J. Fluid Mech. (2005), *vol.* 534, *pp.* 67–95.
doi:10.1017/S0022112005004374 Printed in the United Kingdom

Wave capture and wave–vortex duality

By OLIVER BÜHLER[1] AND MICHAEL E. McINTYRE[2]

[1]Center for Atmosphere Ocean Science at the Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, USA

[2]Centre for Atmospheric Science at the Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Rd., Cambridge CB3 0WA, UK

(Received 7 July 2004 and in revised form 7 January 2005)

New and unexpected results are presented regarding the nonlinear interactions between a wavepacket and a vortical mean flow, with an eye towards internal wave dynamics in the atmosphere and oceans and the problem of 'missing forces' in atmospheric gravity-wave parametrizations. The present results centre around a pre-wave-breaking scenario termed 'wave capture', which differs significantly from the standard such scenarios associated with critical layers or mean density decay with altitude. We focus on the peculiar wave–mean interactions that accompany wave capture. Examples of these interactions are presented for layerwise-two-dimensional, layerwise-non-divergent flows in a three-dimensional Boussinesq system, in the strong-stratification limit.

The nature of the interactions can be summarized in the phrase 'wave–vortex duality', whose key points are firstly that wavepackets behave in some respects like vortex pairs, as originally shown in the pioneering work of Bretherton (1969), and secondly that a collection of interacting wavepackets and vortices satisfies a conservation theorem for the sum of wave pseudomomentum and vortex impulse, provided that the impulse is defined appropriately. It must be defined as the rotated dipole moment of the Lagrangian-mean potential vorticity (PV). This PV differs crucially from the PV evaluated from the curl of either the Lagrangian-mean or the Eulerian-mean velocity. The results are established here in the strong-stratification limit for rotating (quasi-geostrophic) as well as for non-rotating systems. The concomitant momentum budgets can be expected to be relatively complicated, and to involve far-field recoil effects in the sense discussed in Bühler & McIntyre (2003). The results underline the three-way distinction between impulse, pseudomomentum, and momentum. While momentum involves the total velocity field, impulse and pseudomomentum involve, in different ways, only the vortical part of the velocity field.

1. Introduction

Recently we reported a new insight into wave-induced mean forces (Bühler & McIntyre 2003, hereafter BM03). A standard paradigm widely used in atmospheric modelling says that significant – meaning persistent, cumulative – $O(a^2)$ mean forces induced by waves of amplitude $O(a)$ can arise only when and where the waves are generated or dissipated. We showed by detailed analysis of an idealized problem that the standard paradigm is wrong, except for the two-dimensional, translationally or rotationally invariant, backgrounds assumed by classical non-acceleration theorems. For three-dimensional backgrounds the paradigm omits significant mean forces not associated with wave generation or dissipation.

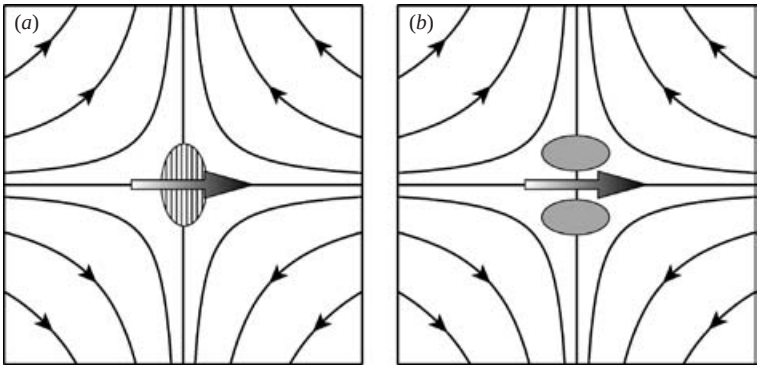

FIGURE 1. (*a*) Wavepacket exposed to pure horizontal strain contracting along the x-axis and extending along the y-axis. The wavecrests align with the extension axis and their spacing is decreased, so that the wavenumber vector $\boldsymbol{k}$ points at right angles to the extension axis and grows in magnitude, as suggested by the large arrow. (*b*) A pair of oppositely signed vortices exposed to same strain. The arrow now indicates the vortex pair's Kelvin impulse.

Such 'missing forces' appear to be generic to a far wider range of problems than the idealized problem studied in BM03, which was confined to weak-refraction asymptotics and shallow-water flow, or equivalent gasdynamic flow. By appealing to results from generalized Lagrangian-mean (GLM) theory and taking note, in particular, of the pioneering work of Bretherton (1969, hereafter B69), we argued in BM03 that similar missing forces will arise in realistic problems of internal inertia–gravity waves in stratified, rotating fluid media. As recalled in BM03, such problems are directly relevant to atmospheric modelling, and possibly to ocean modelling as well. We also noted that essentially the same phenomena may be expected to be characteristic of phonon–vortex interactions in superfluids.†

The forces arise from the persistent advection of vorticity or potential-vorticity anomalies, relative to their surroundings, by $O(a^2)$ wave-induced mean velocities. As shown in BM03, a vortex core can be thus advected – so that the effect of the waves is equivalent to a sideways Magnus force on the core in the absence of waves – even if a refracted wavetrain passes at a distance remote from the core. We therefore described the effective force on the vortex core as a 'remote recoil'. The present paper develops the theory for more realistic three-dimensional stratified problems, and confirms that various kinds of remote recoil are indeed involved. Some of them are even more striking than in the case of BM03.

The simplest and most striking example of all is that of figure 1(*a*), an internal gravity or inertia–gravity wavepacket in a steady background flow consisting of pure horizontal strain. This is an extreme case of *strong* refraction. As pointed out by Jones (1969) and by Badulin & Shrira (1993), hereafter J69 and BS93, such a wavepacket rapidly becomes frozen into the background flow, with the wavecrests behaving as passive tracers. The wavelength and intrinsic group velocity shrink toward zero like $\exp(-|\alpha|t)$, for horizontal strain-rate α, while the wavenumber $\boldsymbol{k}$ goes like $\exp(|\alpha|t)$. The exponential behaviour makes this process a non-trivial variant of classical critical-level absorption. We call it 'wave capture'. At first sight it confronts us with an

† This is a controversial topic. Within a big literature one may cite, for instance, the complementary discussions and bibliographies in Nazarenko, Zabusky & Scheidegger (1995) and Stone (2000).

acute paradox regarding wave-induced mean forces. The so-called momentum of the wavepacket, strictly the packet-integrated pseudomomentum $\boldsymbol{P}$, being equal to $\boldsymbol{k}$ times the packet's conserved wave action, tends toward infinity like $\exp(|\alpha|t)$ until the waves break or otherwise dissipate. The horizontal component $\boldsymbol{P}_H$ likewise tends toward infinity, as hinted by the large arrow in figure 1(*a*). If one were to believe the version of the 'wave momentum myth' sometimes associated with the standard paradigm – saying that the waves, on dissipating, give 'their momentum' to the mean flow – then one would be left with a picture in which a local wave-induced force, exponentially larger than the forces involved in generating the wavepacket, suddenly appears out of nowhere as the waves break.

Here we show in detail how the resolution of the paradox lies precisely in the remoteness of the associated recoil and in the presence of missing forces and momentum fluxes – missing, that is, from the above picture – which operate throughout the evolution of the wavepacket and not just when it is generated and dissipated. These forces and momentum fluxes are anything but local. In fact, the key to understanding the situation is the same as the key to understanding what happens to a two-dimensional vortex pair being pulled apart by pure strain, as shown in figure 1(*b*). Here the wavepacket is replaced by a vortex pair and $\boldsymbol{P}_H$ is replaced by the Kelvin hydrodynamical impulse $\boldsymbol{I}$, namely the first moment of the vorticity distribution rotated through a right angle, i.e. the rotated dipole moment (e.g. p. 529 in Batchelor 1967).

$\boldsymbol{I}$ changes under strain in a manner closely analogous to the way $\boldsymbol{P}_H$ changes for the wavepacket, accompanied by essentially the same remote-recoil effects. In developing a general theoretical framework we shall find it convenient, therefore, to speak of a generalized vortex dynamics involving a *wave–vortex duality*, implying a non-trivial extension of standard vortex dynamics for strongly stratified, layerwise-two-dimensional flow. A central result of this paper is that when vortices and wavepackets are both present they satisfy a conservation theorem for the sum $\boldsymbol{P}_H + \boldsymbol{I}$, (7.6) below, as long as $\boldsymbol{I}$ is defined in a suitable way. Then, for instance, if a wavepacket is being strained by the velocity field of a nearby vortex pair, the resulting changes in its $\boldsymbol{P}_H$ are accompanied by compensating changes in $\boldsymbol{I}$ for the vortex pair. An example of this will be analysed in detail (§ 10 below). One may regard the situation of figure 1(*a*) as a formal limiting case in which the background strain is produced by suitably distributed vortices at infinity. The changing $\boldsymbol{P}_H$ of the wavepacket is accompanied by a remote recoil on the infinitely distant vortices, changing their total $\boldsymbol{I}$ in compensation. A parallel conservation theorem holds, and is similarly useful, in problems of three-dimensional phonon–vortex interactions in superfluids (Bühler & McIntyre, in preparation).†

† The three-dimensional $\boldsymbol{P}_H + \boldsymbol{I}$ conservation theorem greatly simplifies the description and quantification of phonon–vortex interactions, by comparison with momentum-based descriptions such as those cited in the previous footnote. In connection with the quantum-fluids and condensed-matter literature we caution that not only pseudomomentum and momentum, but also momentum and impulse, are sometimes conflated. (One reason may simply be that in some languages, including Russian and German, a single word 'Impuls' tends to be used for all three; an old paper by one of us, *On the wave momentum myth* (McIntyre 1981), has been circulated in Russian translation as *Mif o volnovom impul'sye*. Another reason for the conflation may be a technical error dating back to 1976 and propagated in the well-known book by Donnelly (1991, p. 15), equating rates of change of momentum and impulse through, in effect, mistaking a conditionally convergent integral for an absolutely convergent integral.) The present work is enough to remind us of the need

The wave–vortex duality just indicated stems from the relation between $\boldsymbol{P}$ and the Kelvin circulation for a general material circuit – see (6.3) below – more specifically the relation between $\boldsymbol{P}_H$ and the Kelvin circulation for material circuits lying on isentropes or stratification surfaces. Those relations are most clearly exhibited by the GLM theory, via its exact definitions of $\boldsymbol{P}$ and $\boldsymbol{P}_H$ (Andrews & McIntyre 1978*a*, hereafter AM78*a*; also Gjaja & Holm 1996; Bühler 2000, hereafter B00). The exact definitions account for the contributions to the Kelvin circulation from correlations between wave-induced velocity fluctuations and undulations of the material circuit. In the case of large-scale atmosphere–ocean dynamics the effective forces associated with $\boldsymbol{P}_H$ are therefore related to distributions of Rossby–Ertel potential vorticity (PV), and to the balanced, vortical part of the velocity field derivable from PV inversion (e.g. Hoskins, McIntyre & Robertson 1985, 1987), which to a first approximation is just the layerwise-non-divergent part. It is important to note, therefore, that there is no reason to expect there to be such a thing as a $\boldsymbol{P}_H$-associated force straightforwardly acting on the part of the velocity field not derivable by PV inversion, which to a first approximation is the layerwise-divergent, layerwise-irrotational part.

The latter part is, however, involved in mediating the remote recoil effects. It does so in subtle and complicated ways that depend on the precise parameter regime and on the particular fluid domain geometry and boundaries, or conditions at infinity, as the case may be. It usually involves an unbalanced far field, consisting of fast, large-scale acoustic-gravity-wave motions of amplitude $O(a^2)$, part of a *far-field recoil* of the kind discussed in § 4.4 of BM03. Far-field recoil is the reason why incompressible flow models exhibit conditionally convergent momentum integrals (e.g. Batchelor 1967, p. 529; Bühler & McIntyre, in preparation).

For the strongly stratified, near-incompressible flows of interest we have the aphorism, then, that pseudomomentum is simple and momentum complicated. More precisely, the $\boldsymbol{P}_H + \boldsymbol{I}$ budget of layerwise-two-dimensional vortical flow is simple (though implicitly involving remote recoil, of the kind exemplified by the *vortex-core recoil* already mentioned); and the accompanying momentum budget is complicated (involving further varieties of remote recoil including the far-field recoil). The relative simplicity of the $\boldsymbol{P}_H + \boldsymbol{I}$ budget is important because it is the balanced, vortical flow, and the vortex dynamics, that usually matters most in atmosphere–ocean modelling, for all kinds of reasons from the accuracy of weather forecasting to the correct representation of greenhouse-gas transport. Thus the representation of gravity-wave effects in atmosphere–ocean models, the so-called 'gravity-wave parametrization problem', needs above all to respect the $\boldsymbol{P}_H + \boldsymbol{I}$ budget and vortex-core recoil effects, while leaving the numerical model to handle the momentum budget along with the model-specific, boundary-sensitive far-field recoil effects.

The plan of the paper is as follows. In §§ 2–4 we establish the needed ray-theoretic notation and formalism, showing in § 2 the relevance of passive-tracer behaviour, in § 3 the vulnerability of internal waves to capture, and in § 4 the near-inevitability of wave dissipation after capture. In § 5 we briefly review the seminal work of B69 and show how it leads to the notion of wave–vortex duality, in a simple non-rotating case, and in § 6 we exhibit the underlying exact GLM results, which point toward generalizations including the effects of rotation. In §§ 7–8 we derive the conservation theorem that justifies the notion of a $\boldsymbol{P}_H + \boldsymbol{I}$ budget, and show how it helps make

to keep all three quantities – impulse, momentum, and pseudomomentum – sharply distinguished from one another.

sense of various cases including the peculiar wave–mean interactions and $\boldsymbol{P}_H$ evolution during wave capture, taking account of the evolution of pseudomomentum density and its curl according to ray theory. The type of vortex-core recoil analysed in BM03, involving an effective Magnus force though not wave capture, fits effortlessly into the same framework. Indeed, the $\boldsymbol{P}_H + \boldsymbol{I}$ conservation theorem shows that essentially the same vortex-core recoil phenomena will arise in any problem in which waves refract past a vortex, whether or not weak-refraction asymptotics apply as in BM03. Dissipative effects are discussed in § 9, where we show that in a certain sense wave capture and wave dissipation are equivalent, for wave–mean interaction purposes. In § 9.2 we point out one implication of that equivalence, in the form of a serious caveat regarding recent attempts to estimate the secondary gravity-wave generation caused by dissipating wavepackets. Such estimates will be much too high in some cases, essentially through the neglect of missing forces prior to the onset of wave dissipation. In § 10 an example involving a single vortex pair and a single wavepacket is discussed in detail, to further illustrate the points made in the paper. Concluding remarks are offered in § 11.

2. Analogy between wavecrests and passive tracers

As a prelude to analysing wave capture for internal waves, and for any other wave type, we recall the well-known partial analogy between the ray-tracing equations and the equations governing the evolution of a passive tracer. This emerges most simply in a generic three-dimensional setting.

The ray-tracing equations govern the phase and amplitude evolution of an $O(a)$ linear wavepacket or wavetrain with slowly varying amplitude $a(\boldsymbol{x}, t) \ll 1$ and rapidly varying phase function $\Theta(\boldsymbol{x}, t)$ such that a typical wave field is given by $a(\boldsymbol{x}, t)\exp(\mathrm{i}\Theta)$, with real parts understood. Defining the local wavenumber vector $\boldsymbol{k}$ and frequency ω via

$$\boldsymbol{k} \equiv +\nabla\Theta, \quad \omega \equiv -\Theta_t \tag{2.1}$$

and introducing the absolute frequency function

$$\Omega(\boldsymbol{k}, \boldsymbol{x}, t) = \hat{\omega} + \boldsymbol{U}\cdot\boldsymbol{k}, \tag{2.2}$$

we have Hamilton's equations for the evolution of $\boldsymbol{x} = (x, y, z)$ and $\boldsymbol{k} = (k, l, m)$ as functions of time along group-velocity rays, i.e. wavepacket trajectories,

$$\frac{\mathrm{d}_g\boldsymbol{x}}{\mathrm{d}t} = +\frac{\partial\Omega}{\partial\boldsymbol{k}} \quad \text{and} \quad \frac{\mathrm{d}_g\boldsymbol{k}}{\mathrm{d}t} = -\frac{\partial\Omega}{\partial\boldsymbol{x}}. \tag{2.3}$$

Here $\boldsymbol{U} = (U, V, W)$ is the $O(1)$ background velocity field on which the $O(a)$ waves propagate and $\hat{\omega}$ is the intrinsic frequency, i.e. the frequency intrinsic to the wave dynamics, the Doppler-shifted frequency seen from a local frame moving with velocity $\boldsymbol{U}$. As already indicated we shall be especially interested in strongly stratified, layerwise-two-dimensional, layerwise-non-divergent flows satisfying

$$\boldsymbol{U} = (U, V, 0) \quad \text{and} \quad \frac{\partial U}{\partial x} + \frac{\partial V}{\partial y} = 0, \tag{2.4}$$

where z is vertical; but these restrictions will not be needed until the next section.

The intrinsic frequency $\hat{\omega}$ always depends on $\boldsymbol{k}$, and both $\hat{\omega}$ and $\boldsymbol{U}$ may in addition depend slowly on $\boldsymbol{x}$ and t. The absolute group velocity $\boldsymbol{c}_g = (u_g, v_g, w_g)$ is

$$\boldsymbol{c}_g = \frac{\mathrm{d}_g\boldsymbol{x}}{\mathrm{d}t} = \hat{\boldsymbol{c}}_g + \boldsymbol{U}, \tag{2.5}$$

where the intrinsic group velocity $\hat{\boldsymbol{c}}_g \equiv \partial\hat{\omega}/\partial\boldsymbol{k}$ measures wavepacket propagation relative to the moving fluid background. The time derivative along a ray is equivalent to the operator

$$\frac{\mathrm{d_g}}{\mathrm{d}t} \equiv \frac{\partial}{\partial t} + \boldsymbol{c}_g\cdot\nabla = \frac{\partial}{\partial t} + (\hat{\boldsymbol{c}}_g + \boldsymbol{U})\cdot\nabla \tag{2.6}$$

acting on slowly varying functions of $(\boldsymbol{x}, t)$.

For simplicity, though this is not essential, we will assume that $\hat{\omega}$ depends on $\boldsymbol{k}$ only, so that the explicit dependence of Ω on $\boldsymbol{x}$ and t is contained entirely in the background flow $\boldsymbol{U}(\boldsymbol{x}, t)$. Equation (2.3*b*) then becomes

$$\frac{\mathrm{d_g}\boldsymbol{k}}{\mathrm{d}t} = -(\nabla\boldsymbol{U})\cdot\boldsymbol{k} \quad\Leftrightarrow\quad \frac{\mathrm{d_g}k_i}{\mathrm{d}t} = -\frac{\partial U_j}{\partial x_i}k_j \tag{2.7}$$

where summation is understood; note that $\boldsymbol{k}$ contracts with $\boldsymbol{U}$ and not with ∇. This equation describes the rate of change of $\boldsymbol{k} = \nabla\Theta$ due to mean-flow refraction, i.e. due to the differential advection of the phase lines $\Theta = \text{const.}$ by the background flow $\boldsymbol{U}$.† Let us compare (2.7) to the evolution equation for the gradient $\nabla\phi$ of a passive tracer $\phi(\boldsymbol{x}, t)$ advected by $\boldsymbol{U}$,

$$\mathrm{D}_t\phi \equiv \left(\frac{\partial}{\partial t} + \boldsymbol{U}\cdot\nabla\right)\phi = 0 \quad\Rightarrow\quad \mathrm{D}_t(\nabla\phi) = -(\nabla\boldsymbol{U})\cdot\nabla\phi, \tag{2.8}$$

where again $\nabla\phi$ contracts with $\boldsymbol{U}$ and not with ∇. Comparing (2.7*a*) with (2.8*b*), we see that the only difference is that between $\mathrm{d_g}/\mathrm{d}t$ and D_t, the time derivatives following wavepacket and material particle respectively:

$$\frac{\mathrm{d_g}}{\mathrm{d}t} - \mathrm{D}_t = (\hat{\boldsymbol{c}}_g\cdot\nabla). \tag{2.9}$$

Hence $\boldsymbol{k}$ behaves like the gradient of a passive tracer to the extent that the difference between these two time derivatives does not matter, and to the extent that $-(\nabla\boldsymbol{U})\cdot\boldsymbol{k}$ dominates any other terms that might arise in (2.7) from reintroducing dependence of $\hat{\omega}$ upon $\boldsymbol{x}$. Under suitable circumstances, therefore, the well-studied properties of passive advection can be relevant to dispersive wave dynamics. For instance, in the passive advection case it is well known that $|\nabla\phi|$ grows exponentially in time, on average, in the Batchelor random-straining regime, in which isotropic statistics are assumed along with a scale separation between the length scales of $\boldsymbol{U}$ and ϕ. Haynes & Anglade (1997) find similar behaviour in a scale-separated random-straining model for velocity fields of the anisotropic form (2.4). Could the same exponential growth be characteristic of the evolution of $|\boldsymbol{k}|$, given that the relevant scale separation is already implied in the ray-theoretic or slow-variation assumption?

As pointed out by BS93, the answer is yes provided that $|\hat{\boldsymbol{c}}|$ decreases as $|\boldsymbol{k}|$ increases, as is true, in particular, of all internal gravity, inertia and Rossby waves and their hybrids. Then the analogy between ray-tracing and passive advection will tend to reinforce itself, becoming increasingly accurate as $|\hat{\boldsymbol{c}}|$ and $|\mathrm{d_g}/\mathrm{d}t - \mathrm{D}_t|$ decrease. In such cases, therefore, there is scope for wave capture to occur, with its signature of exponentially large $|\boldsymbol{k}|$. From here on we denote $|\boldsymbol{k}|$ by κ.

† The wave phase Θ moves with the phase velocity $\boldsymbol{k}\omega/|\boldsymbol{k}|^2$. However, for the evolution of $\boldsymbol{k} = \nabla\Theta$ only the gradient of $\boldsymbol{U}$ matters.

3. Wave capture for internal waves

Internal inertia–gravity waves are susceptible to wave capture (J69, BS93) because of the form of their intrinsic dispersion relation,

$$\hat{\omega}^2 = f^2 + (N^2 - f^2)\,\frac{k_H^2}{\kappa^2}, \quad \kappa^2 = k_H^2 + m^2, \quad k_H^2 = k^2 + l^2, \tag{3.1}$$

alternatively

$$\frac{k_H^2}{m^2} = \frac{\hat{\omega}^2 - f^2}{N^2 - \hat{\omega}^2}, \tag{3.2}$$

where $f > 0$ is the Coriolis parameter and $N > 0$ the buoyancy frequency, both assumed constant. Prandtl's ratio f/N is typically small in most of the atmosphere, roughly 10^{-2}, and more variable but still usually well below unity in the oceans. The Boussinesq approximation, well justified for large m^2, has been made. We see that $\hat{\omega}$ depends only on the ratio k_H/m of horizontal to vertical wavenumber, in turn making it clear on dimensional grounds that the intrinsic group velocity $\hat{\boldsymbol{c}}_g = \partial\hat{\omega}/\partial\boldsymbol{k}$ must be inversely proportional to κ for fixed $\hat{\omega}$. Explicitly, the components of $\hat{\boldsymbol{c}}_g$ are

$$(\hat{u}_g, \hat{v}_g) = \frac{(k,l)}{k_H}\,\frac{N^2 - \hat{\omega}^2}{\hat{\omega}\kappa}\sqrt{\frac{\hat{\omega}^2 - f^2}{N^2 - f^2}} \quad \text{and} \quad \hat{w}_g = -\mathrm{sgn}(m)\,\frac{\hat{\omega}^2 - f^2}{\hat{\omega}\kappa}\sqrt{\frac{N^2 - \hat{\omega}^2}{N^2 - f^2}}, \tag{3.3}$$

so that

$$|\hat{\boldsymbol{c}}_g|^2 = \frac{(N^2 - \hat{\omega}^2)(\hat{\omega}^2 - f^2)}{\hat{\omega}^2\kappa^2}, \tag{3.4}$$

confirming the inverse proportionality of $\hat{\boldsymbol{c}}_g$ to κ and the possibility of wave capture.

With $\boldsymbol{U}$ of the form (2.4), the background velocity gradient tensor is

$$\nabla\boldsymbol{U} \equiv \begin{pmatrix} U_x & V_x & W_x \\ U_y & V_y & W_y \\ U_z & V_z & W_z \end{pmatrix} = \begin{pmatrix} U_x & V_x & 0 \\ U_y & -U_x & 0 \\ U_z & V_z & 0 \end{pmatrix} \tag{3.5}$$

say, where suffixes denote partial derivatives and where, for ray theory to be consistently applicable, we require large Richardson number,

$$Ri \;\equiv\; N^2/\big(U_z^2 + V_z^2\big) \gg 1. \tag{3.6}$$

Notice that the evolution of the horizontal wavenumber vector $\boldsymbol{k}_H = (k, l)$ decouples from that of the vertical wavenumber m. Thus (2.7) splits into the two subsystems

$$\frac{\mathrm{d_g}}{\mathrm{d}t}\begin{pmatrix} k \\ l \end{pmatrix} = -\begin{pmatrix} U_x & V_x \\ U_y & -U_x \end{pmatrix}\begin{pmatrix} k \\ l \end{pmatrix} \quad \text{and} \quad \frac{\mathrm{d_g}}{\mathrm{d}t}\,m = -U_z k - V_z l, \tag{3.7}$$

which can be studied in sequence.

Generically, U_x etc. are functions of time along the ray. For the sake of simplicity, we now neglect this time-dependence. This is a severe simplification and we hope to report later on work that goes beyond it, following the Haynes & Anglade work already cited. Here we restrict ourselves to a locally steady linear flow $U = U_0 + U_x x + U_y y + U_z z$ and $V = V_0 + V_x x - U_x y + V_z z$, where (x, y, z) are measured from the starting location of the ray at $t = 0$, and U_x etc. are all taken to be constant.

The evolution of $\boldsymbol{k}_H$ may be contrasted with the evolution of a material line segment initially coincident with $\boldsymbol{k}_H$, so that one can imagine the tip of the line segment moving with the local velocity field $(U - U_0, V - V_0)$, relative to the starting location $\boldsymbol{x} = (0, 0, 0)$. The curl $V_x - U_y$ acts equally on both objects: it rotates their

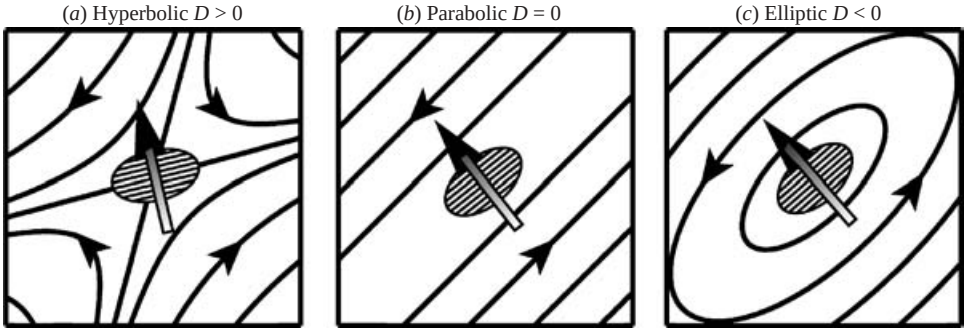

FIGURE 2. Generic contours of horizontal stream function $\psi = -\alpha xy + 0.5\gamma(x^2 + y^2)$ with positive α and γ in principal strain coordinates such that $V_x + U_y = 0$, $D = \alpha^2 - \gamma^2$. (For non-zero $V_x + U_y$ and U_x such coordinates are obtained by a rotation through an angle $0.5 \arctan((V_x + U_y)/(2U_x))$.) Also indicated are captured wavepackets at the orientation of the growing eigenmode in the first two cases and in that of maximal transient amplification in the third, elliptic case. The large arrows indicate $\boldsymbol{k}_H$. (*a*) A hyperbolic case ($\gamma = 0.5\alpha$) similar to that in figure 1 (in which $\gamma = 0$). If $\gamma > 0$ then the axis of extension is turned counterclockwise by $0.5 \arcsin(\gamma/\alpha)$ whilst the axis of contraction is turned *clockwise* by the same angle. With increasing γ the strain axes close like a pair of scissors. The advected wavecrests align with the extension axis; and the growing horizontal wavenumber vector $\boldsymbol{k}_H$, which is always perpendicular to the crests, becomes perpendicular to the extension axis. (*b*) The scissors shut in the parabolic or simple shear case $\gamma = \alpha$. (*c*) An elliptic case ($\gamma = 1.5\alpha$). The ellipses have aspect ratio $\sqrt{(\gamma + \alpha)/(\gamma - \alpha)}$, which equals the maximum transient amplification factor for $|\boldsymbol{k}_H(t)|$.

components by $(V_x - U_y)/2$ radians per unit time around the z-axis. However, the layerwise-irrotational strain flow, which is determined by $U_x = -V_y$ and $U_y + V_x$, acts oppositely on $\boldsymbol{k}_H$ and the line segment: the tip of the $\boldsymbol{k}_H$ vector behaves like the tip of the line segment but with the sign of the irrotational strain reversed. This is a consequence of $\boldsymbol{k}$ being the gradient of a scalar, Θ; in modern geometric language, $\boldsymbol{k}$ is a 1-form rather than a 1-vector.

Most important for our purposes is the long-time behaviour of $\boldsymbol{k}_H(t)$, which is governed by the exponential stretching rates given by the matrix eigenvalues $\pm\sqrt{D}$, where D is the determinant $U_x^2 + V_x U_y$, i.e.

$$D = U_x^2 + \left(\frac{V_x + U_y}{2}\right)^2 - \left(\frac{V_x - U_y}{2}\right)^2. \tag{3.8}$$

Note that a non-zero curl always diminishes the stretching rate. As illustrated in figure 2(*a*), if $D > 0$ then the streamlines are hyperbolic and there is a wavenumber eigenmode that grows exponentially in time (J69) with stretching rate $\sqrt{D}$. The growing eigenmode is excited by almost all initial conditions; asymptotically $\boldsymbol{k}_H(t) \propto (-V_x,\ U_x + \sqrt{D}) \exp(\sqrt{D}\, t)$. This is the case of wave capture.

If $D = 0$ (figure 2*b*) then the flow is a parallel shear flow with a linearly growing wavenumber, such that asymptotically $\boldsymbol{k}_H(t) \propto (-V_x,\ U_x)\, |V_x - U_y|\, t$, i.e. classical critical-layer behaviour. If $D < 0$ (figure 2*c*) then the streamlines are closed ellipses, implying that the wavenumber evolution is bounded, though temporary amplification can still occur, up to a factor equal to the aspect ratio of the ellipse. For $D \geqslant 0$ and for almost all initial $\boldsymbol{k}_H(0)$, the asymptotic orientation and growth rate of $\boldsymbol{k}_H$ depend solely on $\nabla \boldsymbol{U}$. In other words, we have a robust behaviour in which the wavenumber vector of a captured wavepacket forgets about the initial conditions at large time, and asymptotically points in a direction determined by the local velocity gradient alone.

We now consider the evolution of the vertical wavenumber m in the case $D > 0$, which features exponential growth of $\boldsymbol{k}_H$ and therefore κ, with stretching rate $\sqrt{D}$. Turning to (3.7*b*) we see that m will also exhibit exponential growth at large time, unless it so happens that the right-hand side of (3.7*b*) is zero for the growing eigenmode in horizontal wavenumber, implying zero vertical shear in that direction. Specifically, at large time

$$m(t) = -\frac{U_z k(t) + V_z l(t)}{\sqrt{D}} + O(1), \tag{3.9}$$

where $k(t)$ and $l(t)$ correspond to the growing eigenmode, $\boldsymbol{k}_H \equiv (k, l) \propto (-V_x, U_x + \sqrt{D})\exp(\sqrt{D}\,t)$. For a captured wavepacket we therefore have the scaling relation

$$|\boldsymbol{k}_H/m| \sim H/L \tag{3.10}$$

in order of magnitude, if H, L are vertical and horizontal length scales characteristic of the $\boldsymbol{U}$ field. In the real atmosphere we typically have $H/L \sim f/N$. If for the sake of argument we take $|\boldsymbol{k}_H/m| = H/L = f/N$ precisely, in the wave-capture limit, then from (3.2)

$$\hat{\omega}(t)^{-2} \to \tfrac{1}{2}(f^{-2} + N^{-2}); \qquad \text{hence} \quad \hat{\omega}(t)^2 \to 2f^2 \tag{3.11}$$

if we further approximate using $N^2 \gg f^2$. It is noteworthy that values of $\hat{\omega}$ close to f appear to be commonplace in the oceans and the atmosphere (e.g. Garrett & Munk 1975; Polzin 2004; Fritts & Alexander 2003 and references therein), where for long data records the observed wave energy spectrum typically peaks at or near f, so ubiquitous wave capture would not, of itself, greatly distort the frequency spectrum.

It may be anticipated that the robustness of the wave-capture scenario will persist when the neglected time-dependence of the mean-flow gradient $\nabla\boldsymbol{U}$ along rays is restored and taken into account. By virtue of the passive-tracer analogy and the results of Haynes & Anglade (1997), we may reasonably hypothesize that the exponential straining of the wavenumber vector components will typically be slowed down but not eradicated – that is, capture will typically be delayed but not prevented – and that the scaling relation (3.10) will tend to persist.

4. Wave amplitude evolution

Before turning to the nonlinear mean-flow response and recoil effects we need to consider the amplitude evolution during wave capture. We show that the exponentially increasing amplitudes rapidly overtake any reasonable wave-breaking thresholds.

The wave amplitude a along non-intersecting rays is governed by the standard ray-theoretic equation for wave action A per unit mass at $O(a^2)$ (e.g. Bretherton & Garrett 1968; Whitham 1974), which for the Boussinesq system is

$$\frac{\partial A}{\partial t} + \nabla\cdot(A\boldsymbol{c}_g) = \frac{\overline{\boldsymbol{u}'\cdot\boldsymbol{F}'}}{\hat{\omega}} \quad \Leftrightarrow \quad \frac{\mathrm{d}_g A}{\mathrm{d}t} + A\nabla\cdot\boldsymbol{c}_g = \frac{\overline{\boldsymbol{u}'\cdot\boldsymbol{F}'}}{\hat{\omega}}. \tag{4.1}$$

Here $A = E/\hat{\omega}$, and

$$E = \tfrac{1}{2}(\overline{|\boldsymbol{u}'|^2} + \overline{b'^2}/N^2) = \overline{|\boldsymbol{u}'_{||}|^2}, \tag{4.2}$$

the Eulerian-mean wave energy per unit mass, the overbars denoting averages over a wave period at constant $\boldsymbol{x}$, and $\boldsymbol{u}'_{||}$ the projection of $\boldsymbol{u}'$ on to the vertical plane of $\boldsymbol{k}$. The fluctuating body-force field $\boldsymbol{F}'$ will be taken to be zero, making A a conserved density, except in cases of slow wave generation or dissipation (§ 9), where

$\boldsymbol{F}'$ has magnitude $O(a\mu)$ with $\mu \ll 1$ characterizing slow variation. The other symbols are $\boldsymbol{u}' = (u', v', w')$, the $O(a)$ wave velocity, and b', the $O(a)$ disturbance buoyancy acceleration.

Now the non-dimensional amplitude a of a locally plane wave can conveniently be defined as the maximum vertical shear of the horizontal wave velocity $\boldsymbol{u}'$ parallel to $\boldsymbol{k}_H$ divided by the buoyancy frequency N. For example, in the case $\boldsymbol{k} = (k, 0, m)$ the amplitude would be

$$a = \frac{|u'_z|_{\max}}{N}. \tag{4.3}$$

This definition of a has the advantage of flagging the two most relevant wave-breaking thresholds with the same order-of-magnitude criterion $a \sim 1$, to within a factor 2, across the whole frequency range $f \leqslant \hat{\omega} \leqslant N$ implied by (3.2) (e.g. McIntyre 2003, § 11). These are the thresholds for gravitational convective instability and for Kelvin–Helmholtz vertical shear instability. Weaker, slower instabilities with $a \ll 1$ are also possible (e.g. Hasselmann 1967; McEwan 1971; Sonmor & Klaassen 1997). For the geophysically interesting case $N^2 \gg f^2$, the amplitude a is related to A by

$$a^2 = A\kappa^2 \frac{2\hat{\omega}(N^2 - \hat{\omega}^2)^2}{N^6}; \tag{4.4}$$

so knowledge of A and $\boldsymbol{k}$ implies knowledge of a.

During wave capture, the density A becomes approximately constant along a ray if dissipation is negligible, because of the tendency to passive-tracer behaviour. For the ray divergence $\nabla \cdot \boldsymbol{c}_g = \nabla \cdot \hat{\boldsymbol{c}}_g$ by (2.4); also, by assumption, the intrinsic group velocity $\hat{\boldsymbol{c}}_g$ varies by only a negligible amount across the wavepacket (as also assumed in § 3.1 of B69). Therefore $|\nabla \cdot \hat{\boldsymbol{c}}_g|$ is a small quantity to begin with, even before capture. During capture $|\hat{\boldsymbol{c}}_g|$ decays exponentially, whilst spatial derivatives across the footprint of the strained wavepacket grow at most at the same exponential rate. This means that the typical magnitude of $|\nabla \cdot \hat{\boldsymbol{c}}_g|$ is at most constant and therefore remains small by assumption. With no dissipation, (4.1*b*) has zero right-hand side and we are left with $\mathrm{d}_g A/\mathrm{d}t \approx 0$, as anticipated.

With constant A but exponentially growing κ we see from (4.4), with $f \leqslant \hat{\omega} < N$, that the wave amplitude a must also grow exponentially and must overtake all wave-breaking thresholds unless other dissipation mechanisms intervene. Indeed, with the scaling (3.11) we obtain

$$a^2(t) \to A\kappa(t)^2 \frac{2\sqrt{2}f}{N^2} \to \frac{A\sqrt{2}f}{|\hat{\boldsymbol{c}}_g(t)|^2} \tag{4.5}$$

for the asymptotic behaviour of a, confirming that the exponential decay of $|\hat{\boldsymbol{c}}_g|$ goes hand in hand with the exponential growth of a at the same rate.

In summary, wave capture must lead to wave dissipation in one way or another, with exponential rapidity. As hinted in the Introduction, however, we need to consider what happens before and during wave capture if we are to understand the recoil effects.

5. Nonlinear mean-flow response and wave–vortex duality

Linear $O(a)$ waves produce a nonlinear mean-flow response at $O(a^2)$, the topic usually dealt with by small-amplitude wave–mean interaction theory. Using that theory, the joint evolution of waves and mean flow can be studied in great detail. Laboratory experiments show that this can be a good qualitative guide even for large wave amplitude, which usually defies detailed theoretical analysis. The extension to

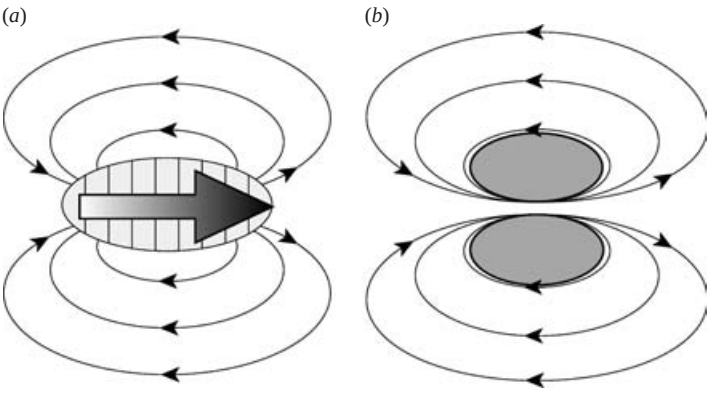

FIGURE 3. Illustration of wave–vortex duality (schematic only, plan view). (*a*) A wavepacket on an isentrope together with streamlines indicating the return branches of the (horizontal) Bretherton flow. As in figure 1, the arrow indicates the direction and magnitude of $\boldsymbol{p}_H$. (*b*) A vortex pair on the same isentrope with qualitatively the same return flow. The shaded areas indicate non-zero PV values with opposite signs.

large amplitude is all the more plausible if we use the GLM theory, in which certain key results are formally valid at arbitrary wave amplitude a, as noted in the next section.

Another important key is the seminal work of B69, in which the mean-flow response to a propagating wavepacket of internal waves is computed in the ray-tracing regime, for $f=0$. The result is sketched in figure 3(*a*). The wavepacket is surrounded by a horizontal, layerwise-two-dimensional, layerwise-non-divergent dipolar return flow, like that of the ordinary two-dimensional vortex pair sketched in figure 3(*b*), with velocities falling off as the inverse square of distance. For convenience we call the layerwise-two-dimensional $O(a^2)$ dipolar flow indicated in figure 3(*a*) the *Bretherton flow* of the wavepacket; the part of the flow outside the wavepacket will also be called the *wavepacket return flow*. In B69 the $O(a^2)$ velocity field was computed in terms of the usual Eulerian-mean velocity $\overline{\boldsymbol{u}}$, defined by averaging over a wave period at fixed $\boldsymbol{x}$. It was pointed out in B69 that the Kelvin impulse of the Eulerian-mean flow $\overline{\boldsymbol{u}}$ in figure 3(*a*) is formally well defined, by an absolutely convergent volume integral – unlike the momentum – and that in the ray-theoretic approximation the impulse of $\overline{\boldsymbol{u}}$ is just equal to the packet-integrated horizontal pseudomomentum. That is, B69 showed that

$$\iiint (y, -x, 0)\left(\frac{\partial \overline{v}}{\partial x} - \frac{\partial \overline{u}}{\partial y}\right) \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = \boldsymbol{P}_H \equiv \iiint \boldsymbol{p}_H\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \tag{5.1}$$

where $\boldsymbol{p}_H$ is the horizontal projection of the pseudomomentum density $\boldsymbol{p}$, with $\boldsymbol{p}$ itself given in the ray-theoretic approximation by the standard formula

$$\boldsymbol{p} = \frac{E}{\hat{\omega}}\,\boldsymbol{k} = A\boldsymbol{k}. \tag{5.2}$$

On the left of (5.1) the impulse integral is absolutely convergent, essentially because the integrand has compact support. To sufficient approximation, the integrand is zero at altitudes z above and below the wavepacket. At the intervening z values, the return flow is layerwise-irrotational outside the wavepacket implying $(\partial\overline{v}/\partial x - \partial\overline{u}/\partial y) \equiv \widehat{\boldsymbol{z}}\cdot\nabla\times\overline{\boldsymbol{u}} = 0$ there. The entire Bretherton flow, inside and outside the wavepacket, is

layerwise-non-divergent,

$$\frac{\partial \overline{u}}{\partial x} + \frac{\partial \overline{v}}{\partial y} = 0, \qquad \text{with} \qquad \overline{w} = 0. \tag{5.3}$$

Bretherton's result (5.1) follows at once from compactness of the wavepacket, which allows us to write

$$\iint \boldsymbol{p}_H \, \mathrm{d}x \, \mathrm{d}y = \iint (y, -x, 0) \, (\widehat{\boldsymbol{z}} \cdot \nabla \times \boldsymbol{p}_H) \, \mathrm{d}x \, \mathrm{d}y \tag{5.4}$$

using integration by parts, together with the relation

$$\widehat{\boldsymbol{z}} \cdot \nabla \times \overline{\boldsymbol{u}} = \widehat{\boldsymbol{z}} \cdot \nabla \times \boldsymbol{p}_H \tag{5.5}$$

which again holds to leading order within the ray-theoretic approximations. We omit Bretherton's derivation of this last result (5.5) since it has an exact counterpart in the GLM theory, (6.6) below, whose simple derivation is recalled at the end of § 6.

The relation (5.5) and its exact GLM counterpart (6.6) form the basis for the idea of wave–vortex duality already mentioned. Each of the two coherent structures in figure 3 can be understood in terms of vorticity inversion, the sole difference being that the effective vorticity field to be inverted is curl $\boldsymbol{p}$, more precisely $\widehat{\boldsymbol{z}} \cdot \nabla \times \boldsymbol{p}$, $= \widehat{\boldsymbol{z}} \cdot \nabla \times \boldsymbol{p}_H$, on the left of the figure, and the ordinary vorticity $\widehat{\boldsymbol{z}} \cdot \nabla \times \boldsymbol{u}$ on the right of the figure.

Of course a wavepacket is not the same thing as a vortex pair. The two structures have different propagation velocities. The significance of wave capture, not considered in B69, is that it makes the difference vanish for all practical purposes. This is further discussed in §§ 7.5 and 9.1. So the idea of wave–vortex duality involves not only the similarity between the two structures in figure 3, but also the fact that wave capture can, in effect, turn a wavepacket into a vortex pair *whether or not wave breaking or other dissipative processes have intervened.*

Before proceeding, we remark that the Bretherton flow supplies us with one of the missing pieces of the remote-recoil jigsaw. If figure 3(*a*) is superposed on figure 1(*a*), we see that there is an $O(a^2)$ advective momentum flux systematically importing positive x-momentum into the square region shown, through correlations between the two superposed velocity fields, namely the $O(a^2)$ Bretherton flow and the $O(1)$ straining flow.† The local, differential form of the $\boldsymbol{P}_H + \boldsymbol{I}$ conservation theorem, (7.9) below, will show that a similar though not identical statement holds for the flux of Kelvin impulse.

6. Circulation, PV, and the exact form of (5.5)

We now recall some theory that is fundamental to the wave–vortex duality concept. It is also needed to motivate the appropriate definition of Kelvin impulse.

† The jigsaw has further pieces, of course, including those known from classical impulse theory (e.g. Batchelor 1967, pp. 519, 529): in the case of figure 1(*a*) the rate of change of momentum within the square region shown is only half the advective rate of import, the difference being accounted for by the rate of export of x-momentum via an $O(a^2)$ horizontal dipolar pressure field (the missing ingredient in the result previously quoted from Donnelly 1991). A similar dipolar pressure field is involved in the inner part of the far-field recoil discussed in § 4.4 of BM03. For reasons of symmetry, the same partitioning between momentum tendencies and advective and pressure-mediated momentum-flux contributions is found for a circular region, though in general the partitioning depends on the shape of the region considered, as is typical of problems involving conditionally convergent integrals. Further discussion is given in our paper in preparation.

As in other problems of wave–mean interaction theory, the key is to recognize the importance of the Lagrangian conservation laws for ideal-fluid flow, whose averaged counterparts appear in the GLM theory. Formally, these averaged conservation laws are valid at finite wave amplitude a but in practice one usually reverts to $O(a^2)$ accuracy, as will for the most part be done in the next section. For simplicity we continue to neglect the background rotation f, as in B69, returning to $f \neq 0$ in § 8.

The Lagrangian conservation laws in question are Kelvin's circulation theorem (whose importance for wave–mean problems was recognized in Rayleigh's *Theory of Sound*, 1896) and the concomitant potential-temperature and potential-vorticity theorems. For the ideal non-rotating, stratified flows under consideration, Kelvin's circulation theorem states that

$$\oint_{\mathscr{C}} \boldsymbol{u} \cdot \mathrm{d}\boldsymbol{x} = \text{constant}, \tag{6.1}$$

where $\mathscr{C}$ is any closed material loop lying on an isentrope or stratification surface, that is, a surface of constant specific entropy or potential temperature θ, with $\mathrm{D}\theta/\mathrm{D}t = 0$. The isentropes are material surfaces moving with velocity $\boldsymbol{u}$.

In the GLM theory (AM78*a*, *b*; for further developments see also Gjaja & Holm 1996; Bühler & McIntyre 1998, hereafter BM98; and B00), the material invariance of θ is inherited by the Lagrangian-mean specific entropy $\overline{\theta}^L$. Thus we have in buoyancy-acceleration units

$$\theta = b(\boldsymbol{x}, t) + N^2 z, \quad \frac{\mathrm{D}\theta}{\mathrm{D}t} = 0; \quad \Rightarrow \quad \overline{\theta}^L = \overline{b}^L + N^2 z, \quad \overline{\mathrm{D}}^L \overline{\theta}^L = 0, \tag{6.2}$$

with no restriction to small b, so that variable background stratification is now included. Here $\overline{\mathrm{D}}^L = \partial/\partial t + \overline{\boldsymbol{u}}^L \cdot \nabla$, the advective derivative based on the Lagrangian-mean (i.e. particle-following mean) velocity $\overline{\boldsymbol{u}}^L$.† This implies that the mean isentropes or constant-$\overline{\theta}^L$ surfaces move with velocity exactly equal to $\overline{\boldsymbol{u}}^L$. There is an inherited mean-circulation theorem

$$\oint_{\mathscr{C}^L} (\overline{\boldsymbol{u}}^L - \boldsymbol{p}) \cdot \mathrm{d}\boldsymbol{x} = \text{constant}, \tag{6.3}$$

which is exact provided that the exact GLM definition of $\boldsymbol{p}$ is used (AM78*a*, eq. (3.1)), with $\mathscr{C}^L$ a closed loop moving with velocity $\overline{\boldsymbol{u}}^L$ and lying on a mean isentrope. Importantly, the loop is advected with velocity $\overline{\boldsymbol{u}}^L$ but the mean circulation is formed from velocity $\overline{\boldsymbol{u}}^L - \boldsymbol{p}$.‡ Similarly, material invariance of the PV, meaning the exact Rossby–Ertel potential vorticity, Q say, is inherited by its Lagrangian mean $\overline{Q}^L$ (AM78*a*, eq. (2.22)); and again we have advection by $\overline{\boldsymbol{u}}^L$ but a vorticity factor formed

† For small-amplitude propagating internal gravity waves $\overline{\boldsymbol{u}}^L \approx \overline{\boldsymbol{u}}$ to sufficient accuracy, where $\overline{\boldsymbol{u}}$ is the usual Eulerian mean flow defined by averaging over a wave period at constant $\boldsymbol{x}$. This is because there is no Stokes drift at leading order for these waves, even for a finite-amplitude plane wave, which in the Boussinesq system is an exact solution. Indeed B69 derived an approximate version of (6.3) based on (5.2) and on $\overline{\boldsymbol{u}}$. However, we will continue to use $\overline{\boldsymbol{u}}^L$ in order to stress the more general validity of (6.3) and (6.4) for other problems, and to prepare for the generalized definition of Kelvin impulse in § 7.

‡ As is now well understood, the appearance of twin velocity fields or 'velocity splitting' is an inescapable consequence of a flow averaging scheme, or any other procedure such as symplectic balancing or symplectic numerics, that preserves material invariance while suppressing some aspect of the structure, such as, in the present case, small-scale wave structure (e.g. Gjaja & Holm 1996; McIntyre & Roulstone 1996).

from $\overline{\boldsymbol{u}}^L - \boldsymbol{p}$. In fact, consistently with (6.3), we have

$$\overline{Q}^L = \tilde{\rho}^{-1} \nabla \overline{\theta}^L \cdot \nabla \times (\overline{\boldsymbol{u}}^L - \boldsymbol{p}) \quad \text{and} \quad \overline{\mathrm{D}}^L \overline{Q}^L = 0. \tag{6.4}$$

Here $\tilde{\rho}$ is the density satisfying the exact GLM mass-conservation relation

$$\partial \tilde{\rho} / \partial t + \nabla \cdot (\tilde{\rho}\, \overline{\boldsymbol{u}}^L) = 0. \tag{6.5}$$

As discussed in § 4 of AM78*a*, $\tilde{\rho}$ is a mean (non-fluctuating) quantity, though not equal to $\overline{\rho}^L$ in general. (Remarkably, not only (6.5) but also the first of (6.4) holds, with suitable initialization, even for flows subject to arbitrary body forces $\boldsymbol{F}$, provided that $\mathrm{D}\theta/\mathrm{D}t = 0$ for all t (BM98).) Rotating counterparts of the above are noted at the end of § 8.

Now in Bretherton's problem the initial state is undisturbed and the PV is exactly zero everywhere. Therefore (6.4) implies that

$$\widehat{\boldsymbol{n}}^L \cdot \nabla \times \overline{\boldsymbol{u}}^L = \widehat{\boldsymbol{n}}^L \cdot \nabla \times \boldsymbol{p}, \tag{6.6}$$

where $\widehat{\boldsymbol{n}}^L = \nabla \overline{\theta}^L / |\nabla \overline{\theta}^L|$, the unit normal to a mean isentrope. This is the exact GLM counterpart of (5.5). The unit vector $\widehat{\boldsymbol{n}}^L$ becomes vertical in the limit of infinitely strong stratification, recovering (5.5).

7. $\boldsymbol{P}_H + \boldsymbol{I}$ conservation with and without mean-flow refraction

7.1. *Preliminaries*

Our aim in this section is to extend Bretherton's problem (B69) to cases in which the wavepacket is refracted and possibly captured by the mean flow. In particular, we seek a generalization of the relation between horizontal pseudomomentum and Kelvin impulse indicated by Bretherton's result (5.1). The key is to replace the left-hand side of (5.1) by the GLM version suggested by the simplicity of (6.4) – that is, the vorticity used in the definition of the impulse is $\nabla \times (\overline{\boldsymbol{u}}^L - \boldsymbol{p})$ rather than $\nabla \times \overline{\boldsymbol{u}}^L$ or $\nabla \times \overline{\boldsymbol{u}}$.

Before proceeding we should sound one note of caution. Unlike the GLM relations displayed in § 6, the relation between momentum and Kelvin impulse holds only in the strong-stratification limit underpinning (5.3) and (5.5). In fact generalization of the Kelvin impulse concept to less restrictive parameter regimes is not straightforward, for subtle reasons related, in part, to the limitations inherent in the notions of balanced flow and PV inversion. More specifically, there is an expectation that balanced flows will not themselves precisely conserve even momentum, let alone impulse (McIntyre & Norton 2000), because of the ubiquity of the weak Lighthill radiation, or spontaneous-adjustment emission, of inertia–gravity waves by unsteady vortical motion – yet another aspect of wave–mean and wave–vortex interaction problems and, in the present context, another aspect of the far-field recoil effects. In the less restrictive parameter regimes there are, therefore, added layers of complication within the aphorism that "momentum is complicated".

So it is at present a matter of speculation how far the impulse concept can usefully be generalized. We content ourselves here with returning for the most part to the strong-stratification limit and the ray-tracing approximations, with $f = 0$ in this section and $f \neq 0$ in § 8. We can then use the concept of impulse in the original Kelvin form, apart from using $\nabla \times (\overline{\boldsymbol{u}}^L - \boldsymbol{p})$ instead of $\nabla \times \overline{\boldsymbol{u}}$, allowing everything to be made analytically clear in a simple way.

7.2. *The $\boldsymbol{P}_H + \boldsymbol{I}$ conservation theorem*

Assuming ideal-fluid flow in the strong-stratification limit, and remembering the absence of Stokes drifts in plane internal waves, we can replace (5.3) by their GLM counterparts

$$\nabla_H \cdot \overline{\boldsymbol{u}}^L = 0 \qquad \text{with} \qquad \overline{w}^L = 0 \tag{7.1}$$

and take the mean isentropes $\overline{\theta}^L = \text{const.}$ to be horizontal planes. Here ∇_H is the horizontal projection of ∇. Given the $\boldsymbol{p}_H$ field, the mean flow $\overline{\boldsymbol{u}}^L$ is then fully determined by the circulation integrals (6.3) on horizontal z-surfaces, or equivalently by the horizontal distributions of the material invariant $\overline{q}^L \propto \overline{Q}^L$:

$$\overline{q}^L = \widehat{\boldsymbol{z}} \cdot \nabla \times (\overline{\boldsymbol{u}}^L - \boldsymbol{p}_H); \qquad \overline{\mathrm{D}}^L \overline{q}^L = 0. \tag{7.2}$$

Far-field recoil effects have been pushed out of sight, to infinity. In components, we have

$$\frac{\partial \overline{u}^L}{\partial x} + \frac{\partial \overline{v}^L}{\partial y} = 0 \quad \text{and} \quad \frac{\partial \overline{v}^L}{\partial x} - \frac{\partial \overline{u}^L}{\partial y} = \overline{q}^L + \widehat{\boldsymbol{z}} \cdot \nabla \times \boldsymbol{p}_H, \tag{7.3}$$

generalizing (5.5), hence applying both to the Bretherton problem $\overline{q}^L \equiv 0$ and to its extensions to arbitrary wavefields and vortices.

If we now define the impulse $\boldsymbol{I}$ and its density $\boldsymbol{i}$ by

$$\boldsymbol{I}(t) = \iiint \boldsymbol{i}(\boldsymbol{x}, t)\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z \qquad \text{where} \qquad \boldsymbol{i} = (y, -x, 0)\, \overline{q}^L, \tag{7.4}$$

then because $\overline{\mathrm{D}}^L \overline{q}^L = 0$ we have $\overline{\mathrm{D}}^L \boldsymbol{i} = (\overline{v}^L, -\overline{u}^L, 0)\, \overline{q}^L$. A little manipulation, using (7.1) and the definition of the horizontal Kronecker delta Δ_{ij} with $\Delta_{i3} = \Delta_{3j} = 0$, puts this equation for $\overline{\mathrm{D}}^L \boldsymbol{i}$ into conservation form apart from a term reminiscent of the right-hand side of (2.7):

$$\frac{\partial i_i}{\partial t} + \frac{\partial}{\partial x_j} \left\{ i_i \overline{u}_j^L + \overline{u}_i^L \overline{u}_j^L - \left(\tfrac{1}{2} |\overline{\boldsymbol{u}}^L|^2 - \overline{\boldsymbol{u}}^L \cdot \boldsymbol{p}_H \right) \Delta_{ij} - p_{Hi} \overline{u}_j^L \right\} = \frac{\partial \overline{u}_j^L}{\partial x_i} p_{Hj} \, (i = 1, 2). \tag{7.5a}$$

Indeed, (2.7) with (4.1) and (5.2) implies

$$\frac{\partial p_{Hi}}{\partial t} + \frac{\partial}{\partial x_j} \left\{ p_{Hi} \hat{c}_{gj} + p_{Hi} \overline{u}_j^L \right\} = -\frac{\partial \overline{u}_j^L}{\partial x_i} p_{Hj} \qquad (i = 1, 2), \tag{7.5b}$$

for non-dissipative wave motion, correct to $O(a^2)$ since $\boldsymbol{U} = \overline{\boldsymbol{u}}^L + O(a^2)$. By comparing the right-hand sides we see, therefore, not only that $\boldsymbol{p}_H + \boldsymbol{i}$ is a conserved density for non-dissipative motion, but also how the individual rates of change of $\boldsymbol{p}_H$ and $\boldsymbol{i}$ are related to wave propagation and horizontal refraction. For instance, it is the source–sink term $-(\partial \overline{u}_j^L / \partial x_i) p_{Hj}$ that accounts for the exponential rate of increase of the packet-integrated pseudomomentum in the situation of figure 1(*a*).

For an arbitrary collection of wavepackets and vortices within a closed system whose boundaries recede to infinity, all the flux terms vanish fast enough to make no contribution at infinity. Remarkably, this remains true even for $\overline{q}^L$ distributions with non-vanishing monopole moment, since with $\overline{q}^L$, $\boldsymbol{i}$ and $\boldsymbol{p}_H$ all compact or sufficiently evanescent the only flux terms not similarly compact or evanescent are those quadratic in $\overline{\boldsymbol{u}}^L$ which, even in monopolar cases, are $O(r^{-2})$ where $r^2 = x^2 + y^2$. So by adding (7.5*a*) to (7.5*b*) and integrating, we have simply

$$\boldsymbol{P}_H + \boldsymbol{I} = \text{constant}, \tag{7.6}$$

while the compensating individual rates of change are non-zero only when there is horizontal refraction somewhere, hence creation or destruction of horizontal pseudomomentum:

$$\frac{\mathrm{d}\boldsymbol{P}_H}{\mathrm{d}t} = -\iiint (\nabla_H \overline{\boldsymbol{u}}^L)\cdot \boldsymbol{p}_H \,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \tag{7.7}$$

and

$$\frac{\mathrm{d}\boldsymbol{I}}{\mathrm{d}t} = \iiint (\nabla_H \overline{\boldsymbol{u}}^L)\cdot \boldsymbol{p}_H \,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z. \tag{7.8}$$

It should be remembered that $\boldsymbol{I}$, though not its rate of change, depends on the choice of coordinate origin when $\overline{q}^L$ has non-vanishing monopole moment.

We note in passing that (7.5*a*) and (7.8) are valid at arbitrary finite amplitude a, if the exact GLM definition of $\boldsymbol{p}_H$ is used, though dependent on the strong-stratification conditions (7.1)–(7.2). By contrast (7.5*b*), (7.6) and (7.7) have been derived correct to $O(a^2)$ only, and depend on the ray-theoretic approximations.† We note also that the conservation theorem does not require the constant background stratification assumed in (2.7). It is valid for z-dependent slowly varying stratification $N^2(z)$, since background z-dependence affects only the z or $i=3$ component of the right-hand side of (2.7).

We may summarize everything so far by displaying the full differential form of the conservation theorem implied by (7.5*a*) and (7.5*b*), remembering always that the strong-stratification conditions (7.1)–(7.2) are needed (banishing far-field recoil to infinity) and noting the interesting fact that the pseudomomentum advection terms cancel:

$$\frac{\partial}{\partial t}(p_{Hi} + I_i) + \frac{\partial}{\partial x_j}\left\{p_{Hi}\hat{c}_{gj} + I_i\overline{u}_j^L + \overline{u}_i^L\overline{u}_j^L - \tfrac{1}{2}|\overline{\boldsymbol{u}}^L - \boldsymbol{p}_H|^2\,\Delta_{ij}\right\} = 0 \quad (i=1,2). \tag{7.9}$$

The Bernoulli terms have been rewritten in an alternative form, correct to $O(a^2)$. This is sometimes more convenient, since outside a compact $\overline{q}^L$ distribution the field $\overline{\boldsymbol{u}}^L - \boldsymbol{p}_H$ is layerwise irrotational, $\overline{q}^L = 0$, with layerwise divergence $-\nabla_H\cdot\boldsymbol{p}_H$.

Together with the absence of pressure terms, dictated by their absence from the vorticity equation $\overline{\mathrm{D}}^L\overline{q}^L = 0$, these features help to simplify the kind of argument used in § 5.2 of BM03 – some of whose complexity stems from using momentum rather than impulse. Specifically, if the system is open in the sense that waves can come in from infinity, as in BM03's problem of waves scattered by a single vortex, then the pseudomomentum flux $p_{Hi}\hat{c}_{gj}$ due to wave propagation becomes significant at infinity. Indeed (7.9) shows at once that vortex-core recoil effects of just the kind illustrated in BM03 will be found in most problems involving waves refracting past a single vortex. Care is still needed in estimating the other flux terms at large r, though this task is simplified, in comparison with the momentum-based analysis of BM03 § 5.2, by the absence of pressure terms and by the layerwise irrotationality of $\overline{\boldsymbol{u}}^L - \boldsymbol{p}_H$. If the wavefield is steady then in most such problems the only rate-of-change term available to balance the net pseudomomentum flux from infinity is $\partial\boldsymbol{I}/\partial t$, integrating to $\mathrm{d}\boldsymbol{I}/\mathrm{d}t$ and representing the vortex-core recoil – most simply manifested as sideways vortex-core translation as in BM03's problem.

† There is an exact GLM counterpart of (7.5*b*) – essentially the difference between (3.8) and (8.7a) of AM78 – but its application is far less simple and will be left aside here. The main complication is that background inhomogeneities, including $\nabla\tilde{\rho}$, can no longer be treated as independent of wavefields. Finite-amplitude wave propagation involves self-refraction.

7.3. *Layerwise pseudomomentum–impulse budget*

In the strong-stratification limit it is natural to consider layerwise $\boldsymbol{P}_H + \boldsymbol{I}$ budgets. Continuing to assume compact or sufficiently evanescent $\overline{q}^L$ and $\boldsymbol{i}$ distributions on each isentrope, we define the layerwise Kelvin impulse, i.e. the impulse per unit altitude z, as

$$\boldsymbol{J}(z,t) \equiv \iint \boldsymbol{i}\,\mathrm{d}x\,\mathrm{d}y, \tag{7.10}$$

where the integral is taken over all (x, y) at constant (z, t). Then $\boldsymbol{I} = \int \boldsymbol{J}\,\mathrm{d}z$. We see from (7.5) that $\boldsymbol{J}$ evolves according to

$$\frac{\partial \boldsymbol{J}}{\partial t} = -\left\{\frac{\partial}{\partial t}\iint \boldsymbol{p}_H\,\mathrm{d}x\,\mathrm{d}y + \frac{\partial}{\partial z}\iint (\boldsymbol{p}_H \hat{w}_g)\,\mathrm{d}x\,\mathrm{d}y\right\} = \iint (\nabla_H \overline{\boldsymbol{u}}^L)\cdot\boldsymbol{p}_H\,\mathrm{d}x\,\mathrm{d}y \tag{7.11}$$

where it has been assumed that $\boldsymbol{p}_H$ is also compact or sufficiently evanescent. This is the layerwise counterpart of (7.6)–(7.8). It is useful, for instance, when considering the effect at given altitude z of a steady wavetrain propagating from below, as with waves generated by flow over bottom topography. With a steady wavetrain, (7.11) shows at once that refraction by *horizontal* gradients of $\overline{\boldsymbol{u}}^L$ must lead to a non-constant *vertical* flux of horizontal pseudomomentum, whose convergence at the given altitude is equal to minus the rate of change of $\boldsymbol{J}$ at that altitude.

7.4. *Extensions of the Bretherton problem*

The $O(a^2)$ mean-flow response to the arrival of a wavepacket on a given mean isentrope $z = \text{const.}$ can be computed from (7.3). That response is initially given by the Bretherton flow, defined as the layerwise-two-dimensional, layerwise-non-divergent flow whose vertical curl is equal to the vertical curl of $\boldsymbol{p}_H$. In general this is not, of course, the entire $O(a^2)$ mean-flow response after the arrival of the wavepacket. (In this respect the situation is like that studied in BM98.) A non-uniform $\overline{q}^L$ with $O(1)$ gradients will itself be changed at $O(a^2)$ by advection with the Bretherton flow, in turn triggering ongoing changes in the self-induced advection of $\overline{q}^L$ (the changes, in some cases, then amplifying if the flow has positive Lyapunov exponents). But we may still say that an $O(a^2)$ dipolar flow qualitatively like the Bretherton flow always accompanies wavepackets.

Figure 4 shows a simple example. A wavepacket, whose Bretherton flow is shown by the dashed streamlines, interacts with a vortex pair brought to rest in a uniform flow, shown by solid streamlines. Other graphical conventions are as in figure 3. The full analytical details are given in §10 below. This is a wave-capture scenario, in which the growth in $|\boldsymbol{P}_H|$ and $|\boldsymbol{p}_H|$ is compensated by the decrease in $|\boldsymbol{I}|$ and $|\boldsymbol{J}|$ due to compression of the vortex pair by the Bretherton flow, as demanded by (7.6) and (7.11). The vortex-core recoil – the effective force $\frac{1}{2}\mathrm{d}\boldsymbol{I}/\mathrm{d}t$ on each vortex core in the absence of waves (BM03) – is directed rightwards; notice that $\boldsymbol{P}_H$ and $\boldsymbol{I}$ are both negative or leftwards. The generic features of wave capture are further discussed in the next subsection.

To prepare for that discussion we make one more remark about the circulation theorem (6.3). If we revert to the original assumption $N^2 = \text{constant}$ then (7.5*b*) holds for $i = 3$ as well as for $i = 1, 2$. Then we have simply

$$\frac{\mathrm{d}}{\mathrm{d}t}\oint_{\mathscr{C}^L} \boldsymbol{p}\cdot\mathrm{d}\boldsymbol{x} \;=\; -\oint_{\mathscr{C}^L} \nabla\cdot(\boldsymbol{p}\hat{\boldsymbol{c}}_g)\cdot\mathrm{d}\boldsymbol{x}, \tag{7.12}$$

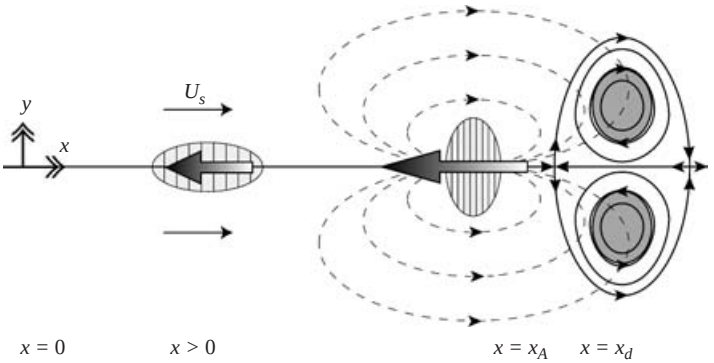

FIGURE 4. View from above of a topographic wavepacket being captured by a blocking dipole or vortex pair; full details are in § 10. Initially the wavepacket (not shown) is generated at $x = y = 0$ and at a z value below the level depicted, with the absolute group velocity vertically upward toward the viewer. Compression in x and stretching in y by the horizontal strain slows the intrinsic propagation, so that asymptotically the wavepacket approaches the level depicted while being swept toward the upwind stagnation point $x = x_A$. The packet-integrated pseudomomentum $\boldsymbol{P}_H$ continually increases in magnitude, as suggested by the large arrows, and does so at an exponential rate as long as the size of the wavepacket remains small relative to background scales. The wavepacket's Bretherton flow, likewise growing exponentially, is indicated schematically by the dashed streamlines. That flow compresses the vortex pair on the right, reducing the magnitude of the vortex pair's impulse at just the rate needed to compensate the exponential growth of $|\boldsymbol{P}_H|$ as demanded by the $\boldsymbol{P}_H + \boldsymbol{I}$ conservation theorem.

for *arbitrary* mean material loops $\mathscr{C}^L$ in three dimensions. As before, the divergence operator contracts with $\hat{\boldsymbol{c}}_g$. Because $\mathscr{C}^L$ moves with velocity $\overline{\boldsymbol{u}}^L$, the source–sink term from the right of (7.5*b*) has disappeared. When $N^2 =$ constant, only intrinsic wave propagation – i.e. only wave propagation relative to the fluid – can change the pseudomomentum circulation around mean material loops. In the absence of such propagation the three-dimensional vector $\nabla \times \boldsymbol{p}$ is frozen into the flow $\overline{\boldsymbol{u}}^L$ in the same way as vorticity is frozen into un-averaged incompressible flow $\boldsymbol{u}$. This has relevance to pseudomomentum evolution in some other classes of wave problems, including the phonon–vortex interaction problem and its extension to roton–vortex interactions, rotons at zero group velocity being somewhat like captured phonons.†

7.5. *Generic behaviour under wave capture*

We use the equations for the vertical curls of $\overline{\boldsymbol{u}}^L$ and $\boldsymbol{p}_H$ implied by (7.2) and by (7.12) with isentropic $\mathscr{C}^L$, valid even if $N^2 = N^2(z)$. The equations are

$$\overline{\mathrm{D}}^L(\widehat{\boldsymbol{z}} \cdot \nabla \times \overline{\boldsymbol{u}}^L) \;=\; \overline{\mathrm{D}}^L(\widehat{\boldsymbol{z}} \cdot \nabla \times \boldsymbol{p}_H) \;=\; -\widehat{\boldsymbol{z}} \cdot (\nabla \times (\nabla \cdot (\boldsymbol{p}_H \hat{\boldsymbol{c}}_g))), \tag{7.13}$$

where again the divergence operator contracts with $\hat{\boldsymbol{c}}_g$. The product $\boldsymbol{p}_H \hat{\boldsymbol{c}}_g$ on the right-hand side has factors with compensating growth and decay during wave capture. So estimating the magnitude of this product requires care in general. But under the previously discussed assumption, below (4.4), that $\hat{\boldsymbol{c}}_g$ is approximately uniform across

† *Note added in proof*: We should acknowledge a careless slip in BM03, where on p. 214 we inadvertently suggested that a classical acoustic wavepacket would have a return flow. Though correct for rotons and for slow (dispersive) phonons, it is incorrect for the classical (non-dispersive) wavepacket.

a wavepacket, the second of (7.13) reduces to

$$(\overline{\mathrm{D}}^L + \hat{\boldsymbol{c}}_g \cdot \nabla)(\widehat{\boldsymbol{z}} \cdot \nabla \times \boldsymbol{p}_H) = 0, \tag{7.14}$$

i.e. this scalar is simply advected by the absolute group velocity. During wave capture the total displacement of the wavepacket due to the exponentially decaying $\hat{\boldsymbol{c}}_g$ is finite and small. This allows us to argue that $\boldsymbol{p}_H \hat{\boldsymbol{c}}_g$ in (7.13) makes only a small contribution during wave capture. Therefore, at the late stages of wave capture (7.13)–(7.14) simplify to a statement that the vorticity and pseudomomentum-curl are frozen-in separately:

$$\overline{\mathrm{D}}^L(\widehat{\boldsymbol{z}} \cdot \nabla \times \overline{\boldsymbol{u}}^L) \;=\; 0 \;=\; \overline{\mathrm{D}}^L(\widehat{\boldsymbol{z}} \cdot \nabla \times \boldsymbol{p}_H). \tag{7.15}$$

A curious implication of (7.15) is that after wave capture the evolution of the mean flow $\overline{\boldsymbol{u}}^L$ appears to be, in a certain sense, no longer directly affected by $\boldsymbol{p}_H$ – even though the packet-integrated pseudomomentum $\boldsymbol{P}_H$ is still increasing exponentially, as in figure 1(*a*), with $\boldsymbol{I}$ and $\boldsymbol{J}$ decreasing as demanded by (7.6) and (7.11). After wave capture the layerwise-non-divergent horizontal mean flow $\overline{\boldsymbol{u}}^L$ is governed by the constancy of the mean-flow circulation, the circulation of $\overline{\boldsymbol{u}}^L$, around mean material loops $\mathscr{C}^L$, just as it would be in the complete absence of the waves! The circulation of $\boldsymbol{p}_H$ around a loop $\mathscr{C}^L$ that intersects the wavepacket is similarly constant in time, allowing (6.3) to hold without the $\boldsymbol{p}_H$ term. Of course this does not mean that wave capture has gone unnoticed by the mean flow, a fact already clear from figure 4. The circulation of $\boldsymbol{p}_H$ is indeed constant after wave capture, but the value of the constant has been changed, by the wave capture process, from the value it would otherwise have had.

Let us look more closely at the early sequence of events. The mean flow has initially responded to the arrival of the wavepacket by manifesting the packet's Bretherton flow, which flow would have disappeared again had the wavepacket immediately left the mean isentrope under consideration. It is in this initial stage that the vertical curl of $\boldsymbol{p}_H$ gets added to the vertical curl of $\overline{\boldsymbol{u}}^L$, as described by (7.3). But now that wave capture has intervened, the wavepacket never leaves the isentrope. Hence the Bretherton flow remains in place, persistently advecting the $\overline{q}^L$ distribution on that isentrope. Indeed since $\boldsymbol{P}_H$ and $\boldsymbol{p}_H$, and therefore $\boldsymbol{I}$ and $\boldsymbol{J}$, all have exponentially growing rates of change, we see that there must be an exponentially growing, albeit $O(a^2)$, mean-flow reponse arising from rearrangement of the isentropic distribution of $\overline{q}^L$. In the example of figure 4 the two vortex cores are advected toward each other at an exponentially increasing rate, accounting for the exponentially increasing rate of reduction of $|\boldsymbol{J}|$. This re-arrangement of $\overline{q}^L$ will in turn change the mean flow and its self-advection. In the case of figure 4 the vortex cores will, perhaps counterintuitively, start moving leftward (but recall (7.4)). Such changes will then affect the refraction of the wavepacket. Indeed, the conclusion from (7.11) is that, ultimately, whatever else happens, the exponential surge in $\boldsymbol{p}_H$ must cease because it cannot consume more than the total impulse $\boldsymbol{J}$ available on the isentrope. Notice that the wavepacket can interact and exchange impulse with the vortices in the fashion familiar from elementary vortex dynamics. Indeed, the observed interaction, in which the wave dipole is stretched in the y-direction, resembles the early stages of the familiar 'leap-frogging' of vortex pairs, or vortex rings in three dimensions (e.g. § 10.4 in Saffman 1993).

So, to summarize, the exponential behaviour characteristic of wave capture in no way contradicts the constancy of circulation remarked on below (7.15). This was clear all along from the wave–vortex duality perspective, since (7.15) and (7.1)–(7.3)

tell us that after wave capture we have nothing more than two-dimensional vortex dynamics with the two kinds of 'vorticity', $\overline{q}^L$ and $\widehat{\boldsymbol{z}}\cdot\boldsymbol{\nabla}\times\boldsymbol{p}_H$, separately frozen into the mean flow and interacting with each other just like the ordinary vorticity in ordinary two-dimensional vortex dynamics. The $O(a^2)$ Bretherton flow and its advective power grows in just the same way as the flow around an ordinary vortex pair grows when the pair is pulled apart. After wave capture the Bretherton flow has irreversibly become part of the mean flow, for all practical purposes. We may highlight this point by imagining ourselves somehow unable to observe the waves directly. We imagine that only observations of the mean flow $\overline{\boldsymbol{u}}^L$ are available. The signature of a captured wavepacket would then be indistinguishable from that of a vortex pair, even before dissipative processes have intervened. This is one of the more striking aspects of wave–vortex duality. A non-captured wavepacket, by contrast, would be detectable by its intrinsic group velocity, which moves the location of the packet's $\widehat{\boldsymbol{z}}\cdot\boldsymbol{\nabla}\times\boldsymbol{p}_H$ through the fluid, i.e. at velocities different from $\overline{\boldsymbol{u}}^L$, as described by (7.14). Generically, then, one has the two kinds of vorticity $\overline{q}^L$ and $\widehat{\boldsymbol{z}}\cdot\boldsymbol{\nabla}\times\boldsymbol{p}_H$ moving at velocities $\overline{\boldsymbol{u}}^L$ and $\overline{\boldsymbol{u}}^L+\hat{\boldsymbol{c}}_g$, respectively, but adding together to determine $\overline{\boldsymbol{u}}^L$ via the inversion problem (7.3).

A technical caveat should be noted at this point. The wave–mean interaction theory used above is based on a regular perturbation analysis valid for times $t=O(1)$; so strictly speaking the mean-flow changes it can describe are bounded by $O(a^2)$. To fully justify the foregoing picture one must suppose that the analysis can be extended to mean-flow changes that grow to be $O(1)$. This supposition is reasonable provided that the wave evolution is continually adjusted as $\overline{\boldsymbol{u}}^L$ evolves. Such an assumption finds some additional support in the fact that the most crucial terms in the evolution equations, especially the refraction term $(\boldsymbol{\nabla}_H\overline{\boldsymbol{u}}^L)\cdot\boldsymbol{p}_H$, have nonlinear, finite-amplitude counterparts in the GLM theory, as can be verified from the references cited; see footnote to § 7.2.

In § 9 we analyse post-capture dissipation. The analysis will support what is implicit in all the foregoing, namely that, in a real sense – even though contrary to the standard paradigm – it is the onset of wave capture rather than of wave dissipation that is most significant in these problems.

8. Pseudomomentum–impulse budget with rotation, $f\neq 0$

The generalization of (6.1) and (6.3) to a frame rotating with constant angular velocity $\boldsymbol{f}/2$ around the $\widehat{\boldsymbol{z}}$-axis consists of using the absolute velocity in the circulation integral, (8.6) below, and in the GLM definition of $\boldsymbol{p}$ (AM78*a*). A nonlinear GLM treatment of the Boussinesq equations that includes a nonlinear definition of mean potential vorticity can be found in BM98, along with the extension to the compressible case. Here we will instead use the standard quasi-geostrophic theory, giving a leading-order-approximate potential vorticity for small-Rossby-number rotating flows. The small-Rossby-number scaling includes the wavepacket envelope scales. Then the Kelvin impulse concept can be carried over to rotating problems with minor changes only, essentially because we keep the strong-stratification conditions $\boldsymbol{\nabla}_H\cdot\overline{\boldsymbol{u}}^L=0$ and $\overline{w}^L=0$ to leading order. The quasi-geostrophic potential vorticity is $f+\overline{q}^L$ where $\overline{q}^L$ is now redefined by

$$\overline{q}^L\equiv\widehat{\boldsymbol{z}}\cdot\boldsymbol{\nabla}\times(\overline{\boldsymbol{u}}^L-\boldsymbol{p}_H)+\frac{\partial}{\partial z}\left(\frac{f\overline{b}^L}{N^2}\right),\quad\text{so that}\quad\overline{\mathrm{D}}^L\overline{q}^L=0 \tag{8.1}$$

under geostrophic advection. Here we allow N^2 to be a function of z as is standard in quasi-geostrophic theory. The vertical derivative will give rise to a new vertical flux

divergence in the equation for the impulse density. The theory assumes that the mean flow is in geostrophic balance (contrast the exceptions in BM98), which implies that, to leading order in Rossby number,

$$\nabla_H \cdot \overline{\boldsymbol{u}}^L = 0, \quad \overline{w}^L = 0, \quad \text{and} \quad f\left(\frac{\partial \overline{u}^L}{\partial z}, \frac{\partial \overline{v}^L}{\partial z}\right) = \left(-\frac{\partial \overline{b}^L}{\partial y}, +\frac{\partial \overline{b}^L}{\partial x}\right), \tag{8.2}$$

this last representing the usual thermal wind-shear relations. The generic small-amplitude definition of $\boldsymbol{p}_H$ in (5.2) is unchanged, and so are the subsequent evolution equations including (7.5*b*). The definition of the mean-flow impulse density $\boldsymbol{i}$ is also unchanged but its flux now contains two additional terms, those displayed on the second line of the following version of the $\boldsymbol{p}_H + \boldsymbol{i}$ conservation theorem, which is the rotating counterpart of (7.9):

$$\frac{\partial}{\partial t}(p_{Hi} + i_i) + \frac{\partial}{\partial x_j}\left\{ p_{Hi}\hat{c}_{gj} + i_i \overline{u}_j^L + \overline{u}_i^L \overline{u}_j^L - \tfrac{1}{2}|\overline{\boldsymbol{u}}^L - \boldsymbol{p}_H|^2 \,\Delta_{ij} \right.$$
$$\left. + \frac{f}{N^2}\,\overline{u}_k^L \overline{b}^L \varepsilon_{3ki}\,\delta_{3j} + \frac{\overline{b}^{L\,2}}{2N^2}\,\Delta_{ij} \right\} = 0 \quad (i = 1, 2;\ j = 1, 2, 3). \tag{8.3}$$

The vertical flux of impulse represented by the first term on the second line is analogous to the familiar vertical Eliassen–Palm flux or so-called isentropic or isopycnic 'form drag' of quasi-geostrophic theory, which also quantifies the mean momentum transport in the vertical. The horizontal flux of impulse represented by the second term on the second line integrates out when we form the layerwise budget, with $\boldsymbol{J} \equiv \iint \boldsymbol{i}\,\mathrm{d}x\,\mathrm{d}y$ as before:

$$\frac{\partial \boldsymbol{J}}{\partial t} + \frac{\partial}{\partial z}\left(\frac{f}{N^2}\iint (-\overline{v}^L, +\overline{u}^L, 0)\overline{b}^L\,\mathrm{d}x\,\mathrm{d}y\right)$$
$$= -\left\{\frac{\partial}{\partial t}\iint \boldsymbol{p}_H\,\mathrm{d}x\,\mathrm{d}y + \frac{\partial}{\partial z}\iint (\boldsymbol{p}_H \hat{w}_g)\,\mathrm{d}x\,\mathrm{d}y\right\} = \iint (\nabla_H \overline{\boldsymbol{u}}^L)\cdot \boldsymbol{p}_H\,\mathrm{d}x\,\mathrm{d}y, \tag{8.4}$$

when $\overline{q}^L$, $\boldsymbol{i}$ and $\boldsymbol{p}_H$ are compact or sufficiently evanescent. Since $\overline{\boldsymbol{u}}^L$ decays by one further power of distance, now three-dimensional distance $(x^2 + y^2 + z^2)^{1/2}$, as does $\overline{b}^L$, consistently with (8.2), the flux terms again integrate to zero over the horizontally infinite domain – even in cases of non-vanishing monopole moment of the quasi-geostrophic $\overline{q}^L$.

Thus, with $\overline{q}^L$ and $\boldsymbol{i}$ redefined as above, the global budgets (7.6)–(7.8) hold on just the same basis as before. Wave capture and horizontal pseudomomentum surges are again compensated by mean-flow impulse changes, just as in the non-rotating case $f = 0$. The main modification is that mean-flow changes are now governed by (8.1)–(8.2), involving three-dimensional quasi-geostrophic PV inversion. In particular, the Bretherton flow is now the three-dimensional geostrophic flow whose quasi-geostrophic PV based on $\overline{\boldsymbol{u}}^L$ alone is equal to the vertical curl of $\boldsymbol{p}_H$. The Bretherton flow therefore extends vertically as well as horizontally around the wavepacket, a stack of horizontal vortex-pair flows with vertical-to-horizontal aspect ratio $\sim f/N$, in accordance with the standard quasi-geostrophic scaling.

We note finally that the exact ideal-fluid results of (6.2)–(6.4) have rotational counterparts (AM78*a*; BM98)

$$\overline{\mathrm{D}}^L \overline{\theta}^L = 0 \tag{8.5}$$

with circulation theorem

$$\oint_{\mathscr{C}^L} (\overline{\boldsymbol{u}}^L - \boldsymbol{p} + \tfrac{1}{2}\boldsymbol{f}\times\boldsymbol{x})\cdot \mathrm{d}\boldsymbol{x} = \text{constant}, \tag{8.6}$$

and the corresponding definition and conservation theorem for PV,

$$\overline{Q}^L \;=\; \tilde{\rho}^{-1}\nabla\overline{\theta}^L\cdot\{\boldsymbol{f} + \nabla\times(\overline{\boldsymbol{u}}^L - \boldsymbol{p})\} \qquad \text{and} \qquad \overline{\mathrm{D}}^L\overline{Q}^L \;=\; 0. \tag{8.7}$$

9. Dissipative effects

Wave dissipation can take many forms, including that of strongly nonlinear wave breaking due to the irreversible overturning of isentropes, but clearcut theoretical results are usually restricted to the case of weak dissipation and laminar flows. Here we will give an account of the interplay between wave capture, dissipation, and mean-flow changes based on weak laminar dissipation. However, to the extent that many of our arguments are based on the exact results of § 6, and their rotational counterparts (8.5)–(8.7), we may hypothesize that our results give a good qualitative guide even for the case of strongly nonlinear wave breaking. The discussion applies word-for-word whether $f = 0$ or $f \neq 0$.

9.1. *Pseudomomentum dissipation and mean-flow response*

Dissipative effects can enter either through a fluctuating body force $\boldsymbol{F}$ per unit mass in the momentum equation or through diabatic heating terms in the buoyancy equation. We will not consider heating terms here, but see BM98 for a discussion. In the ray-tracing regime we see from (4.1) and (5.2) that dissipation adds a decay term $\hat{\omega}^{-1}\boldsymbol{k}\,\overline{(\boldsymbol{u}'\cdot\boldsymbol{F}')} \equiv \mathscr{F}$, say, to the right of the pseudomomentum evolution equation, where $\mathscr{F}$ is on average anti-parallel to $\boldsymbol{p}$ if the force is dissipative, i.e. if $\overline{(\boldsymbol{u}'\cdot\boldsymbol{F}')} < 0$. So in this case $\mathscr{F}$ can be written as $-\boldsymbol{p}/\tau$, where $\tau = -E/\overline{(\boldsymbol{u}'\cdot\boldsymbol{F}')} > 0$, a $\boldsymbol{k}$-dependent dissipation time scale that measures the attenuation of $\boldsymbol{p}$ along the ray. An exact GLM definition of $\mathscr{F}$ can be found in AM78*a*, in terms of which the dissipative form of the GLM circulation theorem (6.3) is (Corollary III in AM78*a*; also B00):

$$\frac{\mathrm{d}}{\mathrm{d}t}\oint_{\mathscr{C}^L} (\overline{\boldsymbol{u}}^L - \boldsymbol{p} + \tfrac{1}{2}\boldsymbol{f}\times\boldsymbol{x})\cdot\mathrm{d}\boldsymbol{x} \;=\; \oint_{\mathscr{C}^L} (\overline{\boldsymbol{F}}^L - \mathscr{F})\cdot\mathrm{d}\boldsymbol{x}. \tag{9.1}$$

For *momentum-conserving forces* $\boldsymbol{F}$ (which derive from the divergence of a momentum flux tensor such as that due to viscosity) it is shown in B00 that $|\overline{\boldsymbol{F}}^L| \ll |\mathscr{F}|$ at $O(a^2)$ for slowly varying waves. Then the last term dominates on the right of (9.1).

We now consider the balanced mean-flow response to dissipation, in the strong-stratification limit. In place of (7.5*b*) we have

$$\overline{\mathrm{D}}^L\boldsymbol{p}_H + \nabla\cdot(\boldsymbol{p}_H\hat{\boldsymbol{c}}_g) + \nabla\overline{\boldsymbol{u}}^L\cdot\boldsymbol{p}_H \;=\; \frac{\boldsymbol{k}_H}{\hat{\omega}}\,\overline{(\boldsymbol{u}'\cdot\boldsymbol{F}')} \;\equiv\; \mathscr{F}_H, \tag{9.2}$$

say. The dissipative forms of (7.2*b*) and (7.13) for momentum-conserving forces are

$$\overline{\mathrm{D}}^L\overline{q}^L = -\hat{\boldsymbol{z}}\cdot\nabla\times\mathscr{F}_H \quad \text{and} \quad \overline{\mathrm{D}}^L(\hat{\boldsymbol{z}}\cdot\nabla\times\boldsymbol{p}_H) + \hat{\boldsymbol{z}}\cdot(\nabla\times(\nabla\cdot(\boldsymbol{p}_H\hat{\boldsymbol{c}}_g))) = +\hat{\boldsymbol{z}}\cdot\nabla\times\mathscr{F}_H. \tag{9.3}$$

It follows that, in the layerwise pseudomomentum-and-impulse budget (7.11) or (8.4), the left-hand equation is unchanged from its non-dissipative form whilst the right-hand equation, involving horizontal refraction, acquires a horizontal integral over $-\mathscr{F}_H$ on its right-hand side. In this way, dissipation irreversibly transmutes horizontal pseudomomentum into impulse.

Locally, we observe that dissipation has the *sole* effect of irreversibly transmuting the vertical curl of pseudomomentum into Lagrangian-mean potential vorticity in (9.3). The dissipation has no effect on the $\overline{\boldsymbol{u}}^L$ field. This is just what was hinted at in § 7.5. Viewed from the wave–vortex duality perspective suggested by figures 3 and 4, it means that dissipation transmutes the wavepacket on the left into the vortex pair on the right. *The transmutation leaves* $\overline{\boldsymbol{u}}^L$ *unchanged* because, in equation (7.3) or its quasi-geostrophic counterpart from (8.1) and (8.2), the quantity $\overline{q}^L + \widehat{\boldsymbol{z}} \cdot \boldsymbol{\nabla} \times \boldsymbol{p}_H$ stays the same and hence the inversion problem to find $\overline{\boldsymbol{u}}^L$ stays the same.

The effect of dissipation is therefore solely a matter of book-keeping, in that only the partitioning between the two terms of $\overline{q}^L + \widehat{\boldsymbol{z}} \cdot \boldsymbol{\nabla} \times \boldsymbol{p}_H$ is changed. So we have a particularly clear counterexample to the myth that the waves necessarily give 'their momentum' to the mean flow when they dissipate.

To further underline this point, one can imagine that $\boldsymbol{p}_H$ and $\overline{q}^L$ are time-evolved by a sequence of alternating non-dissipative and dissipative time steps, as could be the case in a numerical scheme based on operator splitting. During the non-dissipative time step both $\overline{q}^L$ and the vertical curl of $\boldsymbol{p}_H$ are advected by $\overline{\boldsymbol{u}}^L$. The curl of $\boldsymbol{p}_H$ is then, in addition, advected by $\hat{\boldsymbol{c}}_g$. During the following, dissipative time step, both change but in such a way that their sum stays constant in $\overline{q}^L + \widehat{\boldsymbol{z}} \cdot \boldsymbol{\nabla} \times \boldsymbol{p}_H$. Together with (7.1) this implies that there is *no mean-flow acceleration during the dissipative time step*. (Indeed this is true whether measured by $\overline{\boldsymbol{u}}^L$ or by $\overline{\boldsymbol{u}}$, since Stokes drifts are negligible.) What emerges instead is that the essential effect of dissipation is merely to make permanent a mean-flow change that has already occurred during the preceding non-dissipative step of the dynamics, such as the buildup of the Bretherton flow during the arrival of a wavepacket.

9.2. *Failure of the standard paradigm for secondary gravity-wave generation*

The point just made is obscured in the classical wave–mean theories that led to the standard paradigm. These assume two-dimensional, translationally or rotationally invariant backgrounds, i.e. backgrounds that depend on two coordinates only, say y and z, and steady wave fields whose amplitude $a(y, z)$ and other parameters such as wavenumbers or phase gradients likewise depend on y and z only, even though the phase itself varies in all three directions x, y and z. In such cases the regions of pseudomomentum dissipation and persistent mean-flow forcing necessarily coincide, so that persistent, cumulative mean-flow changes are always associated with pseudomomentum dissipation, as predicted by classical non-acceleration theorems (e.g. AM78*a*, p. 627).

The present work has shown, however, that there are three-dimensional wave–mean problems involving transient wavepackets, or steady wavetrains regarded as successions of transient wavepackets, in which the standard paradigm is thoroughly misleading. A case in point is the problem of estimating the generation of secondary, $O(a^2)$ gravity waves by the sudden breaking of a primary, $O(a)$ gravity wavepacket. The problem has been addressed in the atmospheric-dynamics literature (e.g. Zhu & Holton 1987; Vadas & Fritts 2001), using the standard paradigm. Thus, the $O(a^2)$ wave-induced force is modelled in these studies as a horizontal body force (*a*) that is localized at the wavepacket, (*b*) that has time scale equal to the time scale τ for wave dissipation, and (*c*) that has total volume and time integral equal to the packet-integrated horizontal pseudomomentum. Since rapid wave-dissipation events due to wave breaking may take place during a fraction of an intrinsic wave period $2\pi/\hat{\omega}$, the relative shortness of the time scale τ for such events would appear to argue for better matching to gravity-wave time scales, and stronger secondary gravity-wave generation.

However, according to the wave–mean interaction theory presented above, there is no reason why the actual mean-flow forcing time scale T should be equated to the time scale τ of the wave dissipation. Indeed, it is theoretically possible to conceive of a three-dimensional wavepacket that dissipates instantaneously at some time t_0, i.e. $\mathscr{F}_H = -\boldsymbol{p}_H\,\delta(t-t_0)$ in (9.2), making $\tau = 0$, yet generates *no secondary gravity waves at all*. This is because the actual mean-flow forcing time scale T on a given isentrope is not associated with dissipation, but rather with the prior setting-up of the Bretherton flow on that isentrope. The actual time scale T is therefore of the order of the non-dissipative propagation time scale of the wavepacket, i.e. of the order of its vertical extent divided by its vertical group velocity. If dissipation operates, then all it means is that the Bretherton flow is only partly undone if the remnants of the wavepacket leave the isentrope under consideration. That is the lasting, irreversible effect of wave dissipation, as discussed before.

In summary, there is no relation between the actual time scale T and the dissipative decay time τ. One can conceive of situations where there is a link between T and τ, but if the concept of a slowly varying wavepacket is to make sense to begin with – which gravity-wave observations say it often does – then the propagation time scale T will always be large in comparison with the intrinsic wave period (B69). Overall, this means that real secondary wave emission will tend to be much weaker than is suggested by the assumption $T \approx \tau$ based on the standard paradigm.

9.3. *Wave capture versus dissipation*

As already discussed, there is a strong similarity between wave capture and dissipation as far as the mean-flow response is concerned. For instance, consider that a wavepacket has arrived from below, on a given isentrope, and that therefore the Bretherton flow at $O(a^2)$ has been set up on this isentrope. If the wavepacket were to dissipate instantaneously at this stage, then $\overline{q}^L$ would absorb the pseudomomentum curl whilst keeping the right-hand side of (7.3) constant. Clearly, the Bretherton flow would not change during the dissipation, as discussed above. After the wavepacket has disappeared, the implication of (7.2) is that the vertical curl of $\overline{\boldsymbol{u}}^L$ is a mean material invariant, and the flow would evolve according to such invariance together with (7.1).

Let us now consider a situation where rapid wave dissipation is replaced by wave capture, i.e. where wave amplitude a is so small that the final dissipative stage is significantly postponed until well after the freezing of the group velocity. As discussed below (7.15), during wave capture the vertical curl of pseudomomentum becomes a mean material invariant and therefore (7.2) implies that the vertical curl of $\overline{\boldsymbol{u}}^L$ itself becomes a mean material invariant, just as in the dissipative case. Indeed, as far as the mean flow is concerned, this is precisely *as if* the wavepacket had dissipated, even though the wavepacket can still carry significant $O(a^2)$ disturbance energy. Apparently, that energy plays no further role in the mean-flow evolution.

So once again we find a peculiar similarity between wave capture and wave dissipation: in both cases the mean flow is irreversibly changed by adding the vertical curl of $\boldsymbol{p}_H$ to that of $\overline{\boldsymbol{u}}^L$ and then simply evolving the vertical curl of $\overline{\boldsymbol{u}}^L$ via advection by $\overline{\boldsymbol{u}}^L$. In the case of dissipation, the irreversible nature of this process is overtly on display through the changes in $\overline{q}^L$ due to the vertical curl of the dissipative mean force $\mathscr{F}_H$, determined by (9.3). In the case of capture, irreversibility arises less overtly, because of the exponential straining of the wavenumber vector that renders the vertical curl of $\boldsymbol{p}_H$ a mean material invariant, as noted in (7.15). This straining process will usually be irreversible in practice.

Thus the main implication, especially for practical considerations involving gravity-wave parametrizations in general circulation models for the atmosphere, is that the eventual dissipation of a captured wave is of no consequence as far as the mean flow is concerned. Such dissipation can be treated merely as a book-keeping exercise in which the right-hand side of (7.3*b*) or its quasi-geostrophic counterpart, which is already materially invariant, is absorbed into $\overline{q}^L$ and then $\boldsymbol{p}_H$ set to zero.

10. A topographic wavepacket example

We now give the full details for the example of figure 4, reverting to $f = 0$.

10.1. *Background flow set-up and linear ray-tracing*

Consider a steady background flow as sketched in figure 4 above, in the strong-stratification limit, so that the flow-induced undulations in the background isentropes are negligible. The flow consists of a uniform flow $U_s > 0$ along the x-axis plus a horizontal vortex pair located at $x_d > 0$ and symmetric across the x-axis. The dipole strength of the vortex pair is arranged such that the entire background flow is steady, which is straightforward if we assume that the vertical extent of the vortex pair is sufficiently deep to allow two-dimensional fluid dynamics to be accurate for the background flow.† By (7.4) and (7.10), the layerwise impulse $\boldsymbol{J}$ of this vortex pair is proportional to, and at right angles to, the distance between the vortices. It points in the negative x-direction here.

We now consider the evolution of a small-scale wavepacket that initially propagates vertically upwards through this background flow. For definiteness, we could imagine that this wavepacket has initial wavenumber $\boldsymbol{k}_0 = (k_0, l_0, m_0)$ and initial absolute frequency $\omega_0 = 0$, as if it had been generated by the flow of the surface wind $U_0 > 0$ over corrugated bottom topography at $x = y = z = 0$. We assume that the initial value of the wavenumber y-component $l_0 = 0$, as would be the case if the topography were corrugated in the x-direction only; and we also assume that the wavepacket is quasi-hydrostatic in the sense that $k_0^2 \ll m_0^2$ and $\hat{\omega}_0^2 = U_0^2 k_0^2 \ll N^2$. It will be verified shortly that such a wavepacket propagates vertically upward until it begins to be affected by the horizontal refraction.

As long as the background flow remains steady, z-independent, and symmetric across the x-axis, it follows from the ray-tracing equations, including (2.3) and (3.7), that

$$\omega(t) = 0, \quad m(t) = m_0, \quad l(t) = 0, \quad y(t) = 0 \tag{10.1}$$

hold along the wavepacket's group-velocity ray. As $l(t) = 0$, we also have

$$\omega = \hat{\omega} + \boldsymbol{U} \cdot \boldsymbol{k} = \hat{\omega} + Uk + Vl = \hat{\omega} + Uk = 0, \tag{10.2}$$

and therefore the radiation condition $\hat{w}_g > 0$ together with the sign convention $\hat{\omega} > 0$ imply $k(t) < 0$ and $m(t) < 0$. This shows that $\boldsymbol{p}_H$ points in the negative x-direction, as indicated by the large arrows in figure 4.

With (10.2), the horizontal group velocity can be read off from (3.3) and (3.1) with $f = 0$, hence $\kappa = N/U$, as

$$u_g = \frac{\mathrm{d}_g x}{\mathrm{d}t} = U + \hat{u}_g = U\,\frac{\hat{\omega}^2}{N^2} = U\,\frac{U^2 k^2}{N^2}, \quad \text{and} \quad \frac{\mathrm{d}_g k}{\mathrm{d}t} = -\frac{\partial U}{\partial x} k \tag{10.3}$$

† This also implies that vertical mean-flow derivatives are negligible, so in this respect the present example differs from the generic three-dimensional scenario considered earlier and is more similar to well-known examples of horizontal critical layers (see J69, BS93, and references in Staquet & Sommeria 2002).

follows from (2.7). The latter shows that $|k|$ increases if $\partial U/\partial x$ is negative. Again since $\kappa = N/U$ we see that

$$k^2 - k_0^2 = N^2\left(\frac{1}{U^2} - \frac{1}{U_0^2}\right) \tag{10.4}$$

along the ray, where $U_0 \approx U_s$.† Substituting this into (10.3) gives the useful relation

$$u_g = U + \hat{u}_g = U\left(1 - \frac{U^2}{U_0^2}\right) + \frac{U^3 k_0^2}{N^2}. \tag{10.5}$$

10.2. *Wave capturing sequence*

To begin with, $u_g \ll U$ so that the negative intrinsic group velocity in x holds the quasi-hydrostatic wave steady against the wind $U_0 \approx U_s > 0$. However, $|k(t)|$ grows exponentially because on the centreline $y = 0$ the mean flow $U > 0$ decreases with x, which squeezes the phase lines together. Indeed, for this particular flow the x-direction is the axis of contraction on $y = 0$, and therefore $\boldsymbol{k}_H = (k, 0)$ is a growing wavenumber eigenmode. Non-hydrostatic effects become more important as $|k/m|$ grows and this leads to a decrease in $|\hat{u}_g|$ and therefore to a growing $u_g > 0$: the wavepacket begins to drift to the right.

By (10.4), increasing values of $|k|$ match decreasing background wind values U, which are found to the right along $y = 0$. The intrinsic group velocity is continually diminished until we reach the capturing regime $|\hat{u}_g| \ll U$. This is easy to see from (10.5), which implies that $|\hat{u}_g/U| \approx U^2/U_0^2$ and hence that perfect capture with $\hat{u}_g = 0$ would occur at the upstream stagnation point $U = 0$ of the vortex pair, at $x = x_A$, say. Of course, the captured wavepacket would reach this point only in infinite time.‡

The exponentially growing $|k|$ implies the growth of $|\boldsymbol{p}_H|$ during wave capture. Another view of the surge in $|\boldsymbol{p}_H|$ comes from (5.4). Concomitant with the mean-flow squeeze in x is an extension of the wavepacket in y to preserve area (apart from small corrections due to non-zero $\nabla \cdot \hat{\boldsymbol{c}}_g$ and $\hat{w}_g$). This extension in y separates the regions of oppositely signed $\widehat{\boldsymbol{z}} \cdot \nabla \times \boldsymbol{p}_H$ and thus increases the pseudomomentum content of the wavepacket according to (5.4), at an exponentially increasing rate.

During capture the action density A is approximately constant and as $\hat{\omega} \to N$ near the stagnation point this implies that the energy density $E = \hat{\omega}A$ goes to a constant as well. The unbounded growth in k^2 then implies unbounded shear $\overline{w_x'^2} \propto k^2 E$, which, as pointed out in § 4, must lead to instability and nonlinear breakdown of the wavepacket before it reaches the stagnation point.

10.3. *Mean-flow response*

The $O(a^2)$ Bretherton flow is indicated by the dashed streamlines in figure 4. As can be seen by inspection of figure 4, the advection of the background vortices by this flow is such as to decrease the magnitude of the impulse of the background

† This could be made more quantitative by considering the limit of large x_d such that the two-dimensional flow due to the vortex pair, which decays with inverse square distance from the location of the vortex pair, is negligible at $x = y = 0$.

‡ Interestingly, the JWKB approximation underlying ray theory appears to remain valid as the wavepacket approaches x_A. Near x_A the background flow U on $y = 0$ can be Taylor expanded as $U \propto -(x - x_A)$, which implies that the relevant JWKB parameter $|k(x - x_A)| \propto kU \propto \hat{\omega} \approx N$, as k/m becomes large. Therefore this parameter remains finite as the packet approaches the stagnation point, so if ray tracing worked initially then it will continue to work even as $x(t) \to x_A$. However, the survival of the JWKB approximation merely masks the overall breakdown of linear theory near x_A.

vortex pair. This is consistent with the $\boldsymbol{P}_H + \boldsymbol{I}$ conservation theorem: the growth in $|\boldsymbol{P}_H|$ must occur at the expense of $\boldsymbol{I}$ and $\boldsymbol{J}$. A regular perturbation analysis keeps the $O(1)$ vortices fixed, but one can see from the conservation theorem how this must fail at large times $t = O(a^{-2})$: over such long times $\boldsymbol{I}$ and $\boldsymbol{J}$ will have decreased in magnitude by an $O(1)$ amount, and hence the mean flow will have been significantly modified. These mean-flow changes would then have to be allowed to feed back on the wavepacket, and in this way the perturbation analysis could be extended to such long times. The above makes clear that the exponential growth stage in the evolution of $\boldsymbol{P}_H$ does not persist forever: $|\boldsymbol{P}_H|$ can grow only at the expense of the finite vortex impulse, and so there must be growth saturation at $t = O(a^{-2})$.

It is intriguing to speculate about the flow features that might be observed over such long times. The idea of wave–vortex duality discussed in § 5 reminds us that the captured wavepacket behaves like a vortex pair or vortex dipole, with impulse equal to $\boldsymbol{P}_H$. As already remarked in § 7.5, we have something like an incipient 'leap-frogging' situation. However, it is unclear what will happen in the later stages, when the wavepacket would become wrapped laterally around the vortex pair and presumably breaks down into passively advected filaments. This may be accompanied or preceded by wave dissipation and the irreversible generation of potential vorticity anomalies $\overline{q}^L$, as described in § 9.1. Idealized numerical simulations would be useful to test these speculative scenarios, especially if such simulations could be performed in a highly resolved two-dimensional simulation. The need for very high spatial resolution is dictated by the inevitable scale separation between mean-flow vortices and the crests of a captured wavepacket, though it is possible that some progress could be made by a numerical ray-tracing scheme such as that developed and used in Nazarenko *et al.* (1995).

11. Concluding remarks

This work has given us some important new insights into the mechanisms involved in the missing-forces problem and into the general nature of what BM03 termed 'vortex-core recoil'. As pointed out in § 7.2, the problem solved in BM03 comes under the general heading of wave–vortex problems that satisfy the joint conservation of Kelvin impulse $\boldsymbol{I}$ and horizontal wave pseudomomentum $\boldsymbol{P}_H$ – see especially (7.9), (7.11), (8.3) and (8.4) – justifying the notion of a $\boldsymbol{P}_H + \boldsymbol{I}$ *budget*.

We stress that a two-dimensional vortex monopole whose core is advected relative to its surroundings has a well-defined *rate of change* of Kelvin impulse, just equal and opposite to the rate of change of pseudomomentum due to wave refraction. We can therefore say that a 'pseudomomentum rule' holds in just that sense: rates of change of pseudomomentum due to mean-flow refraction are linked to opposite rates of change of impulse, identifiable with equivalent forces on vortex cores in the absence of waves – thus the effective location of the forces need not be the same as the location of the pseudomomentum changes. Action at a distance is involved in the interactions of wave packets and vortices, just as it is in ordinary vortex interactions, and in any motion whatever to which the notions of balance and PV inversion apply.

The next major challenge, already noted in § 7.1, is how to generalize – to what extent we can generalize – the impulse concept to make it applicable to the less idealized flows dealt with in realistic weather and climate models. These are the most general flows to which the notions of balance and PV inversion apply. As it stands, the impulse concept depends on the strong-stratification limit, banishing far-field recoil all the way to infinity, which is much more restrictive. An even greater challenge, partly

for technical reasons relating to the parallelized computer codes used in atmospheric models, is to devise practical 'gravity-wave parametrization schemes' that will respect the appropriately generalized $\boldsymbol{P}_H + \boldsymbol{I}$ budgets.

The guiding principle that has emerged can be stated as an expanded aphorism, focusing on the simple exact relations summarized in §6 and §8: 'What's simple is everything related to PV and Kelvin circulation; what's complicated is everything related to balance, PV inversion, spontaneous-adjustment emission and far-field recoil – the emission of large-scale gravity waves that export pulses of momentum to distant parts of the fluid domain, finite or infinite.'

We thank the referees for constructive comments that produced many significant clarifications in the presentation. Carlo Barenghi, Natalia Berloff, Ulf Leonhardt, Duncan O'Dell and Sergey Nazarenko kindly helped to update us on aspects of the quantum fluids problem. M. E. M. is indebted to the UK Natural Environmental Research Council for support under grant NER/A/S/2000/00452. Part of this work was done whilst O. B. participated in the network "Geometrical methods in geophysical fluid dynamics" funded by the UK Engineering and Physical Sciences Research Council.

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 AM78a).

Andrews, D. G. & McIntyre, M. E. 1978*b* On wave-action and its relatives. *J. Fluid Mech.* **89**, 647–664 (referred to herein as AM78b); and Corrigendum **95**, 796.

Badulin, S. I. & Shrira, V. I. 1993 On the irreversibility of internal waves dynamics due to wave trapping by mean flow inhomogeneities. Part 1. Local analysis *J. Fluid Mech.* **251**, 21–53. (referred to herein as BS93).

Batchelor, G. K. 1967 *An Introduction to Fluid Dynamics.* Cambridge University Press, 615 pp.

Bretherton, F. P. 1969 On the mean motion induced by internal gravity waves. *J. Fluid Mech.* **36**, 785–803 (referred to herein as B69).

Bretherton, F. P. & Garrett, C. J. R. 1968 Wavetrains in inhomogeneous moving media. *Proc. Ry. Soc. Lond.* A **302**, 529–554.

Broutman, D., Rottman, J. W. & Eckermann, S. D. 2004 Ray methods for internal waves in the atmosphere and ocean. *Annu. Rev. Fluid Mech.* **36**, 233–253.

Bühler, O. 2000 On the vorticity transport due to dissipating or breaking waves in shallow-water flow. *J. Fluid Mech.* **407**, 235–263 (referred to herein as B00).

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 BM98).

Bühler, O. & McIntyre, M. E. 2003 Remote recoil: a new wave–mean interaction effect. *J. Fluid Mech.* **492**, 207–230 (referred to herein as BM03).

Donnelly, R. J. 1991 *Quantized Vortices in Helium II.* Cambridge University Press, 346 pp.

Fritts, D. C. & Alexander, M. J. 2003 Gravity-wave dynamics and effects in the middle atmosphere. *Rev. Geophys.* **41**(1), doi:10.1029/2001RG000106.

Garrett, C. J. R. & Munk, W. 1975 Space–time scales of internal waves: a progress report. *J. Geophys. Res.* **80**, 291–297.

Gjaja, I. & Holm, D. D. 1996 Self-consistent Hamiltonian dynamics of wave mean-flow interaction for a rotating stratified incompressible fluid. *Physica D* **98**, 343–378.

Hasselmann, K. 1967 A criterion for non-linear wave stability. *J. Fluid Mech.* **30**, 737–739.

Haynes, P. H. & Anglade, J. 1997 The vertical-scale cascade of atmospheric tracers due to large-scale differential advection. *J. Atmos. Sci.* **54**, 1121–1136.

Hoskins, B. J., McIntyre, M. E. & Robertson, A. W. 1985 On the use and significance of isentropic potential-vorticity maps. *Q. J. R. Met. Soc.* **111**, 877–946.

Hoskins, B. J., McIntyre, M. E. & Robertson, A. W. 1987 Reply to comments by J. S. A. Green. *Q. J. R. Met. Soc.* **113**, 402–404.

Jones, W. L. 1969 Ray tracing for internal gravity waves. *J. Geophys. Res.* **74**, 2028–2033 (referred to herein as J69).

McEwan, A. D. 1971 Degeneration of resonantly-excited internal gravity waves. *J. Fluid Mech.* **50**, 431–448.

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

McIntyre, M. E. 2003 On global-scale atmospheric circulations. In *Perspectives in Fluid Dynamics: A Collective Introduction to Current Research* (ed. G. K. Batchelor, H. K. Moffatt, M. G. Worster), pp. 557–624. Cambridge University Press, 631 pp. (Paperback edition, with important corrections.)

McIntyre, M. E. & Norton, W. A. 2000 Potential-vorticity inversion on a hemisphere. *J. Atmos. Sci.* **57**, 1214–1235; and Corrigendum **58**, 949.

McIntyre, M. E. & Roulstone, I. 1996 Hamiltonian balanced models: constraints, slow manifolds and velocity splitting. *Forecasting Research*, Scientific Paper No. 41, pp. 1–49. UK Met Office.

Nazarenko, S. V., Zabusky, N. J. & Scheidegger, T. 1995 Nonlinear sound–vortex interactions in an inviscid isentropic fluid: A two-fluid model. *Phys. Fluids* **7**, 2407–2419.

Polzin, K. 2004 A heuristic description of internal wave dynamics. *J. Phys. Oceanogr.* **34**, 214–230.

Rayleigh, Lord 1896 *The Theory of Sound, Volume 2.* New York, Dover (reprinted 1945), 504 pp.

Saffman, P. G. 1993 *Vortex Dynamics.* Cambridge University Press, 336 pp.

Sonmor, L. J. & Klaassen, G. P. 1997 Toward a unified theory of gravity wave stability. *J. Atmos. Sci.* **54**, 2655–2680.

Staquet, C. & Sommeria, J. 2002 Internal gravity waves: from instabilities to turbulence. *Annu. Rev. Fluid Mech.* **34**, 559–594.

Stone, M. 2000 Iordanskii force and the gravitational Aharonov–Bohm effect for a moving vortex. *Phys. Rev.* B **61**, 11780–11786.

Vadas, S. L. & Fritts, D. C. 2001 Gravity wave radiation and mean responses to local body forces in the atmosphere. *J. Atmos. Sci.* **58**, 2249–2279.

Whitham, G. B. 1974 *Linear and Nonlinear Waves.* New York, Wiley-Interscience, 620 pp.

Zhu, X. & Holton, J. 1987 Mean fields induced by local gravity-wave forcing in the middle atmosphere. *J. Atmos. Sci.* **44**, 620–630.