, \tag{5.16}$$

$$u'_{,t} - fv' + c_0^2 h'_{,x} = 0 \quad \text{and} \quad v'_{,t} + fu' = 0, \tag{5.17}$$

where $h_0(x)$ enters only in the continuity equation (5.16). The mean disturbance energy density per unit mass at $O(a^2)$ can be read off from (5.13) as

$$E = \frac{1}{2}\left(\overline{\boldsymbol{u}'^2} + \frac{c_0^2}{h_0}\overline{h'^2}\right). \tag{5.18}$$

The local average is such that $\overline{h_0} = h_0$ to sufficient approximation. The expression in (4.9) is used for $\mathsf{p}$, i.e.

$$\mathsf{p} = \overline{h'u'}/h_0 \tag{5.19}$$

to $O(a^2)$. Using (5.18)–(5.19) in (5.13) then gives the energy conservation law

$$(h_0 E)_{,t} + (c_0^2 h_0{}^2 \mathsf{p})_{,x} = 0 \tag{5.20}$$

at leading order. The system can be closed by making the JWKB assumption, which assumes a slowly varying wavetrain containing small-scale sinusoidal waves moving to the right with local intrinsic frequency $\hat{\omega} > 0$ and local zonal wavenumber $k > 0$ that satisfy the local dispersion relation

$$\hat{\omega} = +\sqrt{c_0^2 h_0 k^2 + f^2}. \tag{5.21}$$

The $O(1)$ background state (i.e. $h_0(x)$) is steady but $x$-dependent and correspondingly

$$\hat{\omega} = \text{const.} \quad \text{and} \quad h_0(x)k^2 = \text{const.} \tag{5.22}$$

along group-velocity rays. Assuming a single-frequency wave source then implies that $\hat{\omega}$ and $h_0(x)k^2$ are the same across all rays and hence everywhere in the domain. The generic relation $\mathsf{p} = kE/\hat{\omega}$ holds for propagating waves and using this relation, (5.20), and (5.22) in combination gives the pseudomomentum evolution equation

$$\boxed{(h_0 \mathsf{p})_{,t} + (h_0 E[1 - f^2/\hat{\omega}^2])_{,x} = -h_{0,x}\frac{E}{2}[1 - f^2/\hat{\omega}^2]} \tag{5.23}$$

at leading order. The specific form of (5.23) is useful for the numerical diagnostics described below but the usual ray-tracing form is recovered by noting that the group

| $\hat{\omega}$ | $k$ | $p$ | $E$ | $\lvert u'\rvert$ | $\lvert h'\rvert$ | $a$ |
|---|---|---|---|---|---|---|
| const. | ${h_0}^{-1/2}$ | ${h_0}^{-2}$ | ${h_0}^{-3/2}$ | ${h_0}^{-3/4}$ | ${h_0}^{-1/4}$ | ${h_0}^{-5/4}$ |

TABLE 2. Steady-state scalings of local wavetrain parameters with $h_0(x)$. For example, if $h_0$ decreases from $h_0 = 1$ to $h_0 = \frac{1}{3}$, then the local pseudomomentum density $p$ increases ninefold. The scaling $k \propto {h_0}^{-1/2}$ is also the scaling of the amount of pseudomomentum contained between two points moving with the local group velocity, which shows that pseudomomentum is not conserved unless $h_0$ is constant. Note that $|u'|$ and $|h'|$ stand for local amplitudes in these fields. The wave amplitude $a$ is defined as $a = \max(|h'|/h_0)$, which in JWKB theory implies $a = \max(|u'|k/\hat{\omega})$ and $a = \sqrt{2E}k/\hat{\omega}$.

| Parameter | Symbol | Case IIa non-rotating | Case IIb weakly rotating | Case IIc strongly rotating |
|---|---|---|---|---|
| Coriolis parameter | $f$ | 0 | $\frac{1}{40}$ | 1 |
| Wave amplitude | $a$ | 0.02 | 0.02 | 0.1 |
| Wavenumber magnitude | $\kappa$ | 1 | 1 | 1 |
| Meriodional envelope scale | $d$ | 40 | 40 | 40 |
| Maximum of $h_B(x)$ | | $\frac{2}{3}$ | $\frac{2}{3}$ | $\frac{2}{3}$ |
| Maximum of $\lvert \mathrm{d}h_B(x)/\mathrm{d}x\rvert$ | | 0.056 | 0.056 | 0.056 |
| Domain size in $x$ and $y$ | $(L, D)$ | $(30\pi, 320)$ | $(30\pi, 320)$ | $(30\pi, 320)$ |
| Resolution | $(L/\Delta x, D/\Delta y)$ | (856, 76) | (856, 76) | (856, 76) |
| CFL number | $c_0\Delta t(1/\Delta x + 1/\Delta y)$ | 0.87 | 0.87 | 0.7 |

TABLE 3. Physical and numerical parameters for cases IIa–c. The shape of $h_B(x)$ is shown in figure 6 and the boundary conditions are described in the caption of figure 5. Because of prohibitive computational cost, it was not possible to use the substantially increased resolution of case Ic in case IIc here.

velocity $c_g$ satisfies $c_g k/\hat{\omega} = [1 - f^2/\hat{\omega}^2]$. Equation (5.23) makes clear that pseudomomentum is created wherever the background depth $h_0(x)$ decreases, and vice versa.

The local wavetrain parameters change from place to place as $h_0(x)$ changes. Specifically, for a steady wavetrain the scalings in table 2 have been deduced from the above JWKB relations. The ratio $f/\hat{\omega}$ is constant, and hence the relative importance of rotation for the waves is unaffected by changes in $h_0$. It is noteworthy that $a$, $|u'|$, and $|h'|$ all scale differently as $h_0$ changes, with the wave amplitude $a$ showing the greatest sensitivity. This is important for judging the strengthening of nonlinear effects as $h_0$ is decreased, because these effects scale with $a$.

Let us return now to the two-dimensional set-up. Three cases IIa–c have been studied and the parameters are summarized in table 3. The type of wave-forcing potential $\phi_w$ used was the same as in BM, i.e. for numerical robustness a wavetrain moving in both $x$-directions was generated continually. The forcing was centred half a wavelength away from the left-hand solid-wall boundary, which led to a reflection of the left-going waves and their desired constructive interference with the right-going waves in the bulk of the domain. Various centreline sections of the wavetrain are presented in figures 6 and 7 (see captions for details). A simple $x$-average over one wavelength has been used to calculate the mean fields. The wave amplitude $a$ was chosen such that shock formation was localized in the centre region, i.e. the waves were able to travel to the centre before breaking. Note that in the strongly rotating case IIc, $a$ had to be chosen five times larger than in the other cases due to the resilience to breaking of strongly rotating inertia–gravity waves.

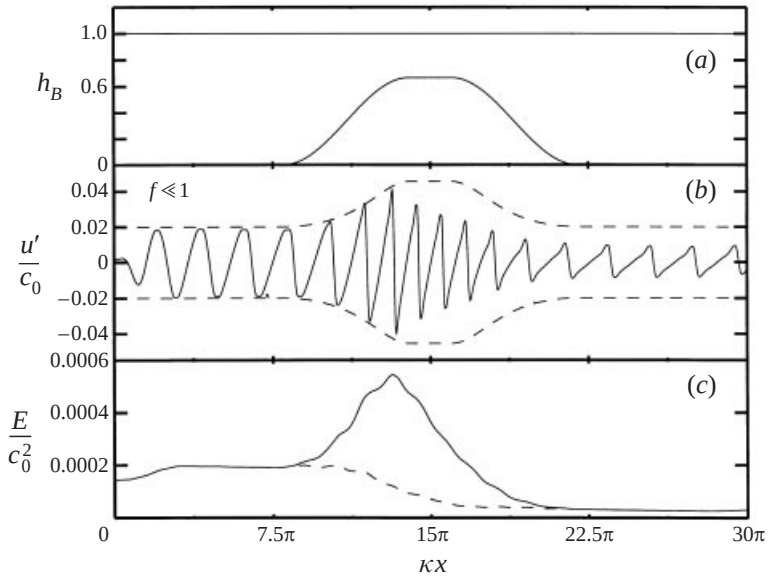


FIGURE 6. Snapshot along centreline $y = D/2$ at $t = 60\pi$ in case IIa, whose disturbance fields are virtually identical to case IIb. (*a*) Hump-shaped bottom topography $h_B(x)$, which is raised over three wavelengths to its centre maximum $h_B = \frac{2}{3}$ (corresponding to $h_0 = \frac{1}{3}$); also shown is the undisturbed free surface such that $h_0$ is the distance between the two curves. (*b*) Disturbance velocity $u'$ showing slow initial steepening, rapid amplitude increase, shock formation, and decay to reduced amplitude. The dashed line is the $u'$ amplitude scaling based on (non-dissipative) JWKB theory. (*c*) Corresponding disturbance energy density $E$. The dashed line is $Eh_0{}^{3/2}$, which is constant as long as (non-dissipative) JWKB theory is valid.

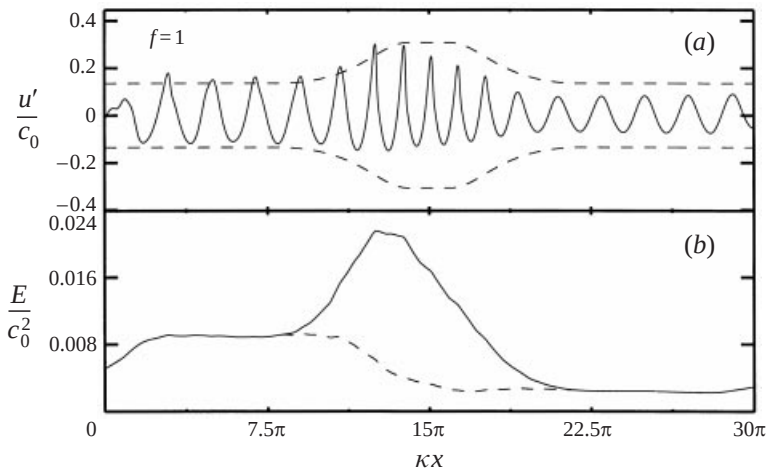


FIGURE 7. Snapshot along centreline $y = D/2$ at $t = 60\pi$ in case IIc, using the same bottom topography as in IIa,b but note the increased wave amplitude. (*a*) Disturbance velocity $u'$ showing an initial peak–trough asymmetry characteristic of large-amplitude inertia–gravity waves. The dashed line is the $u'$ amplitude scaling based on (non-dissipative) JWKB theory and an initial amplitude $a = \sqrt{2E}k/\hat{\omega} \approx 0.1$, which gives $\max|u'| \approx 0.14$. (*b*) Corresponding disturbance energy density $E$. The dashed line is $Eh_0{}^{3/2}$, which is constant as long as (non-dissipative) JWKB theory is valid.

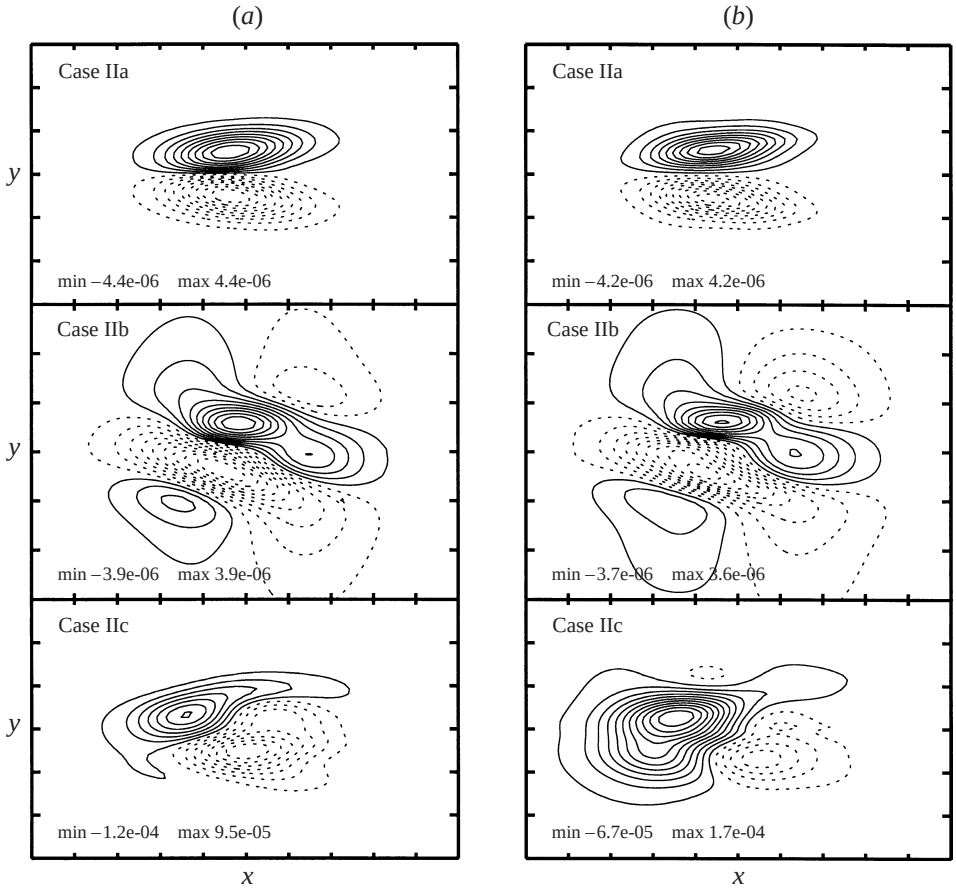


FIGURE 8. Snapshots of $\overline{q}^L_{,t}$ at $t = 60\pi$. A centred region ten wavelengths wide in $x$ and six envelope scales $d$ wide in $y$ is shown. Solid and dashed contours mark positive and negative values, respectively. (*a*) $\overline{q}^L_{,t}$ diagnosed based on the difference of $\overline{q}^L$ at $t = 80\pi$ and $t = 40\pi$. (*b*) $\overline{q}^L_{,t}$ estimated from the pseudomomentum rule and plotted with the same contour levels as the diagnosed $\overline{q}^L_{,t}$; see text for details.

Having discussed the essentially one-dimensional disturbance dynamics along the centreline, it is now possible to turn to the mean-flow dynamics. In the early stages the forced–dissipative flows studied here are amenable to simple $O(a^2)$ diagnostics of the type introduced earlier and used for the initial-value problems Ia–c. The flow starts from rest and hence the initial mean fields $\{\tilde{h}, \overline{\boldsymbol{u}}^L, \overline{q}^L, \boldsymbol{p}\}$ are $\tilde{h} = h_0$, $\overline{\boldsymbol{u}}^L = 0$, $\overline{q}^L = f/h_0$, and $\boldsymbol{p} = 0$. Note that $\overline{q}^L$ varies due to $h_0(x)$ in the rotating cases IIb,c; indeed $\overline{q}^L_{,x} = O(\mu f)$. This means that material advection of background PV contributes to $\overline{q}^L_{,t}$ on top of the wave-breaking contributions summarized by the pseudomomentum rule. The leading-order evolution equation for $\overline{q}^L$ hence reads

$$\overline{q}^L_{,t} + \overline{u}^L \left(\frac{f}{h_0}\right)_{,x} = \frac{\mathscr{F}_{,y} - \mathscr{G}_{,x}}{h_0} + O(a^3) \tag{5.24}$$

where $(\mathscr{F}, \mathscr{G})$ are the components of $\mathscr{F}_i$. It is clear that $|\mathscr{G}| \ll |\mathscr{F}|$ due to the $y$-symmetry of the wavetrain, and hence it only remains to find a diagnostic expression for $\mathscr{F}$. However, using the $x$-component of (3.17) as before brings no direct result here due to various $\overline{(\ldots)}_{,x}$ terms in that equation. Those terms vanished under zonal

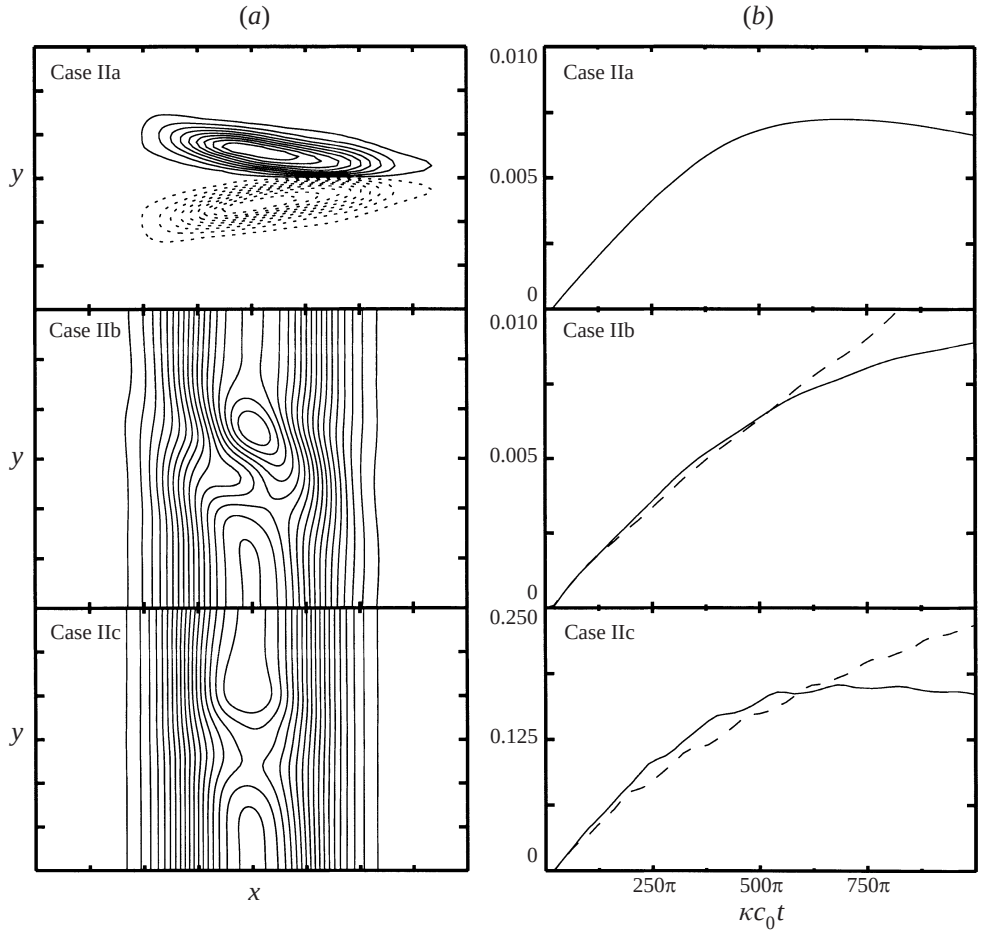


FIGURE 9. (*a*) Snapshots of $\overline{q}$ at large time, i.e. at $t = 1000\pi$. The same centre region as in figure 8 is shown. In the rotating cases IIb,c the background PV has been significantly affected by material displacements, which breaks the *y*-symmetry observed in IIa. The reason for the asymmetry is that the wave breaking occurs at a maximum of PV and hence advection of PV can only deepen negative anomalies. (*b*) Time evolution of extremal values of the PV anomaly $\overline{q} - f/h_0$. The solid line shows maximal values and the dashed line shows (sign-reversed) minimal values. Again, cases IIb,c become asymmetrical at large times, with dominant minimal values.

averaging, but they persist under the small-scale averaging used here. Evaluating these terms would involve $\boldsymbol{\xi}'$, which is not directly available from the numerics. Therefore, a simplistic diagnostic approach based on JWKB theory is adopted here, in which $\mathscr{F}$ is defined as the residual of the steady pseudomomentum equation (5.23), i.e.

$$(h_0 E[1 - f^2/\hat{\omega}^2])_{,x} + h_{0,x}\frac{E}{2}[1 - f^2/\hat{\omega}^2] = h_0\mathscr{F}, \tag{5.25}$$

or

$$\frac{1}{{h_0}^{3/2}}({h_0}^{3/2} E[1 - f^2/\hat{\omega}^2])_{,x} = \mathscr{F}. \tag{5.26}$$

In JWKB theory this would be justified as equating $\mathscr{F}$ to the appropriately scaled pseudomomentum flux convergence. While JWKB theory is clearly not formally applicable here, the fact remains that $\mathscr{F}$ is indeed equal to an appropriate pseudo-

momentum flux convergence, even outside JWKB theory. Equation (5.26) is used as a simplistic first approximation to this flux convergence.

The advective term in (5.24) requires $\overline{u}^L$, which is diagnosed as $\overline{u}^L = \overline{u} + \overline{h'u'}/h_0$ to $O(a^2)$. Figure 8 compares the directly diagnosed $\overline{q}^L_{,t}$ (using $\overline{q}^L = \overline{q} + \overline{h'q'}/h_0$ at different times; see caption) to a prediction based on (5.24) at $t = 60\pi$. It can be seen that the agreement is very good in IIa,b, but markedly worse in case IIc. With considerable effort, this has be traced back to poor convergence of $\overline{u}$ in the advective term in (5.24). There are two reasons for this: first, the $u'$ field shows signs of sub-harmonic pairing in the wavetrain, which contributes to a 'blotchy' $\overline{u}$. Second, in case IIc the Rossby deformation length $L_R \ll d$ and hence the non-dissipative wave-induced $\overline{u}^L$ ought to be very small compared to $\overline{u}^S$ (cf. § 8.2 in BM for an analysis). This means that $\overline{u}^L$ is the small difference between two large quantities, $\overline{u}$ and $-\overline{u}^S$, which leads to further diagnostic difficulties. The localized wave breaking relies on strong $h_{0,x}$, which invariably implies large $\overline{q}^L_{,x}$ if $f = 1$, and hence the advective term in (5.24) is invariably important in this scenario. Hence, it appears that the pseudomomentum rule can be demonstrated clearly in IIa,b, but that it cannot be demonstrated beyond all doubt in IIc. Significantly more costly numerical simulations, beyond the scope of this study, appear to be necessary to test clearly the pseudomomentum rule in the strongly rotating case. Still, I cannot think of a theoretical reason why the pseudomomentum rule should fail there.

Figure 9 illustrates the large-time behaviour with snapshots of $\overline{q}$ at time $t = 1000\pi$. The rotating cases IIb,c show marked $y$-asymmetries at this time, which is commented on in the caption. Also shown there is a time series of the maxima and minima of the created PV anomalies, which in all cases shows cumulative growth in the early stages, which is modified at larger times when the created PV anomalies affect the dynamics; again see caption.

Finally, it can be noted that the dissipative changes in $\overline{q}^L$ arising in these simulations are linked in an essential way to the lateral envelope shape of the wavetrain. In genuinely one-dimensional simulations, as described recently by Kuo & Polvani (1999), the lateral envelope shape is effectively infinite and no such dissipative change in $\overline{q}^L$ occurs.

## 6. Comparison with related weakly nonlinear asymptotic theories

The results described here and in BM can be compared with results that have been obtained elsewhere using asymptotic theories of weakly nonlinear wave–mean interactions. This cross-comparison is helpful because it provides an asymptotic check on both the theory and the numerics presented here and it also illuminates certain aspects of both theories that would otherwise be perhaps less obvious.

The attraction of weakly nonlinear asymptotic theory lies in the fact that, within the restrictions of the chosen asymptotic scaling regime, it can produce results with rigorous mathematical error control. For an unforced flat-bottom shallow-water system on a periodic domain, such a theory has recently been investigated independently by two groups: Babin, Mahalov & Nicolaenko (1997) and references therein, and Majda & Embid (1997) and references therein. (Both groups have extended their approach to the significantly more complicated case of three-dimensional stratified flow.) The brief pedagogical account of the basic results in Majda & Embid (1997) is sufficient for this section. The set-up of the theory involves an asymptotic sequence

of flows ordered by an asymptotic parameter $\epsilon \ll 1$ such that†

$$c_0,\, f \propto \epsilon^{-1} \quad \text{and} \quad L_R = c_0/f = \text{const.} \tag{6.1}$$

The initial conditions are independent of $\epsilon$ provided that the flow is described using the variables $U = \{u, v, c_0 \Delta h\}$, which implies that $\Delta h = O(\epsilon)$. This down-scaling of the depth disturbance $\Delta h$ is necessary to render the initial flow energy in (5.13) $\epsilon$-independent. After suitable non-dimensionalization, the formal structure of the equations of motion can be usefully summarized as (cf. (10) in Majda & Embid 1997)

$$\frac{\partial U}{\partial t} + \frac{1}{\epsilon}\mathscr{L}(U) + \mathscr{B}(U, U) = 0, \quad U(x, y, 0) = O(1), \quad t \leqslant O(1), \tag{6.2}$$

where $\mathscr{L}$ is a linear operator with constant coefficients that describes the usual '$2+1$' dynamics of two gravity-wave modes and one steady balanced mode, and $\mathscr{B}$ contains the nonlinear terms. The structure in (6.2) makes obvious that $\epsilon \ll 1$ corresponds to a weakly nonlinear asymptotic regime, in which the magnitude of the nonlinear terms is formally $O(1)$, but is $O(\epsilon)$ when compared to the magnitude of the linear terms. The theory proceeds by decomposing (6.2) into spectral modes, introducing fast and slow time scales to capture linear and nonlinear terms, and solving the resultant set of equations with rigorous error control for $t \leqslant O(1)$. The technically demanding solution procedure involves averaging over the fast, linear time scale whilst taking proper account of all possible cubic resonant interactions between spectral modes. The validity for times up to $t = O(1)$ corresponds to an asymptotically large number of linear gravity-wave oscillations (whose frequency is $\hat{\omega} \propto \epsilon^{-1}$; cf. (5.6)), and to a fixed nonlinear time span of, say, one vortex turnover time.

The asymptotic results for the mean flow and the gravity waves are as follows. The mean flow exhibits standard quasi-geostrophic dynamics, i.e. the mean PV is advected by a non-divergent mean velocity field that is in geostrophic balance with the mean depth field. The functional relation between the mean PV and the other mean fields is that of the linearized steady balanced mode, which implies, e.g., the absence of cyclostrophic corrections to the structure of vortices. Most notably, the mean flow is completely unaffected by the gravity waves as $\epsilon \to 0$. As for the gravity waves themselves, there is no nonlinear inter-scale dynamics between modes with different wavenumber vector magnitudes $\kappa$. The only nonlinear dynamics that does occur produces energy exchanges within 'shells' of constant $\kappa$, for which the mean flow is acting as a catalyzer. By implication, shock formation is impossible for $t \leqslant O(1)$.‡ All the usual JWKB effects (e.g. focusing, Doppler shifting) are negligible in the asymptotic limit $\epsilon \to 0$.

How do these results compare with the wave–mean theories studied here and in BM? To begin with, the averaging operators of the different theories agree when applied to wavetrains containing small-scale waves: they simply reduce to the average over the rapidly varying phase of the waves. Now, discarding the Stokes corrections for the moment, (3.14) shows that the functional relation between mean PV and the other mean fields contains wave-related terms such as $\nabla \times \boldsymbol{p}$. The same occurs in BM's explicit JWKB result (1.3), where the $O(a^2)$ balance relation contains $\nabla \times \boldsymbol{p}$ and a further term $\propto E/(L_R c_0)$ due to $O(a^2)$ changes in the mean depth. At first sight, this appears to disagree with the aforementioned asymptotic mean-flow results, in

† This is what is called the 'low Froude, low Rossby number' in Majda & Embid (1997); the alternative 'low Froude, fixed Rossby number' limit is not considered here.

‡ According to Babin *et al.* (1997) an even stronger regularity result can be proven, albeit subject to certain technical restrictions: shock formation is impossible for time intervals that go to infinity as $\epsilon \to 0$.

which no such terms were present. However, it turns out that the pseudomomentum $\boldsymbol{p} = O(\epsilon)$, and hence its presence in (3.14) is indeed negligible for $t \leqslant O(1)$. This scaling can be read off, for instance, from the JWKB relation $\boldsymbol{p} = \boldsymbol{k}E/\hat{\omega}$ and by noting that $\hat{\omega} \propto \epsilon^{-1}$ whilst $\boldsymbol{k}$ and $E$ are $\epsilon$-independent. The same argument also shows that the term $\propto E/(L_R c_0) = O(\epsilon)$. Another way to derive this scaling, not using JWKB theory, is via (4.9) and recalling that $h' = O(\epsilon)$. The same argument also shows that all Stokes corrections are in fact $O(\epsilon)$, and hence that Eulerian and Lagrangian averaging must produce the same result asymptotically. This justifies the previous discarding of Stokes corrections in (3.14).

Most fundamentally, $h' = O(\epsilon)$ implies that $\boldsymbol{\xi} = O(\epsilon)$. Therefore, in the limit $\epsilon \to 0$ both the disturbance-associated displacement field $\boldsymbol{\xi}$ and the pseudomomentum field $\boldsymbol{p}$ go to zero. Physically, the fluid layer becomes increasingly 'stiff' as $\epsilon$ is decreased, and material oscillations must hence reduce in amplitude for any fixed energy level. This also explains why the asymptotic mean-flow dynamics can be initialized by simply taking the initial PV as the initial mean-flow PV, a procedure that neglects undulations of PV contours by gravity waves: such undulations are again described by $\boldsymbol{\xi}$ and are hence $O(\epsilon)$.

The restriction of the asymptotic results to $t \leqslant O(1)$ cannot in general be relaxed to $t \leqslant O(1/\epsilon)$, which would cover many nonlinear vortex turnover times. For instance, the various wave-related $O(\epsilon)$ terms described above would become important over such a long time period. BM's example of a resonantly growing Rossby wave driven by a gravity-wave wavetrain demonstrates this point. As is well-known, other $O(\epsilon)$ effects not related to gravity waves (such as cyclostrophic corrections to vortex structures) are also important over such long time periods.

Finally, the asymptotic theory makes an interesting prediction for the shock formation exhibited in the numerical simulations here: the asymptotic theory forbids shock formation for $t \leqslant O(1)$ even for small-scale gravity waves satisfying $\kappa L_R \gg 1$. This is despite the fact that the linear structure of such waves is virtually unaffected by rotation. This asymptotic result clearly does not yet apply to the waves in the weakly rotating case Ib in § 5. However, for fixed initial conditions, $\epsilon$ can always be reduced sufficiently far to reach the asymptotic regime. Numerical experimentation for case Ib has shown that $c_0$ and $f$ need to be increased by a factor of $\approx 1500$ whilst keeping the initial $\{u', v', c_0 h'\}$ constant before the dispersion begins to suppress the shock formation significantly. This suggests that for these wavelike initial conditions $\epsilon \approx \max|h'|(\kappa L_R)^2$, which gives $\epsilon \approx 0.2$ after the increase of $c_0$ and $f$.

## 7. Concluding remarks on atmospheric applications

Work is currently underway to adapt the present shallow-water theory to three-dimensional stratified flow systems. This is not an entirely straightforward process, if only because of the presence of diabatic heating effects in stratified flows. Such diabatic effects are important for geophysical applications: for instance, gravity-wave dissipation by diabatic heating contributes significantly to the gravity-wave budget of the middle atmosphere (e.g. MN). There are also further technical complications because diabatic heating interferes with the kinematical workings of finite-amplitude GLM theory (cf. BM).

Certain broad conclusions for the numerical treatment of atmospheric gravity waves can already be drawn from the present work. High-resolution general circulation models (GCMs) for the atmosphere are becoming increasingly 'gravity-wave-permitting', meaning that more and more small-scale gravity-wave activity is becoming part of

the internal dynamics of these models. Therefore, the numerical issues discussed in § 4.4 are becoming to some extent relevant for GCM design.

Specifically, sufficient numerical accuracy in local mass and momentum conservation will be required to obtain correct wave–mean interaction results in GCMs. This holds for a wider range of wave–mean interactions than just the pseudomomentum rule. One example is the so-called Lighthill radiation, or spontaneous emission of gravity waves by unsteady vortex motion. It is well known theoretically that Lighthill radiation has a comparatively weak quadrupolar source type in the standard compact source limit, but this result depends crucially on the local conservation of mass and momentum and hence non-conservative numerical models might show spuriously strong Lighthill radiation.

Despite the growing importance of direct numerical simulations of gravity waves, GCMs will continue for a long time to rely heavily on the so-called parametrization of gravity waves, in which the effect of unresolved gravity waves is approximately represented in the model via suitable body forces. Current parametrization schemes are based on fluxes of pseudomomentum derived from one-dimensional JWKB theory for the waves and on zonally symmetric mean flows for the interaction theory. The theory developed here might help in devising parametrization schemes that go beyond these restrictions in a consistent way, especially if the generation of pseudomomentum by spatial mean-flow variability is to be taken into account.

It is important in this connection that the effective mean force $\tilde{\boldsymbol{F}}$ in the vorticity budget (3.16) is only part of the mean momentum budget. Other, irrotational mean forces exist that complete the mean momentum budget, unless there are special mean field symmetries (AM78*a*). These irrotational mean forces are not in general amenable to a simple analysis of the type that has led to the pseudomomentum rule. Still, the existence of these irrotational mean forces implies, for instance, that there is no obvious bound on $|\tilde{\boldsymbol{F}}|$ that can be derived from global momentum conservation. If the pseudomomentum rule applies, then this is consistent with the fact that pseudomomentum can be freely generated by spatial mean-flow variability, thereby changing the effective mean force that is available for dissipative mean vorticity transport. For example, the effective mean force in case II of the numerical simulations was a factor of $\sqrt{3}$ larger than would have been expected by simply comparing the incoming and outgoing fluxes of pseudomomentum (cf. table 2).

Part of this work was supported by the Engineering and Physical Sciences Research Council under grant GR/K75583. This general problem area was first pointed out to me by M. E. McIntyre, and it is a pleasure to acknowledge many stimulating conversations in this connection. Further stimulating correspondence on breaking waves with O. Bokhove and A. C. Kuo as well as further fluid-dynamical input by JBB near the end of this study are also gratefully acknowledged.

## Appendix

### A.1. *Proof of lemma* (3.12)

Consider a closed contour $C^\xi$ and the circulation of $\boldsymbol{B}$ around it, which is given by

$$\Gamma = \oint_{C^\xi} \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{x} = \oint_C (\boldsymbol{B} \cdot \mathrm{d}\boldsymbol{x})^\xi. \tag{A 1}$$

The second integral is around the contour $C$ composed of the mean positions belonging to the particles on $C^\xi$ (cf. figure 4 in BM). The equality of both integrals follows from the properties of the lifting map $(\ldots)^\xi$, which provides the appropriate

transformation of the integrand. Averaging the second integral (which only involves the integrand because $C$ is a mean contour), using the definition of $\tilde{\boldsymbol{B}}$, and applying Stokes's theorem then gives

$$\overline{\Gamma} = \int\int_A \nabla \times \tilde{\boldsymbol{B}}\, \mathrm{d}x\, \mathrm{d}y, \tag{A 2}$$

where $A$ is the area enclosed by $C$. On the other hand, Stokes's theorem applied directly to the first integral in (A 1) results in

$$\Gamma = \int\int_{A^\xi} \nabla \times \boldsymbol{B}\, \mathrm{d}x\, \mathrm{d}y = \int\int_A (\nabla \times \boldsymbol{B})^\xi (\mathrm{d}x\, \mathrm{d}y)^\xi \tag{A 3}$$

$$= \int\int_A \left(\frac{\nabla \times \boldsymbol{B}}{h}\right)^\xi \tilde{h}\, \mathrm{d}x\, \mathrm{d}y, \tag{A 4}$$

where $A^\xi$ is the area enclosed by $C^\xi$ and where the last integral uses the definition of $\tilde{h}$. Averaging the last integral gives another expression for $\overline{\Gamma}$, and demanding that both expressions must agree for arbitrary $C^\xi$ then implies the second equality in (3.12).

### A.2. *Approximate expressions for $\tilde{h}$ and $\boldsymbol{p}$*

Consider an $O(1)$ background state with zero velocity and a slowly varying mean depth $h_0$ satisfying $h_0 + h_B = 1$. It is assumed that disturbances have small amplitude $a \ll 1$ and that all mean-field derivatives are weak, as measured by $\mu \ll 1$. The following two integrals of the disturbance continuity equation and of the disturbance vorticity equation then hold approximately:

$$h' + h_0 \nabla \cdot \boldsymbol{\xi}' = O(\mu a, a^2), \tag{A 5}$$

$$v'_{,x} - u'_{,y} - f h'/h_0 = O(\mu a, a^2), \tag{A 6}$$

using the fact that $\boldsymbol{\xi}'_{,t} = \boldsymbol{u}' + O(\mu a, a^2)$.

Using (A 5), (3.6), (4.8), and integration by parts, the effective mean depth $\tilde{h}$ defined in (3.9) can be approximated via

$$\tilde{h} = h^\xi J = \overline{h^\xi J} = \overline{h^\xi (1 + \nabla \cdot \boldsymbol{\xi} + \xi_{,x}\eta_{,y} - \xi_{,y}\eta_{,x})} \tag{A 7}$$

$$\Rightarrow \qquad \tilde{h} = \overline{h}^L + \overline{h' \nabla \cdot \boldsymbol{\xi}'} + h_0 \overline{(\xi'_{,x}\eta'_{,y} - \xi'_{,y}\eta'_{,x})} + O(\mu a^2, a^3) \tag{A 8}$$

$$\Rightarrow \qquad \tilde{h} = \overline{h} + \overline{h}^S - \overline{h'^2}/h_0 + h_0 \left( \overline{(\xi' \eta'_{,y})}_{,x} - \overline{(\xi' \eta'_{,x})}_{,y} \right) + O(\mu a^2, a^3) \tag{A 9}$$

$$\Rightarrow \qquad \tilde{h} = \overline{h} + O(\mu a^2, a^3). \tag{A 10}$$

Using (A 5) and (A 6), the $x$-component of the pseudomomentum vector $\boldsymbol{p}$ defined in (3.15) can be approximated via

$$p = -\overline{\xi_{j,x} [\boldsymbol{u}^\ell + \tfrac{1}{2} \boldsymbol{f} \times \boldsymbol{\xi}]_j} = -\overline{\xi'_{,x} u'} - \overline{\eta'_{,x} v'} + \tfrac{1}{2} f (\overline{\xi'_{,x} \eta'} - \overline{\eta'_{,x} \xi'}) + O(\mu a^2, a^3) \tag{A 11}$$

$$\Rightarrow \qquad p = -\overline{\xi'_{,x} u'} + \overline{\eta' v'_{,x}} + f\, \overline{\xi'_{,x} \eta'} + O(\mu a^2, a^3) \tag{A 12}$$

$$\Rightarrow \qquad p = -\overline{\xi'_{,x} u'} - \overline{\eta'_{,y} u'} + f\, \overline{\eta' h'}/h_0 + f\, \overline{\xi'_{,x} \eta'} + O(\mu a^2, a^3) \tag{A 13}$$

$$\Rightarrow \qquad p = \overline{h' u'}/h_0 - f\, \overline{\eta' \eta'_{,y}} + O(\mu a^2, a^3) \tag{A 14}$$

$$\Rightarrow \qquad \boldsymbol{p} = \overline{h' \boldsymbol{u}'}/h_0 + O(\mu a^2, a^3), \tag{A 15}$$

where the final result holds for $\boldsymbol{p}$ because $\boldsymbol{p}$ and $\boldsymbol{u}'$ are vectors.

*Note added in proof*. The exact nonlinear waves mentioned near the foot of p. 249, see Bühler (1993), were described earlier by V. I. Shrira in the Russian literature. A summary of that work and further references can be found in Grimshaw *et al*. 1998.

## REFERENCES


Andrews, D. G. & McIntyre, M. E. 1978*a* An exact theory of nonlinear waves on a Lagrangian-mean flow. *J. Fluid Mech.* **89**, 609–646 (referred to herein as AM78*a*).

Andrews, D. G. & McIntyre, M. E. 1978*b* On wave-action and its relatives. *J. Fluid Mech.* **89**, 647–664 (and corrigendum **95**, 796) (referred to herein as AM78*b*).

Babin, A., Mahalov, A. & Nicolaenko, B. 1997 Global splitting and regularity of rotating shallow-water equations. *Eur. J. Mech.* B/*Fluids* **16**, 725–754.

Bretherton, F. P. 1969 On the mean motion induced by internal gravity waves. *J. Fluid Mech.* **36**, 785–803.

Bühler, O. 1993 A nonlinear wave in rotating shallow water. In *Woods Hole*, *WHOI*-94-12 (ed. R. Salmon), pp. 153–159. Woods Hole, Woods Hole Oceanographic Inst.

Bühler, O. 1996 Waves and balanced mean flows in the atmosphere. PhD thesis, Cambridge University.

Bühler, O. & McIntyre, M. E. 1998 On non-dissipative wave–mean interactions in the atmosphere or oceans. *J. Fluid Mech.* **354**, 301–343 (referred to herein as BM).

Grimshaw, R. H. J., Ostrovsky, L. A., Shrira, V. I. & Stepanyants, Yu. A. 1998 Long nonlinear surface and internal gravity waves in a rotating ocean. *Surveys in Geophysics* **19**, 289–338.

Holton, J. R., Haynes, P. H., McIntyre, M. E., Douglass, A. R., Rood, R. B. & Pfister, L. 1995 Stratosphere–troposphere exchange. *Rev. Geophys.* **33**, 403–439.

Kuo, A. C. & Polvani, L. M. 1999 Wave–vortex interactions in rotating shallow water. Part 1. One space dimension. *J. Fluid Mech.* **394**, 1–27.

Lighthill, M. J. 1978 *Waves in Fluids*. Cambridge University Press.

Majda, A. J. & Embid, P. 1997 Averaging over fast gravity waves for geophysical flows with unbalanced initial data. *Theor. Comput. Fluid Dyn.* **11**, 155–169.

McIntyre, M. E. 1981 On the 'wave-momentum' myth. *J. Fluid Mech.* **106**, 331–347.

McIntyre, M. E. 1988 A note on the divergence effect and the Lagrangian-mean surface elevation in water waves. *J. Fluid Mech.* **189**, 235–242.

McIntyre, M. E. & Norton, W. A. 1990 Dissipative wave-mean interactions and the transport of vorticity or potential vorticity. *J. Fluid Mech.* **212**, 403–435 (and corrigendum **220**, 693) (referred to herein as MN).

Mo, R. 1994 Studies of wave–mean interaction relevant to the middle atmospheric circulation. PhD Thesis, Cambridge University.

Peregrine, D. H. 1998 Surf zone currents. *Theor. Comput. Fluid Dyn.* **10**, 295–310.

Whitham, G. B. 1974 *Linear and Nonlinear Waves*. Wiley-Interscience.