J. Fluid Mech. (2026), *vol.* 1031, A40, doi:10.1017/jfm.2026.11336

JFM PAPERS

Trapping of near-inertial waves and critical layer formation in baroclinic currents

James P. Hilditch[1] and **Leif N. Thomas**[1]

[1]Department of Earth System Science, Stanford University, Stanford, CA 94305, USA

Corresponding author: James P. Hilditch, hilditch@stanford.edu

(Received 10 September 2025; revised 10 January 2026; accepted 16 February 2026)

The behaviour of near-inertial waves (NIWs) in baroclinic currents is investigated, with a focus on wave trapping and critical layer dynamics. We present theoretical analysis, supported by numerical simulations, of the Sawyer–Eliassen equation, which describes waves in the plane perpendicular to an arbitrary balanced background flow. Gradients of the background velocity and buoyancy field modulate the wave properties, in particular defining the range of frequencies for which the Sawyer–Eliassen equation is hyperbolic, i.e. wave-like. Variations in the lower limit of this range, the local minimum frequency, lead to the trapping of low-frequency, typically sub-inertial, waves through either total internal reflection or the formation of critical layers. Both mechanisms are studied. We introduce a local coordinate rotation that not only elucidates these dynamics by simplifying the governing equation, but also allows us, through direct analogy, to draw upon theory and intuition developed for barotropic problems. In the majority of physically relevant cases, the transformed coordinates are aligned with and perpendicular to isopycnals, and are thus easily utilised. Employing the coordinate transformation, we consider along-isopycnal modes, and study the behaviour of waves approaching a critical level without the need for a full ray-tracing approximation. Finally, we find qualitative differences in the critical layers that form in strongly and weakly baroclinic flows, most notably in their location. In the weakly baroclinic case, NIW energy accumulates about the minimum in the relative vorticity of the background flow, whereas in the strongly baroclinic case, we find slantwise critical layers concentrated in fronts.

Key words: baroclinic flows, internal waves, waves in rotating fluids

1. Introduction

Critical layers are a prominent feature of wave dynamics. They arise when the governing wave equations become singular, signifying that the physical assumptions made in deriving them – e.g. that the wave dynamics are linear and inviscid – are no longer justifiable (Haynes *et al.* 2015). Solutions may be regularised by reintroducing the neglected physics in a thin region – the critical layer – about the singularity. Here, we consider inertia-gravity wave critical layers, which are associated with secular growth in the wave shear, wave breaking, elevated dissipation and enhanced mixing (Kunze 1986; Qu *et al.* 2021*b*). Furthermore, the convergence of the wave momentum flux across the critical layer drives a mean flow response. This forcing is an important component of inertia-gravity wave parametrisations in climate models (e.g. Kruse *et al.* 2023). There are two predominant mechanisms through which inertia-gravity wave critical layers form, both of which may be classified as wave–mean flow interactions. The first involves the Doppler shifting of wave frequency in a vertically sheared flow. For example, Booker & Bretherton (1967) studied the critical layers of the Taylor–Goldstein equation that form whenever the phase velocity of the waves matches the fluid velocity of the background flow. However, here we are interested in the second mechanism, in which we consider the range of intrinsic frequencies supported by a given background flow. In rotating, stratified flows, the supported range of frequencies has finite upper and lower bounds that depend on the properties of the background flow. Certain frequencies of waves may be supported only in some subset of the entire domain. These waves are trapped and can propagate only so far until they either undergo a reflection or enter a critical layer.

In the ocean, the foremost examples of these critical layer dynamics are found at the base of anticyclones (Kunze 1986), eddies with vertical vorticity, $\zeta := \partial v_g/\partial x - \partial u_g/\partial y$, opposite in sign to the Coriolis frequency f. Here, the vertical vorticity of the background flow determines the effective Coriolis frequency f_{eff}, given for small Rossby number $|\zeta/f| \ll 1$ by

$$f_{eff} = f + \frac{1}{2}\zeta \tag{1.1}$$

(Weller 1982; Kunze 1985). In weakly baroclinic flows, the effective Coriolis frequency represents the lower limit of the range of frequencies of freely propagating waves. As a result, sub-inertial waves are supported by anticyclones. However, these waves are trapped in the region of low f_{eff} both in the horizontal and the vertical. As the vertical vorticity of the background flow decays with depth, the waves' vertical propagation will stall as they approach a critical level that they cannot cross (Kunze 1986). While amplified wave energy and elevated dissipation at the base of anticyclones, suggestive of critical layers, have been observed (Kunze, Schmitt & Toole 1995; Joyce *et al.* 2013; Lelong, Cuypers & Bouruet-Aubertot 2020; Lazaneo *et al.* 2024), the fate of this energy has not been conclusively resolved (Kunze *et al.* 1995; Thomas *et al.* 2024).

Extending this picture to the submesoscale regime where baroclinic effects – i.e. the physics associated with horizontal buoyancy gradients – play a more prevalent role is far from trivial. Not only does this modify the wave dynamics, but it greatly complicates the mathematical treatment of the problem since the waves can no longer be projected onto uncoupled vertical modes, as is typical for the barotropic problem (e.g. Balmforth *et al.* 1998; Rocha, Wagner & Young 2018; Asselin *et al.* 2020). One approach that has proved very profitable across a number of problems is to consider the ageostrophic dynamics in a two-dimensional (2-D) plane perpendicular to a front. It is generally justified by the inherently anisotropic structure of fronts and filaments. Neglecting the along-front variability precludes some processes such as baroclinic or barotropic instability, as well as the effects of front curvature. However, this is a small price to pay in order to make the problem mathematically tractable. This approach has been widely used to study frontogenesis (Hoskins 1982; McWilliams 2017), inertial-symmetric instability (Taylor &

Ferrari 2009; Grisouard 2018), parametric subharmonic instability (Thomas & Taylor 2014; Hilditch & Thomas 2023) and inertia-gravity waves (Mooers 1975; Whitt & Thomas 2013), to name just a few examples.

In this 2-D setting, the ageostrophic dynamics about a geostrophically balanced background state, including near-inertial waves (NIWs), is described by the time-dependent Sawyer–Eliassen equation (Sawyer 1956; Eliassen 1962). Mooers (1975) investigated the propagation of inertia-gravity waves in this model, and highlighted differences between the barotropic and baroclinic dynamics. First, in baroclinic flows there exists an anomalously low frequency band in which waves with frequencies less than the effective Coriolis frequency may exist. The waves in this band have the properties that their characteristics have the same sign slope, and that when employing ray-tracing, the shallower of these two characteristics, or rays, has the unusual behaviour that the vertical group and phase velocities have the same sign. Furthermore, waves at the minimum frequency have slantwise particle displacements and phase lines. However, under the hydrostatic approximation, the displacements remain aligned with isopycnals as in the barotropic case. Continuing these ray-tracing and group velocity ideas, Whitt & Thomas (2013) argued for the existence of slantwise critical layers whenever isopycnals align with the separatrix – defined in this context as the curve along which the wave frequency is equal to the minimum supported frequency. Qu *et al.* (2021*b*) employed this theory to show that critical layers for inertial waves – waves with frequency equal to f – form over sloping bathymetry, a mechanism with some observational support (Qu *et al.* 2021*a*). More generally, enhanced dissipation associated with NIWs has been observed in baroclinic currents (Inoue, Gregg & Harcourt 2010; Fer *et al.* 2018).

Here, we present a mathematical framework that reconciles the baroclinic and barotropic dynamics, allowing us to apply ideas from the vast literature of barotropic wave dynamics to the baroclinic case. Moreover, it elucidates the physics of inertia-gravity wave trapping via both total internal reflections and the formation of slantwise critical layers. The paper is organised as follows. First, in § 2, we derive the Sawyer–Eliassen equation and its key properties, largely following Mooers (1975). Then the main theoretical thrust of this work – the introduction of a locally rotated coordinate system – is described in § 3. To demonstrate the utility of the rotated coordinate system and gain some intuition, we consider, in § 4, the problem of a wavepacket undergoing total internal reflection in a simple model. Then in § 5, we consider the trapping of NIWs in critical layers in a more complicated background flow representative of the submesoscale – i.e. with $O(1)$ Rossby numbers and balanced Richardson numbers. Finally, in § 6, we conclude with a discussion of the implications of our results and directions for future research.

2. The Sawyer–Eliassen equation

The Sawyer–Eliassen equation describes inviscid, linear, ageostrophic motions about a geostrophically and hydrostatically balanced background state in two dimensions. The background flow is defined by a geostrophic velocity $V(x, z, t)$ (oriented along the y axis) and buoyancy field $B(x, z, t)$, where x, y, z are Cartesian coordinates with z vertical, and t is time. The assumption of geostrophic and hydrostatic balance gives rise to the thermal wind relation

$$\frac{\partial B}{\partial x} = f\frac{\partial V}{\partial z}, \tag{2.1}$$

but V and B may otherwise be arbitrary functions of x, z and t. However, we will later restrict attention to steady background flows. Here, f is taken to be constant, and the background buoyancy is related to the background density $\rho(x, z, t)$ by $B := g(\rho_0 - \rho)/\rho_0$, where g is the gravitational acceleration, and ρ_0 is the Boussinesq reference density.

2.1. *Derivation*

We linearise the non-hydrostatic, inviscid, adiabatic Boussinesq equations about this background state, and make the assumption that the dynamics are independent of the along-front coordinate y. The governing equations are

$$\frac{\partial u}{\partial t} - fv + \frac{\partial \phi}{\partial x} = 0, \tag{2.2a}$$

$$\frac{\partial v}{\partial t} + u\frac{\partial V}{\partial x} + w\frac{\partial V}{\partial z} + fu = 0, \tag{2.2b}$$

$$\frac{\partial w}{\partial t} - b + \frac{\partial \phi}{\partial z} = 0, \tag{2.2c}$$

$$\frac{\partial b}{\partial t} + u\frac{\partial B}{\partial x} + w\frac{\partial B}{\partial z} = 0, \tag{2.2d}$$

$$\frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} = 0, \tag{2.2e}$$

where u, v, w are the perturbation velocities, b is the perturbation buoyancy, and ϕ is the perturbation (kinematic) pressure. We introduce a streamfunction ψ such that $u = -\partial\psi/\partial z$ and $w = \partial\psi/\partial x$, and form an equation for the evolution of the y component of the perturbation vorticity due to the thermal wind imbalance:

$$-\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2}\right)\frac{\partial \psi}{\partial t} = f\frac{\partial v}{\partial z} - \frac{\partial b}{\partial x}. \tag{2.3}$$

The evolutions of v and b are governed by the material conservation of absolute momentum $V + v + fx$ and total buoyancy $B + b$, such that after linearising,

$$\frac{\partial v}{\partial t} = -\mathcal{J}(\psi, V + fx), \tag{2.4a}$$

$$\frac{\partial b}{\partial t} = -\mathcal{J}(\psi, B), \tag{2.4b}$$

where the operator $\mathcal{J}(\psi, \cdot) = (\partial\psi/\partial x)\,\partial/\partial z - (\partial\psi/\partial z)\,\partial/\partial x$ captures the advection of the background flow by the perturbations. Equation (2.3) can then be reduced to a second-order in time equation for the streamfunction that may be written in the concise form

$$\left(\frac{\partial^2}{\partial t^2}\frac{\partial^2}{\partial x_i^2} + \frac{\partial}{\partial x_i}A_{ij}\frac{\partial}{\partial x_j}\right)\psi = 0, \tag{2.5a}$$

where the A_{ij} are the components of

$$\boldsymbol{A}(\boldsymbol{x}, t) := \begin{pmatrix} \dfrac{\partial B}{\partial z} & -\dfrac{\partial B}{\partial x} \\ -f\dfrac{\partial V}{\partial z} & f\left(f + \dfrac{\partial V}{\partial x}\right) \end{pmatrix}, \tag{2.5b}$$

for $\boldsymbol{x} := (x, z)^{\mathrm{T}} \equiv (x_1, x_2)^{\mathrm{T}}$, and summation over repeated indicies is understood here and throughout the paper.

Equation (2.5) is the time-dependent, non-hydrostatic Sawyer–Eliassen equation with non-constant coefficients. The assumption that the background flow is in thermal wind balance (2.1) not only makes $\boldsymbol{A}$ symmetric but also means that $\partial A_{ij}/\partial x_i = 0$, thus the Sawyer–Eliassen equation can be written as

$$\left(\delta_{ij}\frac{\partial^2}{\partial t^2} + A_{ij}\right)\frac{\partial^2 \psi}{\partial x_i\,\partial x_j} = 0, \tag{2.6}$$

where δ_{ij} is the Kronecker delta.

Henceforth, we make the assumption that the background flow is steady, thus it is often useful to work in the frequency domain in which the different components are uncoupled. Taking a Fourier transform in time, $\hat{\psi} = \int \psi \exp(\mathrm{i}\omega t)\,\mathrm{d}t$, or when solving for the normal modes, the Sawyer–Eliassen equation becomes

$$(A_{ij} - \omega^2\delta_{ij})\frac{\partial^2 \hat{\psi}}{\partial x_i\,\partial x_j} = 0. \tag{2.7}$$

2.2. *Alternative formulations in terms of displacements*

The mathematical symmetries of the problem, which is invariant under the changes $x \leftrightarrow z$, $u \leftrightarrow w$, $fv \leftrightarrow b$ and $V + fx \leftrightarrow B$, motivates reinterpreting the problem as purely 2-D in x and z. To do so, we treat the Coriolis force fv in the across-front momentum equation analogously to the buoyancy force in the vertical momentum equation. In this 2-D treatment, v is considered as not a velocity, but merely the perturbation of a conserved quantity, namely the absolute momentum. Furthermore, the conservation of absolute momentum will give rise to a potential energy in the same way that the conservation of buoyancy gives rise to the gravitational potential energy. Moreover, given the steady background flows, we can dispense with v and b by working with the displacements X and Z instead.

We define

$$\boldsymbol{X} := \begin{pmatrix} X \\ Z \end{pmatrix}, \quad \boldsymbol{u} := \begin{pmatrix} u \\ w \end{pmatrix} \equiv \frac{\partial \boldsymbol{X}}{\partial t}, \quad \boldsymbol{b} := \begin{pmatrix} fv \\ b \end{pmatrix}, \quad \boldsymbol{M} := \begin{pmatrix} f\left(f + \dfrac{\partial V}{\partial x}\right) & \dfrac{\partial B}{\partial x} \\ \dfrac{\partial B}{\partial x} & \dfrac{\partial B}{\partial z} \end{pmatrix}. \tag{2.8a–d}$$

Note that the matrix $\boldsymbol{M}$ is proportional to $\boldsymbol{A}^{-1}$. With these definitions the conservation of absolute momentum and total buoyancy (2.4) can be written as

$$\frac{\partial}{\partial t}(\boldsymbol{b} + \boldsymbol{M}\boldsymbol{X}) = \boldsymbol{0}. \tag{2.9}$$

Then assuming that the Ertel potential vorticity of the background flow,

$$Q := \left(f + \frac{\partial V}{\partial x}\right)\frac{\partial B}{\partial z} - \frac{\partial V}{\partial z}\frac{\partial B}{\partial x}, \tag{2.10}$$

is non-zero such that $\mathrm{Det}(\boldsymbol{M}) \equiv fQ \neq 0$, we define the displacements by

$$\boldsymbol{X} := -\boldsymbol{M}^{-1}\boldsymbol{b} \equiv -\frac{1}{fQ}\boldsymbol{A}\boldsymbol{b}. \tag{2.11}$$

Expressing the across-front and vertical momentum equations, (2.2a) and (2.2c), in terms of the displacements, we have

$$\frac{\partial^2 \boldsymbol{X}}{\partial t^2} + \boldsymbol{M}\boldsymbol{X} + \nabla\phi = \boldsymbol{0}, \tag{2.12}$$

where ∇ is the 2-D gradient operator. Integrating the continuity equation (2.2e) in time introduces an arbitrary function of space

$$\nabla \cdot \boldsymbol{X} = \mathcal{D}(\boldsymbol{x}). \tag{2.13}$$

It is tempting to set $\mathcal{D}(\boldsymbol{x}) = 0$ as would be appropriate for wave solutions. The Sawyer–Eliassen equation, subject to suitable boundary conditions (e.g. periodic in the horizontal and $\psi = 0$ in the vertical), does not admit steady solutions, i.e. a vortical mode. A steady solution to (2.5a) satisfies

$$\psi\left(\frac{\partial}{\partial x_i} A_{ij} \frac{\partial}{\partial x_j}\right)\psi = 0 \implies \iint A_{ij}\frac{\partial\psi}{\partial x_i}\frac{\partial\psi}{\partial x_j}\,\mathrm{d}\boldsymbol{x} = 0, \tag{2.14}$$

where the integral is taken over the whole domain, and we have integrated by parts. Assuming that $\boldsymbol{A}$ is positive definite everywhere (see § 3.1), the only solution is ψ constant. However, even with $u = w = 0$, the linearised Boussinesq equations admit steady solutions in the form of v and b perturbations in thermal wind balance. These steady solutions correspond to non-zero X and Z, and thus a potentially non-zero $\mathcal{D}(\boldsymbol{x})$.

The system (2.12)–(2.13) may be derived from a variational principle with Lagrangian density

$$\mathcal{L} := \frac{1}{2}\left|\frac{\partial \boldsymbol{X}}{\partial t}\right|^2 - \frac{1}{2}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{X} + \phi\left(\nabla\cdot\boldsymbol{X} - \mathcal{D}\right). \tag{2.15}$$

Through Whitham averaging (Whitham 1965), i.e. averaging the Lagrangian over wave phase, the dispersion relation and conservation of wave action may be derived. In particular, the slowly varying waveguide theories of Bretherton (1968) and Hayes (1970) that we call upon later are formulated in terms of the phase-averaged Lagrangian. Most usefully, the variational formulation is agnostic to the choice of coordinate system.

Finally, the conservation of energy immediately follows from contracting (2.12) with $\boldsymbol{u}$:

$$\frac{\partial \mathcal{E}}{\partial t} + \nabla\cdot\boldsymbol{\mathcal{F}} = 0, \tag{2.16a}$$

where

$$\mathcal{E} := \frac{1}{2}\left|\boldsymbol{u}\right|^2 + \frac{1}{2}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{X}, \quad \boldsymbol{\mathcal{F}} := \boldsymbol{u}\phi. \tag{2.16b,c}$$

As highlighted by Mooers (1975), this definition of the energy may be interpreted as the sum of a kinetic energy $(1/2)\left|\boldsymbol{u}\right|^2$ and a potential energy $(1/2)\boldsymbol{X}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{X}$ with the desirable property that for wave modes, the kinetic and potential energies satisfy equipartition. We emphasise that in this 2-D framework, the kinetic energy depends only on u and w. Any contributions from v are wrapped up in the potential energy, which in addition to the usual gravitational potential energy also has a horizontal displacement term and a mixed term. Finally, we note that we can write the potential energy in terms of v and b as

$$\frac{1}{2}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{X} \equiv \frac{1}{2}\frac{1}{fQ}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{b}. \tag{2.17}$$

3. Transformation into principal coordinates

3.1. *Diagonalisation*

The Sawyer–Eliassen equation, when written in terms of the matrices $\boldsymbol{A}$ or $\boldsymbol{M}$, has a remarkably simple form. Its complexity is contained in the spatial variation of the matrices of background gradients. However, these matrices are symmetric, thus may be diagonalised by a rotation

$$\boldsymbol{M} = \boldsymbol{R}\boldsymbol{D}\boldsymbol{R}^{\mathrm{T}}, \tag{3.1a}$$

where

$$\boldsymbol{D} = \begin{pmatrix} \omega_{min}^2 & 0 \\ 0 & \omega_{max}^2 \end{pmatrix}, \quad \boldsymbol{R} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \quad \tan 2\theta = \frac{-2\dfrac{\partial B}{\partial x}}{\dfrac{\partial B}{\partial z} - f\left(f + \dfrac{\partial V}{\partial x}\right)}. \tag{3.1b–d}$$

In general, there is no guarantee that the eigenvalues, which we have suggestively called ω_{min}^2 and ω_{max}^2, satisfy $\omega_{min}^2 < \omega_{max}^2$, or indeed are both positive. In particular, background flows unstable to convective, inertial or symmetric instability will have at least one negative eigenvalue in some region of the domain. This is the Hoskins (1974) condition for instability that in some region of the domain the Ertel potential vorticity of the background flow, Q, has the opposite sign to the Coriolis parameter:

$$fQ \equiv \mathrm{Det}(\boldsymbol{M}) \equiv \mathrm{Det}(\boldsymbol{A}) \equiv \mathrm{Det}(\boldsymbol{D}) \equiv \omega_{min}^2\omega_{max}^2 < 0. \tag{3.2}$$

However, in this paper, we are interested in the case $0 < \omega_{min}^2 \ll \omega_{max}^2$, implying $fQ > 0$ everywhere. We now use this diagonalisation to define a coordinate system that elucidates the dynamics of the Sawyer–Eliassen equation.

3.2. *Coordinate definition*

We construct an orthogonal curvilinear coordinate system (ξ, η), the principal coordinates, using the principal axes or eigenvectors of $\boldsymbol{M}$ as a basis. By specifying the unit tangent vectors everywhere, the background flow completely determines the coordinate curves of ξ and η, although we have the freedom to label the coordinate curves however we wish. The covariant basis vectors of the principal coordinate system are given by

$$\boldsymbol{e}_1 := \frac{\partial \boldsymbol{x}}{\partial \xi} = h_1 \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \quad \boldsymbol{e}_2 := \frac{\partial \boldsymbol{x}}{\partial \eta} = h_2 \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}, \tag{3.3a,b}$$

where h_1 and h_2 are the scale factors associated with ξ and η respectively. The introduction of a second coordinate system raises an important notational point. Throughout this paper, we have made – and will continue to make – no distinction between vector or tensor quantities and their representation in the standard Cartesian basis. In many cases, including (3.3), we define quantities through their Cartesian components. Nevertheless, we will be interested in the representation of vector quantities, such as the group velocity, in the principal basis. However, on those occasions, we will be explicit and express the vectors as linear combinations of basis vectors. In particular, we define the normalised basis vectors $\hat{\boldsymbol{x}} := (1, 0)^{\mathrm{T}}$ and $\hat{\boldsymbol{z}} := (0, 1)^{\mathrm{T}}$ for the Cartesian coordinate system, and $\hat{\boldsymbol{\xi}} := \boldsymbol{e}_1/h_1$ and $\hat{\boldsymbol{\eta}} := \boldsymbol{e}_2/h_2$ for the principal coordinate system. We also define the displacements

$$\varXi := \boldsymbol{X} \cdot \hat{\boldsymbol{\xi}}, \quad H := \boldsymbol{X} \cdot \hat{\boldsymbol{\eta}}, \tag{3.4a,b}$$

https://doi.org/10.1017/jfm.2026.11336 Published online by Cambridge University Press

and associated velocities

$$u_\xi := \frac{\partial \varXi}{\partial t} \equiv \boldsymbol{u} \boldsymbol{\cdot} \hat{\boldsymbol{\xi}} \equiv -\frac{1}{h_2}\frac{\partial \psi}{\partial \eta}, \quad u_\eta := \frac{\partial H}{\partial t} \equiv \boldsymbol{u} \boldsymbol{\cdot} \hat{\boldsymbol{\eta}} \equiv \frac{1}{h_1}\frac{\partial \psi}{\partial \xi}, \tag{3.5a,b}$$

in principal coordinates. More details about the principal coordinate system can be found in Appendix A, but we give the major results here. Under this coordinate transformation, (2.12) becomes

$$\frac{\partial^2 \varXi}{\partial t^2} + \omega_{min}^2 \varXi + \frac{1}{h_1}\frac{\partial \phi}{\partial \xi} = 0, \tag{3.6a}$$

$$\frac{\partial^2 H}{\partial t^2} + \omega_{max}^2 H + \frac{1}{h_2}\frac{\partial \phi}{\partial \eta} = 0. \tag{3.6b}$$

Taking a time derivative, using (3.5) and eliminating the pressure we get the Sawyer–Eliassen (2.5*a*) in principal coordinates:

$$\frac{\partial}{\partial \xi}\left[\left(\frac{\partial^2}{\partial t^2} + \omega_{max}^2\right)\frac{h_2}{h_1}\frac{\partial \psi}{\partial \xi}\right] + \frac{\partial}{\partial \eta}\left[\left(\frac{\partial^2}{\partial t^2} + \omega_{min}^2\right)\frac{h_1}{h_2}\frac{\partial \psi}{\partial \eta}\right] = 0. \tag{3.7}$$

Through the coordinate transformation, we have written the Sawyer–Eliassen equation in a form with no mixed derivatives. In the frequency domain, we have

$$\frac{\partial}{\partial \xi}\left[(\omega_{max}^2 - \omega^2)\frac{h_2}{h_1}\frac{\partial \hat{\psi}}{\partial \xi}\right] - \frac{\partial}{\partial \eta}\left[(\omega^2 - \omega_{min}^2)\frac{h_1}{h_2}\frac{\partial \hat{\psi}}{\partial \eta}\right] = 0. \tag{3.8}$$

This form highlights the change in character of the equation from hyperbolic (permitting wave propagation) to elliptic (no wave propagation) and vice versa, depending on the values of $\omega^2 - \omega_{min}^2$ and $\omega_{max}^2 - \omega^2$.

Finally, the conservation of energy becomes

$$\frac{\partial}{\partial t}(h_1 h_2 \mathcal{E}) + \frac{\partial}{\partial \xi}(h_2 \mathcal{F}_\xi) + \frac{\partial}{\partial \eta}(h_1 \mathcal{F}_\eta) = 0, \tag{3.9a}$$

where

$$\mathcal{F}_\xi := \boldsymbol{\mathcal{F}} \boldsymbol{\cdot} \hat{\boldsymbol{\xi}} \equiv u_\xi \phi, \quad \mathcal{F}_\eta := \boldsymbol{\mathcal{F}} \boldsymbol{\cdot} \hat{\boldsymbol{\eta}} \equiv u_\eta \phi. \tag{3.9b,c}$$

3.3. *Strongly stratified limit*

A particularly important limit, due to both its frequent occurrence and practical utility, is the 'strongly' stratified limit, i.e. the limit in which $\partial B/\partial z \gg |\partial B/\partial x|$ and $\partial B/\partial z \gg f(f + \partial V/\partial x)$. Hence $\omega_{max}^2 \approx \partial B/\partial z$. In this case, (3.1*d*) becomes

$$\tan 2\theta \equiv \frac{2\tan\theta}{1 - \tan^2\theta} \approx \frac{-2(\partial B/\partial x)}{\partial B/\partial z} \ll 1 \quad \Longrightarrow \quad \tan\theta \approx -\frac{\partial B/\partial x}{\partial B/\partial z}, \tag{3.10}$$

thus the eigenvectors of $\boldsymbol{M}$ and the (ξ, η) coordinate curves are aligned with and perpendicular to the local isopycnal slope. In this limit, we recover the hydrostatic results of Whitt & Thomas (2013) that the minimum frequency perturbations are aligned with the isopycnal slope, and furthermore, that the local minimum frequency is given by

$$\omega_{min}^2 \approx \frac{fQ}{\partial B/\partial z} \equiv f\left(f + \frac{\partial V}{\partial x}\right) - \frac{(\partial B/\partial x)^2}{\partial B/\partial z} \equiv f^2(1 + Ro_\ell - Ri_\ell^{-1}), \tag{3.11a}$$

where

$$Ro_\ell := \frac{1}{f}\frac{\partial V}{\partial x}, \quad Ri_\ell := \frac{f^2(\partial B/\partial z)}{(\partial B/\partial x)^2} \tag{3.11b,c}$$

are the local Rossby number and balanced Richardson number.

The conditions for the background flow to be 'strongly' stratified are easily met in much of the world's oceans even in situations that an oceanographer might not term strongly stratified. If we consider fronts with $|\partial B/\partial x| \gg f(f + \partial V/\partial x)$ and then impose that the fronts be stable to symmetric instability, $fQ > 0$, then we are automatically in the strongly stratified limit since the stability criterion implies

$$f\left(f + \frac{\partial V}{\partial x}\right)\frac{\partial B}{\partial z} > \frac{\partial B}{\partial x}^2 \implies \frac{\partial B}{\partial z} \gg \left|\frac{\partial B}{\partial x}\right| \gg f\left(f + \frac{\partial V}{\partial x}\right). \tag{3.12}$$

That is not to say that weakly stratified conditions do not exist. Exceptions may be found giving rise to gyroscopic waves (e.g. van Haren & Millot 2005). Despite these exceptions, we only consider the strongly stratified limit in this paper. Furthermore, since in this limit the η coordinate curves and isopycnals are aligned, we will refer to ξ as the along-isopycnal, or just isopycnal, coordinate, and to η as the diapycnal coordinate. Conversely, we reserve the terms horizontal and vertical for the x and z coordinates. However, whenever we refer to a velocity as upwards or downwards, it will be with respect to a particular coordinate system.

Finally, we comment on the commonly used buoyancy or isopycnal coordinates where one takes the standard Cartesian coordinate system and makes a change of vertical coordinate. The resulting coordinate system is not orthogonal and retains the horizontal coordinates from Cartesian coordinates. Consequently, while the principal coordinate system shares one coordinate (η) with buoyancy coordinates, the ξ coordinate curves differ from the x coordinate curves in buoyancy coordinates. At shallow isopycnal slopes, i.e. small values of θ, this difference is often immaterial but can lead to subtle differences. For example, consider figure 1 in which we plot the ξ and η coordinate curves, which are orthogonal, for an example background flow. However, due to the small aspect ratio – the height of the domain is 60, while the width is 1000 – one has to look closely to differentiate the ξ coordinate curves from the x coordinate curves. In this case, we have that $\xi = x$ on $z = 0$, but that at depth, e.g. $z = -60$, $\xi > x$.

4. Total internal reflection

4.1. *Ray-tracing in the principal coordinate system*

To gain some intuition about the dynamics in principal coordinates, we first assume that the background flow varies on a larger scale than the waves whose behaviour may then be described by ray-tracing theory. Given the wave phase Θ, the slowly varying wavenumbers are $k_\xi := \partial\Theta/\partial\xi$ and $k_\eta := \partial\Theta/\partial\eta$. However, we can utilise the slowly varying nature of the scale factors, and define $\check{k}_\xi := h_1^{-1}k_\xi$ and $\check{k}_\eta := h_2^{-1}k_\eta$, which allows us to recover the classic inertia-gravity wave dispersion relation

$$\omega^2 = \omega_{min}^2 \frac{\check{k}_\eta^2}{\check{k}_\xi^2 + \check{k}_\eta^2} + \omega_{max}^2 \frac{\check{k}_\xi^2}{\check{k}_\xi^2 + \check{k}_\eta^2} \tag{4.1a}$$

from (3.8), and write the group velocity as

$$\boldsymbol{c}_g := \frac{\partial\omega}{\partial k_\xi}\boldsymbol{e}_1 + \frac{\partial\omega}{\partial k_\eta}\boldsymbol{e}_2 = h_1\frac{\partial\omega}{\partial k_\xi}\hat{\boldsymbol{\xi}} + h_2\frac{\partial\omega}{\partial k_\eta}\hat{\boldsymbol{\eta}} = \frac{\partial\omega}{\partial \check{k}_\xi}\hat{\boldsymbol{\xi}} + \frac{\partial\omega}{\partial \check{k}_\eta}\hat{\boldsymbol{\eta}}. \tag{4.1b}$$

Moreover, standard ray-tracing results from the barotropic case can be generalised to the baroclinic case provided that they are applied in the rotated coordinate system.

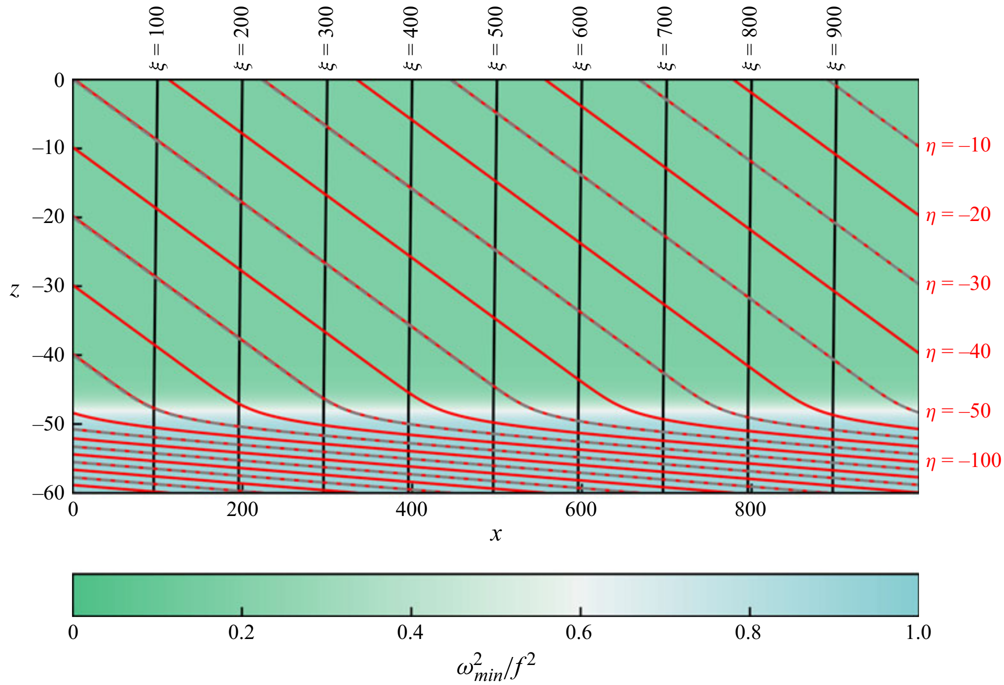

Figure 1. Set-up for the variably stratified frontal zone. Principal coordinate curves are contoured in black (ξ) and red (η). Isopycnals are contoured in dashed grey. In accordance with (3.10), the coordinate curves of η follow isopycnals to very good approximation. Furthermore, although the principal coordinate system is orthogonal, the small aspect ratio of this set-up and submesoscale flows in general mean that the coordinate curves of x and ξ are very similar when plotted.

For example, the oft-used property that upward phase propagation implies downward energy propagation (e.g. Alford *et al.* 2016) is strictly only true for barotropic flows, and in baroclinic cases, it is possible for the vertical phase and group velocities to have the same sign (Mooers 1975; Whitt & Thomas 2013). However, making the coordinate transformation, we see that the more general result is that the diapycnal phase and energy propagation have opposite signs, as we now demonstrate.

4.2. *Variably stratified frontal zone*

Perhaps the simplest example of a baroclinic flow that supports both hyperbolic and elliptic behaviour is the variably stratified frontal zone. In a frontal zone, we consider a background state with constant horizontal buoyancy gradient $\partial B/\partial x = f^2\Gamma$, and impose periodic boundary conditions on the perturbations. Such models are a useful tool for studying a range of processes that occur on scales much shorter than the front width. Here, we allow the stratification to increase with depth, creating an idealisation of the transition between the surface mixed layer and the pycnocline:

$$f^{-2}\frac{\partial B}{\partial z} = \Pi_\infty^2 + \left(\Pi_0^2 - \Pi_\infty^2\right)\sigma(z+d), \tag{4.2a}$$

where

$$\sigma(s) = \frac{1}{1+\mathrm{e}^{-s}} \tag{4.2b}$$

https://doi.org/10.1017/jfm.2026.11336 Published online by Cambridge University Press

is the logistic function. We take $\Gamma = 9$, $d = 50$, $\Pi_0 = 10$ and $\Pi_\infty = 30$ such that the upper region, $\partial B/\partial z = f^2\Pi_0^2$, is stable ($\omega_{min}^2 = 0.19 f^2$) and the pycnocline has minimum frequency $\omega_{min}^2 = 0.9 f^2$.

As the background flow is invariant with x, it is straightforward to compute the scale factors analytically. Choosing $\partial\xi/\partial x = \partial\eta/\partial z = 1$ on $z = 0$, the Jacobian matrix of the (inverse) coordinate transformation is

$$\begin{pmatrix} \dfrac{\partial\xi}{\partial x} & \dfrac{\partial\xi}{\partial z} \\ \dfrac{\partial\eta}{\partial x} & \dfrac{\partial\eta}{\partial z} \end{pmatrix} = \begin{pmatrix} 1 & \tan\theta(z) \\ -\tan\theta_0 & \tan\theta_0\cot\theta(z) \end{pmatrix}, \tag{4.3}$$

where $\theta_0 := \theta|_{z=0}$. We label the coordinates along $z = 0$ as $\xi = x$ and $\eta = -\tan\theta_0(x - L_x)$. Figure 1 demonstrates the set-up, coordinate curves and the approximate equivalence between isopycnals and the η-coordinate curves.

4.3. *Anomalously low frequency regime*

We initialise a wavepacket of the form

$$\psi = A\cos(k(x + \alpha^{-1}z) - \omega t), \qquad \alpha = \frac{\Gamma \pm \sqrt{\Gamma^2 - \left(1 - f^{-2}\omega^2\right)\left(\Pi_0^2 - f^{-2}\omega^2\right)}}{\left(\Pi_0^2 - f^{-2}\omega^2\right)} \tag{4.4a,b}$$

towards the bottom of the weakly stratified upper layer. Here, the amplitude A is a slowly varying Gaussian in z (figure 2*b*). As the background flow is invariant of x, k will be constant throughout the evolution of the wavepacket. According to ray-tracing theory, the wavepacket will propagate with group velocity

$$\boldsymbol{c_g} = \frac{\partial\omega}{\partial\alpha}\frac{\alpha}{k}\left(\hat{\boldsymbol{x}} - \alpha\hat{\boldsymbol{z}}\right) = \frac{\pm\alpha\sqrt{\Gamma^2 - \left(1 - f^{-2}\omega^2\right)\left(\Pi_0^2 - f^{-2}\omega^2\right)}}{k\omega\left(1 + \alpha^2\right)}\left(\hat{\boldsymbol{x}} - \alpha\hat{\boldsymbol{z}}\right). \tag{4.5}$$

The simulation is run for 15 inertial periods using an implicit pseudo-spectral solver for the Sawyer–Eliassen equation (*SawyerEliassenSolver.jl,* Hilditch 2025).

We choose α such that $\omega^2 = 0.6 f^2$, placing us in the 'anomalously low frequency' regime (Mooers 1975; Whitt & Thomas 2013), so named because the frequency is less than the effective Coriolis frequency

$$f_{eff} = \sqrt{f(f + \partial V/\partial x)}, \tag{4.6}$$

which in this case is simply equal to f. This is the frequency at which horizontal perturbations oscillate, and consequently, in the anomalously low frequency regime, both characteristics have the same signed slope, i.e. $\alpha > 0$, with respect to the Cartesian coordinate system. In our initial conditions, we take the negative square root such that the wavepacket propagates on the shallower of the two characteristics. On this characteristic, we have the unusual behaviour that the vertical group and phase velocities take the same sign – here negative as $k < 0$ (see figure 2(*b*) and supplementary movie 1 is available at https://doi.org/10.1017/jfm.2026.11336).

However, when viewed in principal coordinates, this regime displays more familar behaviour. The group velocity is

$$\boldsymbol{c}_g = \frac{\partial\omega}{\partial\alpha}\frac{\alpha}{k}\left(\hat{\boldsymbol{x}} - \alpha\hat{\boldsymbol{z}}\right) \equiv \frac{\partial\omega}{\partial\alpha}\frac{\alpha}{k}\cos\theta\left[\,(1 - \alpha\tan\theta)\,\hat{\boldsymbol{\xi}} - (\tan\theta + \alpha)\,\hat{\boldsymbol{\eta}}\right]. \tag{4.7}$$

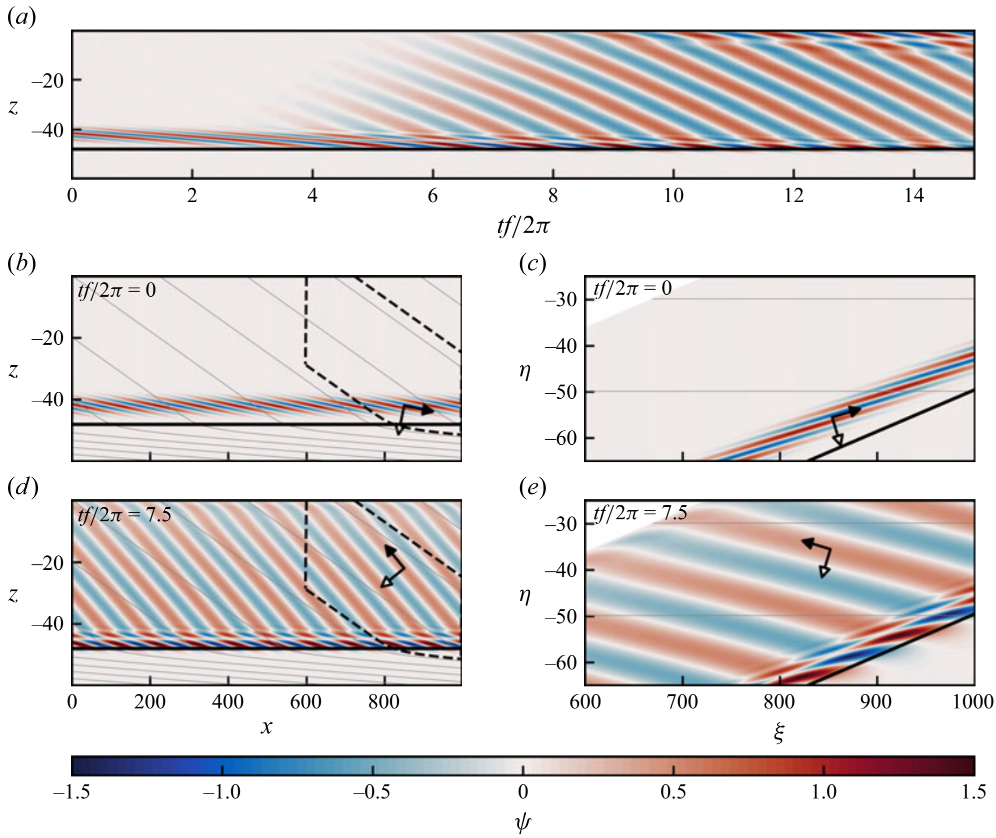

Figure 2. Evolution of a wavepacket of frequency $\omega_{min}^2 = 0.6f^2$ in the variably stratified frontal zone. (*a*) Hovmöller plot of the streamfunction at $x = 0$. The thick black contour is the line $\omega_{min}^2 = 0.6f^2$. (*b*,*d*) Snapshots of the streamfunction in the Cartesian coordinate system at the start and middle of the simulation. Background buoyancy is contoured in grey. (*c*,*e*) Same as (*b*,*d*), but a subset of the domain, indicated by dashed black lines in (*b*,*d*), plotted in the principal coordinate system. Closed black arrows indicate the direction, but not magnitude, of the group velocity (4.5), and open arrows indicate the direction of phase propagation for the downward (*b*,*c*) and upward (*d*,*e*) propagating wavepackets. Supplementary movie 1 shows the evolution of u and w in the two coordinate systems.

We highlight the change in behaviour between the coordinate systems. In the anomalously low frequency regime, the shallow characteristic is shallower than the η coordinate curves, $\alpha + \tan\theta < 0$, thus the group velocity, while downwards in Cartesian coordinates, is upwards in the principal coordinate system (figure 2*c*), i.e. the wavepacket is moving into more buoyant water.

The direction of the group velocity is reflected in the phase-averaged energy flux, here defined via an x-average. The horizontal $\overline{\mathcal{F}_x}$ and along-isopycnal $\overline{\mathcal{F}_\xi}$ energy fluxes display the same behaviour – positive on the shallow characteristic, and negative on the steep (figures 3*b*,*d*). However, while the vertical energy flux $\overline{\mathcal{F}_z}$ is negative on the shallow characteristic and positive on the steep characteristic (figure 3*c*), the diapycnal energy flux $\overline{\mathcal{F}_\eta}$ is positive throughout (figure 3*e*).

4.4. *Turning point reflection*

When the wavepacket reaches the separatrix, $\omega_{min}^2(z) = \omega^2$, it exhibits turning point behaviour and reflects back along the steep characteristic with a rapidly decaying

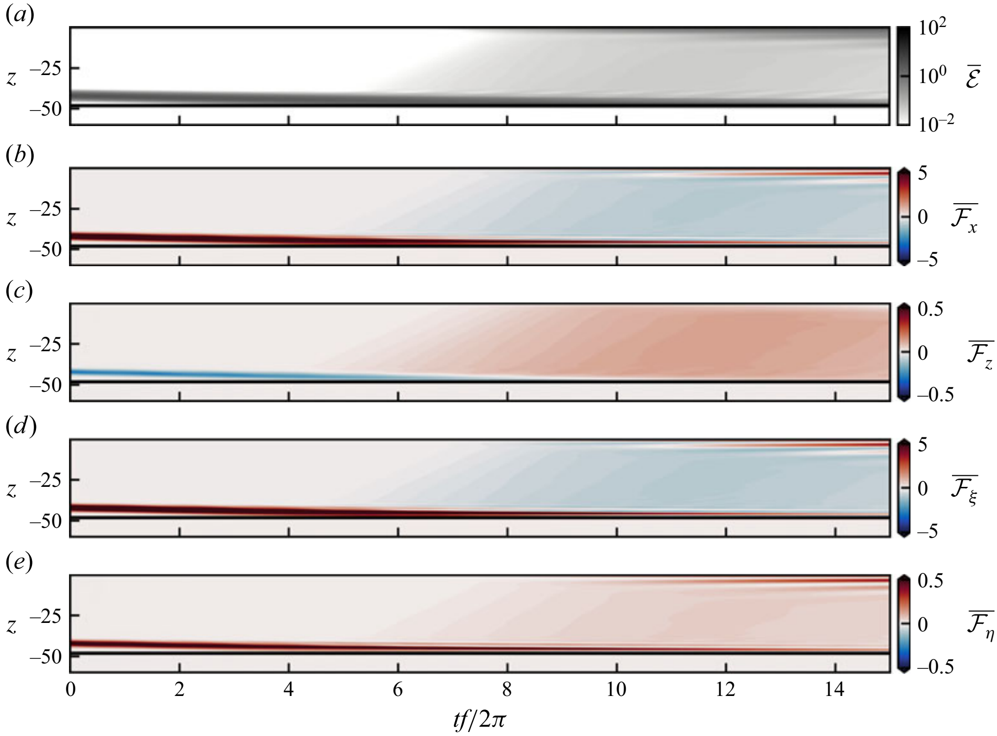

Figure 3. Hovmöller plots of (*a*) the energy and (*b–e*) energy fluxes for the variably stratified frontal zone simulation. Here, phase-averaging, denoted by overlines, is achieved via an x-average. The thick black contour is the line $\omega_{min}^2 = 0.6 f^2$.

evanescent tail extending beneath the separatrix (figure 2). On the steep characteristic, the reflected wavepacket has a much longer vertical wavelength, which not only results in much more rapid propagation, but also has consequences for the magnitude of the horizontal velocity u. While the amplitudes of w and hence the streamfunction are approximately equal between the upward and downward propagating wavepackets, u is much weaker on the steep characteristic (see supplementary movie 1).

In the principal coordinate system, the phase lines approach the isopycnals, indicating minimum-frequency waves, as the wavepacket approaches the separatrix. Under ray-tracing, a wavepacket can approach the separatrix in two vastly different ways. First, when the separatrix does not align with the isopycnals, as is the case here, the waves reach the separatrix in finite time, with a finite diapycnal wavenumber, and undergo a total internal reflection (Whitt & Thomas 2013). Solutions in the vicinity of the separatrix may be found using WKBJ turning point theory. A crucial aspect of total internal reflection is that the along-isopycnal wavenumber becomes zero before switching sign as the wavepacket propagates away on the other characteristic. The isopycnal (diapycnal) group velocity is the familiar odd (even) function of the isopycnal wavenumber, i.e. differentiating the dispersion relation (4.1*a*),

$$\frac{\partial \omega}{\partial \check{k}_\xi} = \frac{\check{k}_\xi}{\omega}\frac{\omega_{max}^2 - \omega^2}{\check{k}_\xi^2 + \check{k}_\eta^2}, \qquad \frac{\partial \omega}{\partial \check{k}_\eta} = -\frac{\check{k}_\eta}{\omega}\frac{\omega^2 - \omega_{min}^2}{\check{k}_\xi^2 + \check{k}_\eta^2}. \tag{4.8a,b}$$

Therefore, when undergoing a total internal reflection, it is the isopycnal group velocity that changes sign and not the diapycnal group velocity. Again, this is reflected in the energy flux, which displayed a change in sign in the isopycnal component (figure 3*d*) but not the diapycnal component (figure 3*e*).

It should be noted that this is true only for total internal reflections and not reflections off a domain boundary such as the ocean surface or partial reflections off sharp changes in the background flow. For domain boundaries, the choice of characteristic and corresponding change in group velocity depend on whether the reflection is sub-critical, critical or super-critical. The behaviour of NIWs at the ocean surface in the presence of fronts was detailed by Grisouard & Thomas (2015). In this simulation, the reflection off the surface is super-critical, or 'backward' in the language of Grisouard & Thomas (2015), and the wavepacket continues propagating upwards in the principal coordinate system. An interesting quirk or weakness of the frontal zone set-up is that the background buoyancy is not continuous. This has two related consequences. First, the wavepacket may continue propagating upwards across diapycnals indefinitely thanks to the periodic boundary conditions that allow the wavepacket to leave the domain at $x = L_x$ and instantly reappear at $x = 0$, where the background buoyancy is less (or vice versa). Second, the isopycnals are always sloped (in the Cartesian coordinate system) and never align with the separatrix. As a result the wavepacket is never truly trapped, and the second way of approaching the separatrix, namely in a critical layer, is not realised.

5. Trapped isopycnal modes and critical layers

5.1. *Along-isopycnal modes*

In order to discuss wave trapping and critical layers, is it helpful to loosen the constraints of ray-tracing. Implicit in the discussions in § 4 and the results of Whitt & Thomas (2013) is the assumption that the background flow is varying slowly such that ray-tracing methods may be applied, and concepts such as the group velocity are clearly defined. In reality, this assumption, at least in the horizontal, is a poor one. The issue is particularly acute in the submesoscale regime, but is also relevant at mesoscales where the spatial extent of the waves may still exceed the eddy length scales. One of the most notable properties of NIWs, and inertia-gravity waves in general, is that under the ray-tracing assumption, their frequency is determined solely by their aspect ratio and not their wavelength. Therefore, to achieve the minimum frequency, it is equally valid to make the diapycnal wavelength very short or the along-isopycnal wavelength very long. It is therefore beneficial to relax the WKBJ assumption in the along-isopycnal direction. However, we will continue to assume that the diapycnal wavelength is short such that we may retain the ray-tracing approximation in η.

Slowly varying waveguide theory (Bretherton 1968; Hayes 1970) describes the propagation of a wavepacket in η. We recall the key results here. After making the WKBJ approximation in η, the Sawyer–Eliassen equation becomes a second-order ordinary differential equation (ODE) in ξ with parametric dependence on ω, η and the diapycnal wavenumber k_η:

$$\frac{\partial}{\partial \xi}\left[\left(\omega_{max}^2-\omega^2\right)\frac{h_2}{h_1}\frac{\partial \hat{\psi}}{\partial \xi}\right]-k_\eta^2\left[\left(\omega_{min}^2-\omega^2\right)\frac{h_1}{h_2}\right]\hat{\psi}=0, \tag{5.1}$$

where $\hat{\psi}(\xi;\eta,t)$ is the mode structure,

$$\psi(\xi,\eta,t)=\hat{\psi}(\xi;\eta,t)\,\mathrm{e}^{\mathrm{i}\,\Theta(\eta,t)}+\text{c.c.} \tag{5.2}$$

This ODE, subject to suitable boundary conditions, defines an eigenvalue problem admitting a discrete spectrum of frequencies ω_n and associated eigenfunctions. That the along-isopycnal modes are discrete is an important part of the slowly varying waveguide theory as it prohibits energy transfer between different along-isopycnal modes as the wavepacket evolves (Bretherton 1968). Crucially, for each mode, this eigenvalue problem defines a dispersion relation and group velocity

$$\omega_n = \omega_n(k_\eta, \eta), \quad c_{g,n} := \frac{\partial \omega_n}{\partial k_\eta}. \tag{5.3a,b}$$

Not only does the change of coordinate system allow us to make this partial WKBJ approximation, but the resulting eigenvalue problem is very similar to the Klein–Gordon eigenvalue problem that arises in the study of barotropic background flows (Balmforth *et al.* 1998; Hilditch, Taylor & Thomas 2025). In particular, if we focus on the low frequency waves $\omega^2 \ll \omega^2_{max}$, then (5.1) is in Sturm–Liouville form. There are subtle differences between the barotropic problem and (5.1), such as the presence of the non-trivial weight function h_1/h_2, but qualitatively the key dynamics are the same.

Notably, we can use scaling arguments to identify an important diapycnal wavenumber $k^\star_\eta$. Suppose that $1/\beta_\xi$ is the along-isopycnal scale of the background flow. Then we define $k^\star_\eta$ by assuming that the waves have the same along-isopycnal scale, i.e. $\partial/\partial\xi \sim \beta_\xi$:

$$k^{\star 2}_\eta := \beta_\xi^2 \left[\!\left[\frac{h_2}{h_1}\right]\!\right]^2 \frac{[\![(\omega^2_{max} - \omega^2)]\!]}{[\![(\omega^2 - \omega^2_{min})]\!]}, \tag{5.4}$$

where $[\![\cdot]\!]$ denotes a scaling. For most of these terms, simply taking the (potentially weighted) mean is sufficient to construct a scaling. In particular, since we are interested in $\omega^2_{max} \gg \omega^2$, we may define $[\![(\omega^2_{max} - \omega^2)]\!]$ to be the mean of ω^2_{max}. The only complicated term is $[\![(\omega^2 - \omega^2_{min})]\!]$, since for the low frequency waves, $\omega^2 \sim \omega^2_{min}$. Rather than the mean, the standard deviation of ω^2_{min} provides a better scaling. Thus letting $\langle\cdot\rangle_\xi$ denote an along-isopycnal average, we have

$$k^{\star 2}_\eta := \beta_\xi^2 \left\langle \frac{h_2}{h_1} \right\rangle_\xi^2 \langle \omega^2_{max} \rangle_\xi \left\langle \left(\omega^2_{min} - \langle \omega^2_{min} \rangle_\xi\right)^2 \right\rangle_\xi^{-1/2}. \tag{5.5}$$

This scaling for the diapycnal wavenumber is very useful for heuristically understanding the behaviour of the along-isopycnal modes through analogy with the barotropic problem studied by Balmforth *et al.* (1998) and Hilditch *et al.* (2025). The eigenvalue problem (5.1) can be thought of as a competition between the derivatives in the first term, which act to smooth the eigenfunction, and the variations in ω^2_{min}, which act to generate spatial variability through refraction. For $k_\eta \ll k^\star_\eta$, the first term dominates and the solutions are smooth. Conversely, for $k_\eta \gg k^\star_\eta$, refraction dominates and the solutions display smaller-scale variations.

Finally, slowly varying waveguide theory gives a conservation law for wave action

$$\frac{\partial}{\partial t}\mathcal{A}_n + \frac{\partial}{\partial \eta}(c_{g,n}\mathcal{A}_n) = 0, \tag{5.6a}$$

where

$$\mathcal{A}_n = \int \frac{\overline{\mathcal{E}}_n}{\omega_n} h_1 h_2 \,\mathrm{d}\xi \tag{5.6b}$$

is the wave action defined in terms of the phase-averaged energy density of the nth mode $\overline{\mathcal{E}}_n$. Given the steady background flow, we may multiply $\mathcal{A}_n$ in (5.6a) by ω_n to achieve a conservation law for the phase-averaged mode energy.

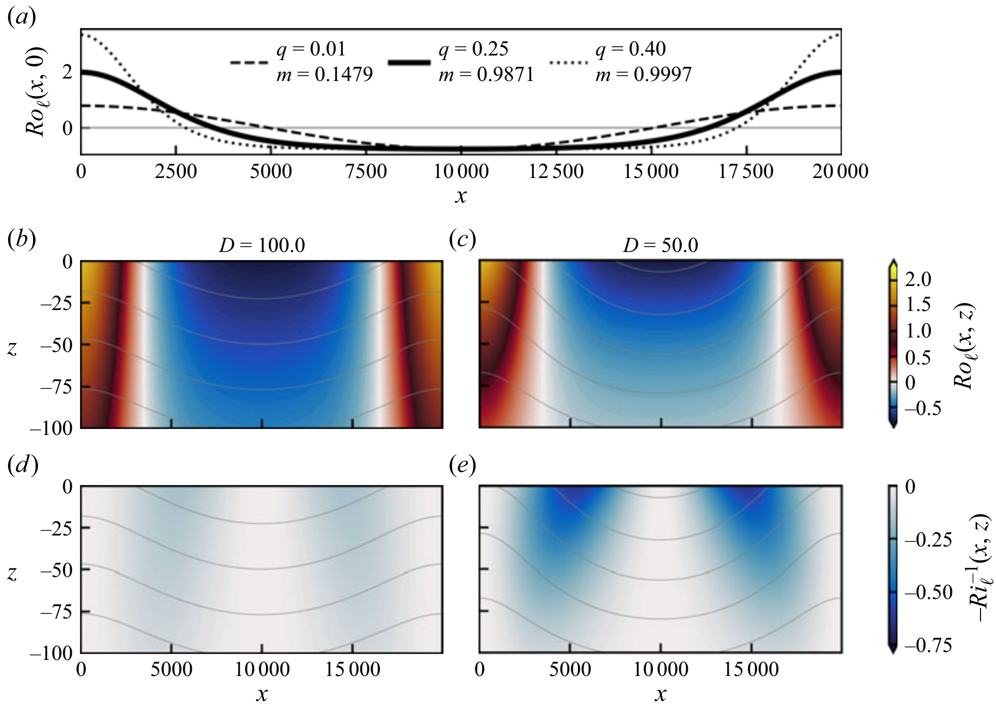

Figure 4. (*a*) Local Rossby number at the surface for $Ro = 0.75$ and three values of q. Local Rossby number (*b*,*c*) and minus the inverse local Richardson number (*d*,*e*) for $D = 60$ and $D = 30$ in the top part of the domain.

5.2. *Numerical simulations*

5.2.1. *Background flow*

To illustrate these concepts, we utilise numerical simulations (see supplementary movie 2 for the evolution of the along-front vorticity). We construct a periodic background flow that decays with depth and displays the skewness in vorticity typical of submesoscale flows (Shcherbina *et al.* 2013). The background flow consists of a filament, i.e. a double-sided front, with a cyclonic region and a broader, weaker anticyclonic region (figure 4). We now give the mathematical details of the background flow construction.

A convenient set of functions for constructing a flow with these properties comprises the Jacobi elliptic functions. These functions are typically given in terms of theta functions, for which we follow the notation of Whittaker & Watson (2000). The theta functions are defined in terms of a parameter q called the nome, which we take to be real with $0 < q < 1$. Choosing a value of q fixes the squared elliptic modulus $m = \theta_2(0, q)^4/\theta_3(0, q)^4$ and the complete elliptic integrals of the first and second kind, $K(m)$ and $E(m)$, respectively. The Jacobi elliptic functions have period $2K(m)$, thus we define $x' := 2K(m)\, x/L_x$, where L_x is the width of the domain. We also define $z' := z/D$, where D is a vertical scale. The lateral shear is given by

$$\frac{1}{f}\frac{\partial V}{\partial x} = 2Ro\left(m\,\Lambda(m)\,\mathrm{cn}^2(x', m) - 1\right)\sigma(z'), \tag{5.7a}$$

where cn is the elliptic cosine, and $\Lambda(m) := K(m)/[E(m) - (1 - m)\,K(m)]$. Oceanographers may recognise the x' dependence of (5.7*a*) as the functional form of the surface elevation of a cnoidal surface gravity wave (Korteweg & de Vries 1895).

D	125	100	75	50
max Ri_ℓ^{-1}	0.10	0.16	0.28	0.62

Table 1. Simulation parameters.

The role of the nome q or equivalently the squared elliptic modulus m is to control the skewness of the lateral shear. As $q \to 0$, the Jacobi elliptic functions tend to sinusoids, and the surface value of $f^{-1}\,\partial V/\partial x$ tends to $Ro\cos(2\pi x/L_x)$. Increasing q makes the cyclonic region sharper and stronger, with the anticyclonic region becoming broader (figure 4*a*). In this paper, we use $q = 0.25$ and $Ro = 0.75$ such that the lateral shear ranges from $-0.75f$ to $2f$.

The horizontal buoyancy gradient is given by

$$\frac{1}{f^2}\frac{\partial B}{\partial x} = 2\,Ro\,\frac{L_x}{D}\,\frac{\Lambda(m)}{2K(m)}\,\mathrm{zn}(x', m)\,\sigma(z')\,(1-\sigma(z')), \tag{5.7b}$$

where zn is the Jacobi zeta function, and the vertical buoyancy gradient is

$$\frac{1}{f^2}\frac{\partial B}{\partial z} = \Pi^2 + 2\,Ro\,\frac{L_x^2}{D^2}\,\frac{\Lambda(m)}{4K(m)^2}\,\log\Theta(x', m)\,\sigma(z')\,(1-\sigma(z'))(1-2\sigma(z')), \tag{5.7c}$$

where $\Theta(x', m) \equiv \theta_4(\pi x/L_x, q)$ is the Jacobi theta function. The other key parameters defining the background flow are the aspect ratio L_x/D and the strength of the background stratification Π. In addition to fixing $Ro = 0.75$, we also fix $\Pi = 60$, $L_x = 20\,000$ and the domain depth $L_z = 400$. These lengths may be interpreted as having units of metres, in which case the background flow would be representative of an oceanic submesoscale flow. However, the Sawyer–Eliassen equation, being of homogeneous spatial degree, does not require such an interpretation. We vary D, the vertical scale, in order to adjust the strength of the baroclinic effects. A useful metric for the strength of these effects is the maximum of the inverse local Richardson number, which varies from 0.1 to 0.62 (see table 1). The local Rossby number and inverse Richardson number for two of the simulations are shown in figure 4.

Finally, we must also decide how to label the ξ and η coordinate curves. We choose $\xi = x$ on $z = 0$, and $\eta = z$ on $x = 0$. With these choices, the scale factors h_1 and h_2 are approximately equal to 1.

5.2.2. *Initial conditions*

Our initial conditions are a thin slab of across-front velocity,

$$u(x, z)|_{t=0} = \sigma\left(\frac{z + d_s}{\lambda_s}\right), \tag{5.8}$$

with $d_s = 10$, $\lambda_s = 2$, and all other variables set to zero. However, these initial conditions introduce a small complication with the boundary conditions. When we numerically solve the Sawyer–Eliassen equation, we impose that ψ is zero on the top and bottom boundaries. In addition to imposing a rigid lid ($w = 0$) boundary condition, this also imposes the additional constraint that the domain average across-front velocity must vanish. Our initial conditions do not have zero mean, so we must subtract off this part and solve for its dynamics separately. Fortunately, these dynamics are very simple. Taking the domain average of (2.12) and noting that $Z = 0$ gives

$$\frac{\partial^2 X}{\partial t^2} + f^2 X = 0, \tag{5.9}$$

thus the dynamics of the domain average component are simply inertial oscillations $X = f^{-1} u_0 \sin ft$, $u = u_0 \cos ft$, $Z = w = 0$, where u_0 is the domain average of the initial condition. This induces v, b and ϕ perturbations that inherit their spatial structure from the background flow: $v = -(f + \partial V/\partial x) u_0 \sin ft$, $b = -f^{-1}(\partial B/\partial x) u_0 \sin ft$ and $\phi = -V u_0 \sin ft$.

5.3. *Critical layer location*

The parametric dependence of (5.1), and hence (5.3), on η introduces the possibility of wave trapping. Typically, ocean flows are surface intensified and decay with depth. While sub-inertial waves may be supported at the surface, at depth where the background flow is negligible, the minimum frequency is f. Therefore, there must exist a critical level beyond which the sub-inertial waves cannot pass. We note that there must be a critical level and not a turning point style reflection as in § 4, as a turning point reflection does not involve a change in sign of the diapycnal group velocity, so a wave propagating downwards in buoyancy space will continue to propagate into less buoyant water after undergoing a turning point reflection.

The key question for determining the location of the critical level is: for a given η level, what range of frequencies will the background flow support? To answer this question, we have the lower bound

$$\omega^2 > \Omega^2(\eta) := \min_{\xi} \omega^2_{min}(\xi, \eta). \tag{5.10}$$

This lower bound can be seen by considering $\int \hat{\psi}^*$ (5.1) dξ, where * denotes complex conjugation, and integrating by parts:

$$k_\eta^{-2} \int \left(\omega^2_{max} - \omega^2\right) \frac{h_2}{h_1} \left| \frac{\partial \hat{\psi}}{\partial \xi} \right|^2 \mathrm{d}\xi = \int \left(\omega^2 - \omega^2_{min}\right) \frac{h_1}{h_2} \left| \hat{\psi} \right|^2 \mathrm{d}\xi. \tag{5.11}$$

If we were to have $\omega^2 < \min_\xi \omega^2_{min}$, then the above equation cannot be satisfied as the left-hand side would be positive, while the right-hand side would be negative. Furthermore, we can see that the solution must be weighted towards the region where $\omega^2_{min} < \omega^2$ to ensure that the right-hand side remains positive. In general, for fixed mode number n, as $k_\eta \to \infty$, $\omega^2_n \to \Omega^2(\eta)$, and the modes become increasingly localised in ξ.

We observe the accumulation of energy into the minima of ω^2_{min} in our numerical simulations (figure 5). When D is large (figure 5*a*,*b*), the energy accumulates in the centre of the domain where the lateral shear $\partial V/\partial x$ is minimal. This behaviour is familiar from the mesoscale regime where the accumulation of wave energy at the base of anticyclones is something that has been observed in both models and observations (e.g. Joyce *et al.* 2013). However, as we shrink the vertical scale of the background flow and increase the horizontal buoyancy gradients, two things change. First, the isopycnals and hence η coordinate curves become more curved, and second, we see a greater baroclinic contribution to the local minimum frequency. As the local Richardson number decreases, the structure of the local minimum frequency contours (in the principal coordinate system) change from a single well to a double well structure (figures 5*b*,*d*,*f*,*h*). The double well structure forms because the baroclinic contribution to the local minimum frequency is concentrated not at the centre of the anticyclonic region, but on the edges. In particular, if one constructs

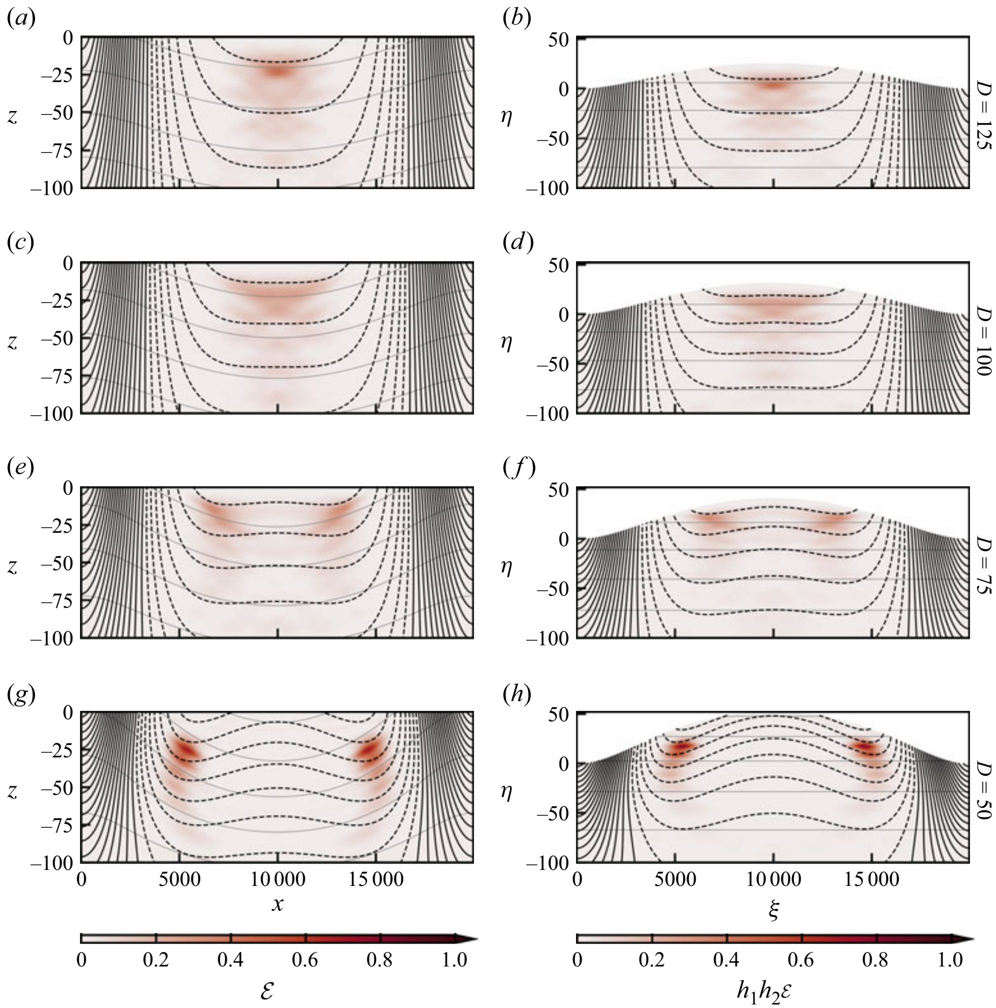

Figure 5. Snapshot at $tf/2\pi = 20$ of the energy density in Cartesian and principal coordinates for four different values of D. Isopycnals are contoured in light grey, and the local minimum frequency squared, ω_{min}^2, is contoured in black with spacing $0.1\,f^2$. Dashed contours indicate sub-inertial values.

a separable background flow, i.e. $V(x, z) = V_x(x)\, V_z(z)$, as we did for (5.7), then the baroclinic effects will be strongest at the extrema of $V_x(x)$ where the lateral shear is zero. In general, for strongly baroclinic flows, we might expect to find the critical layers forming in fronts and at the edges of dense filaments rather than where the flow is most anticyclonic.

5.4. *Localised modes*

To further illustrate this localisation of the wave energy, we solve for the lowest eigenvalues ω_0^2 and eigenmodes $\hat{\psi}_0$ of (5.1) along $\eta = 15$, the approximate location where the wave energy accumulates in the simulations (figures 5*b*,*d*,*f*,*h*). The $\eta = 15$ curve intersects the surface, thus we have the boundary conditions $\hat{\psi} = 0$ at $\xi = \xi_a$ and $\xi = \xi_b$, where ξ_a and ξ_b are the values of ξ at which $\eta = 15$ intersects the surface. We solve the eigenvalue problem as a function of k_η (figures 6*a*,*c*,*e*,*g*) using a Prüfer shooting method (Pryce 1993). We also mark the value of $k_\eta^\star$, as defined by (5.5) for each value of D, taking $\beta_\xi := \pi/(\xi_b - \xi_a)$,

https://doi.org/10.1017/jfm.2026.11336 Published online by Cambridge University Press

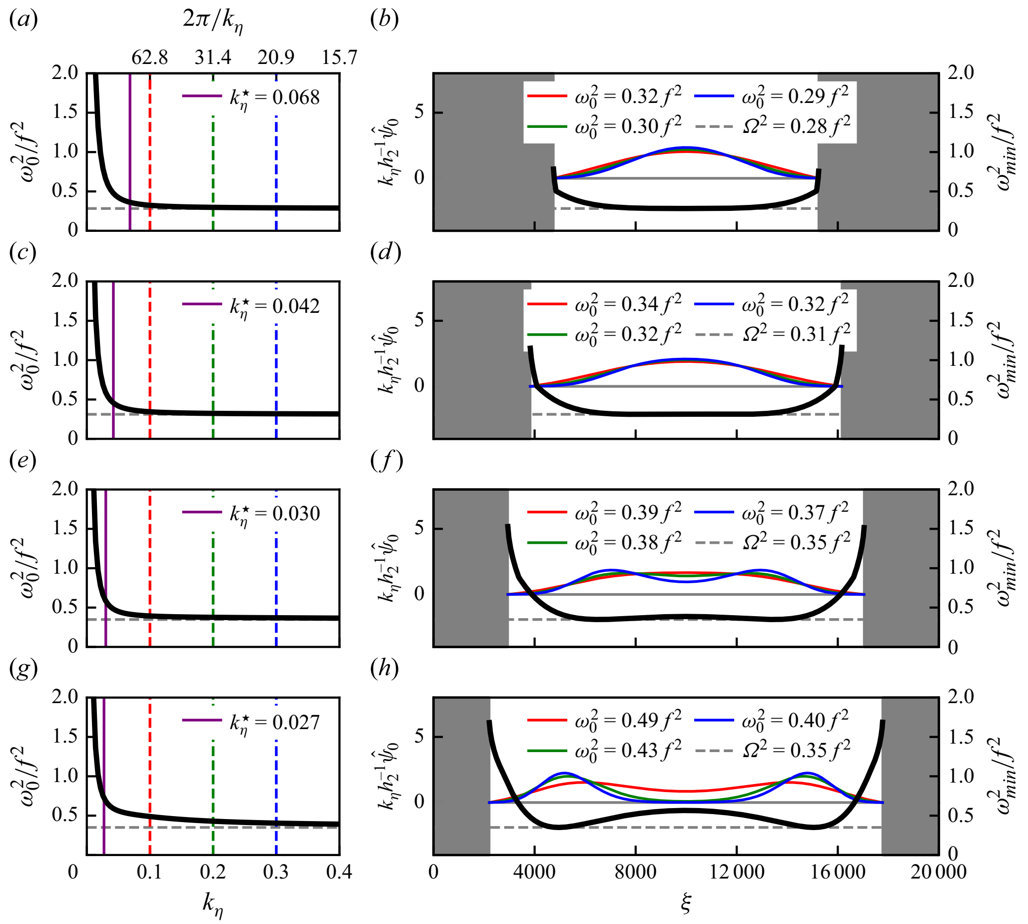

Figure 6. Lowest (*a*,*c*,*e*,*g*) eigenvalues and (*b*,*d*,*f*,*h*) eigenmodes of (5.1) along $\eta = 15$ for (*a*,*b*) $D = 125$, (*c*,*d*) $D = 100$, (*e*,*f*) $D = 75$ and (*g*,*h*) $D = 50$. The eigenmodes for $k_\eta = 0.1$ (red), 0.2 (green) and 0.3 (blue) are plotted as $k_\eta\, h_2^{-1}(\xi)\, \hat{\psi}_0(\xi)$, which is the amplitude of the along-isopycnal component of the velocity (3.5*a*). The eigenmodes are normalised such that $(1/L_x)\int_{\xi_a}^{\xi_b} \overline{\mathcal{E}}_0 h_1 h_2 \,\mathrm{d}\xi = 1$. Purple lines in (*a*,*c*,*e*,*g*) mark $k_\eta^\star$ as defined by (5.5). The black lines in (*b*,*d*,*f*,*h*) are $\omega_{min}^2(\xi)$, and in both the eigenvalue and eigenmode plots, the dashed horizontal grey line indicates the theoretical minimum value (5.10) of ω_0^2, $\Omega^2(\eta = 15)$.

i.e. a sinusoidal scaling. We find that this rough scaling does a good job of predicting the wavenumber at which $\omega_0^2(k_\eta)$ changes behaviour (figures 6*a*,*c*,*e*,*g*). We also plot the lowest eigenmodes for $k_\eta = 0.1$, 0.2, 0.3 (figures 6*b*,*d*,*f*,*h*).

While the shapes of the $\omega_0^2(k_\eta)$ curves are remarkably similar for all four values of D (figures 6*a*,*c*,*e*,*g*), the structure of the eigenmodes varies significantly both as a function of D but also as a function of k_η. The salient variation with D is due to aforementioned change in structure of $\omega_{min}^2(\xi)$ (black lines in figures 6*b*,*d*,*f*,*h*) from a single well to a double well. The eigenmodes reflect this change and similarly transition from a single peak to a double peak structure. The variation with k_η reflects the competition between the smoothing effect of the first term in (5.1) and the refractive effect of the second term. For larger k_η, refraction becomes more dominant, and the eigenmodes become more localised. However, perhaps the most notable aspect of this eigenmode analysis is how sensitive the eigenmodes are to the parameter values. We are varying both D and k_η by at most a factor

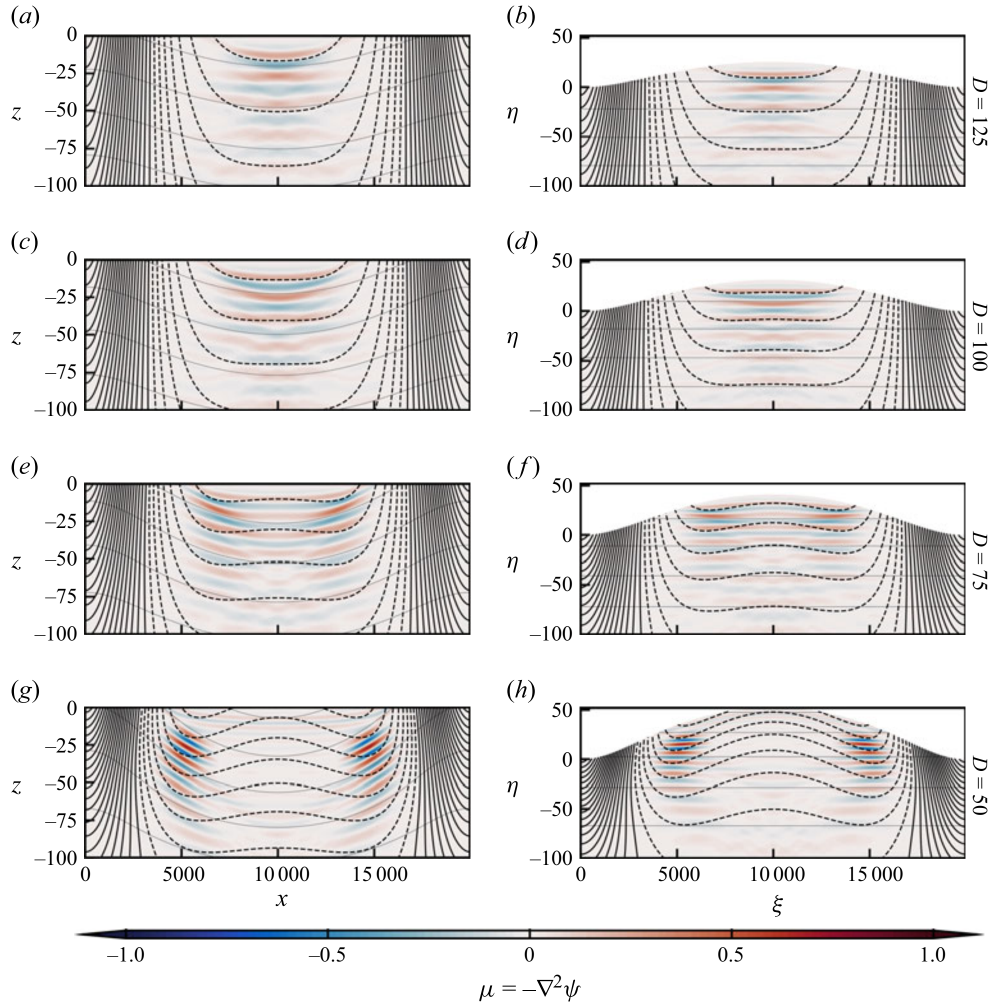

Figure 7. Snapshot at $tf/2\pi = 20$ of the along-front vorticity μ in Cartesian and principal coordinates for four different values of D. Isopycnals are contoured in light grey, and the local minimum frequency ω^2_{min} is contoured in black with spacing $0.1\,f^2$. Dashed contours indicate sub-inertial values.

of 3, but this is sufficient to make significant, qualitative changes to both the structure of the background flow (in the case of D) and the eigenmodes.

5.5. *Secular growth in along-front vorticity*

To observe the along-isopycnal structure of the waves as they approach the critical level, we look at the along-front vorticity

$$\mu := -\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2}\right)\psi \equiv -\frac{1}{h_1 h_2}\left(\frac{\partial}{\partial \xi}\frac{h_2}{h_1}\frac{\partial}{\partial \xi} + \frac{\partial}{\partial \eta}\frac{h_1}{h_2}\frac{\partial}{\partial \eta}\right)\psi. \tag{5.12a}$$

The phase lines of vorticity are aligned with isopycnals and are thus increasingly slanted in physical space as D decreases. However, when plotted in principal coordinates the vorticity forms flat bands (figure 7, supplementary movie 2). These bands of vorticity are spread over a wide range of η values, as indeed to a slightly lesser extent is the energy

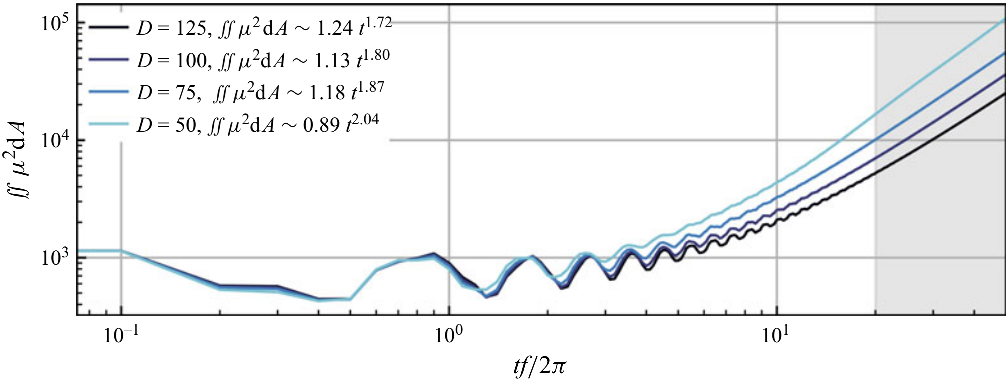

Figure 8. Evolution of the domain integrated along-front enstrophy μ^2 over time for four values of D. Power-law fits over $tf/2\pi > 20$, shaded in grey, are indicated in the legend.

(figure 5). This is likely a consequence of our localised initial conditions. They project onto a range of frequencies, hence the solution does not accumulate at a single critical level.

For NIWs, in particular those approaching a critical level, the vorticity is dominated by the final term of (5.12a) such that

$$\mu \approx -\frac{1}{h_1 h_2}\frac{\partial}{\partial \eta}\frac{h_1}{h_2}\frac{\partial}{\partial \eta}\psi = \frac{1}{h_1 h_2}\frac{\partial (h_1 u_\xi)}{\partial \eta} \approx \frac{\partial u}{\partial z}, \tag{5.12b}$$

where the final approximation follows from the strongly stratified assumption and the smallness of θ. The phase-averaged enstrophy satisfies

$$\overline{\mu^2} \approx \frac{k_\eta^2}{h_2^2}\overline{u_\xi^2} \approx \frac{k_\eta^2}{h_2^2}\overline{\mathcal{E}}, \tag{5.13}$$

where we have used the equipartition of energy to say that $\overline{\mathcal{E}} = \overline{|\boldsymbol{u}|^2} \approx \overline{u_\xi^2}$. Since the domain integrated energy is conserved, the domain integrated enstrophy will scale like k_η^2/h_2^2.

In the numerical simulations, we observe that the domain integrated enstrophy grows faster as the vertical scale is decreased (figure 8). After 10 inertial periods, the domain integrated enstrophy has increased by a factor of 5 for $D = 50$ compared to a factor of 2 for $D = 125$. Furthermore, at later times, we observe power-law growth but with different exponents for each simulation. The exponent also increases as the vertical scale decreases. Interestingly, for the largest D, the growth is significantly slower than quadratic, but for the smallest D, the growth is faster than quadratic. The differing power-law exponents is a signature of the non-trivial dependence on the along-isopycnal structure of the modes. As the mode propagates towards the critical level, not only will the diapycnal wavenumber vary, but so will the eigenmodes. In other words, the entire eigenvalue problem (5.1) has parametric dependence on η. As a result, the dispersion relation and group velocity (5.3) are sensitive to the details of the background flow. Unfortunately, this means that while it is a relatively simple calculation to determine the location of the critical levels, it is much more difficult to predict the behaviour of the waves as they approach the critical level.

Critical layers are often associated with elevated mixing as the secular growth in the wave shear leads to shear instabilities. For these instabilities, we are interested in the vertical shear of both the along- and across-front velocities. For a wave with along-isopycnal displacements, i.e. $H = 0$, the across-front perturbation velocity is

$$v = -f^{-1}\omega_{min}^2 \Xi \cos\theta \Longrightarrow |\hat{v}| \approx \frac{\omega_{min}^2}{f\omega}|\hat{u}_\xi|, \tag{5.14}$$

where $|\hat{v}|$ and $|\hat{u}_\xi|$ are the amplitudes of the along-front and along-isopycnal perturbation velocities, respectively, and we have used $\cos\theta \approx 1$. Hence in the critical layer, the waves will be elliptically polarised with the along-front perturbation velocity weaker than the across-front. On the other hand, the along-front geostrophic shear of the background flow preconditions the flow for shear instability. This is a notable difference between the critical layers that form at fronts and those at the base of anticyclones. In the latter case, the vertical shear of the background flow is very weak, so the wave shear alone must be sufficient to trigger a shear instability.

6. Discussion and conclusions

In this work, we considered the trapping of near-inertial waves (NIWs) in baroclinic background flows governed by the Sawyer–Eliassen equation. Introducing a locally rotated coordinate system, in which the new basis vectors align with the principal axes of $\boldsymbol{M}$, simplifies the mathematical formulation of the problem. The phase lines and particle displacements of minimum frequency waves align with η coordinate curves. An important and very convenient approximation is that for strongly stratified flows, the principal coordinate system is aligned with and perpendicular to isopycnals, justifying the description of ξ and η as the along-isopycnal and diapycnal coordinates.

The introduction of the principal coordinate system makes it clear that the inclusion of baroclinic effects does not fundamentally alter the dynamics of inertia-gravity waves. Indeed, under a ray-tracing approximation, the mathematical formulations of the barotropic problem and the baroclinic problem in principal coordinates are identical (4.1*a*). Perhaps more usefully, the partial WKBJ approximation in the diapycnal coordinate reduces the Sawyer–Eliassen equation to an eigenvalue problem in ξ that is analogous to the Klein–Gordon equation for vertical modes in the barotropic case. We can therefore utilise results from the barotropic case to make informed predictions about the baroclinic dynamics. For example, we know that in barotropic background flows, NIW energy will preferentially accumulate in anticyclones even if it is not trapped (e.g. Balmforth *et al.* 1998; Danioux, Vanneste & Bühler 2015; Rocha *et al.* 2018). We should therefore expect that NIW energy will accumulate in the regions of low ω_{min}; i.e. in strongly baroclinic flows, NIW energy should accumulate in fronts. That being said, while we may use intuition from the barotropic case to inform the qualitative behaviour of NIWs in baroclinic flows, the details of the resulting eigenvalue problem and thus the wave propagation are sensitive to the details of the background flow.

There are notable differences between the critical layers that form in weakly baroclinic and strongly baroclinic flows. First, the location of the critical layer shifts as the baroclinic contributions to the local minimum frequency increase. For sufficiently strong baroclinic flows, the location of the critical layer will be no longer at the base of the anticyclone region, but instead towards the edges where the horizontal buoyancy gradients are strong. Second, the stronger gradients in the background flow mean that the growth rate of the diapycnal wavenumber and hence shear will be larger in the more strongly baroclinic cases. Third, these two effects combine to mean that the NIWs will reach the criterion for shear instability more rapidly in the strongly baroclinic cases. Not only does the wave shear grow more rapidly, but the critical layer has formed in a region where the background flow is already vertically sheared.

Finally, we comment on the broader impacts of the trapping of NIWs in fronts. In addition to the enhanced mixing associated with wave breaking, these waves are efficient at redistributing tracers through shear dispersion. The slantwise orientation of the trapped waves at fronts means that this shear dispersion results in a vertical flux. The NIWs at fronts could therefore play an important role in modifying the vertical distribution of oxygen, nutrients and other tracers in the ocean. Quantifying the role of NIWs in the redistribution of tracers at fronts could be an interesting avenue for future work.

Supplementary movies. Supplementary movies are available at https://doi.org/10.1017/jfm.2026.11336.

Acknowledgements. The authors thank N. Grisouard, E. Atkinson and two anonymous reviewers for their helpful comments on the manuscript.

Funding. J.P.H. and L.N.T. were supported by a grant from the National Science Foundation, award no. OCE-1851450.

Declaration of interests. The authors report no conflict of interest.

Appendix A. Principal coordinate system

We define an orthogonal curvilinear coordinate system using the rotation matrix $\boldsymbol{R}$ that diagonalises $\boldsymbol{M}$. The Jacobian matrix of the coordinate change is

$$\boldsymbol{J} = \begin{pmatrix} \dfrac{\partial x}{\partial \xi} & \dfrac{\partial x}{\partial \eta} \\ \dfrac{\partial z}{\partial \xi} & \dfrac{\partial z}{\partial \eta} \end{pmatrix} = \begin{pmatrix} h_1 \cos\theta & -h_2 \sin\theta \\ h_1 \sin\theta & h_2 \cos\theta \end{pmatrix}. \tag{A1}$$

A.1. *Consistency conditions*

For the coordinate change to be well-defined, the mixed partial derivatives of one set of coordinates with respect to the other must be equal. Enforcing this constraint provides a pair of relationships between the change in angle of the coordinate system and the scale factors:

$$\frac{\partial^2 x}{\partial \eta\, \partial \xi} = \frac{\partial^2 x}{\partial \xi\, \partial \eta} \quad \Longrightarrow \quad \frac{\partial h_1}{\partial \eta} \cos\theta - h_1 \sin\theta \frac{\partial \theta}{\partial \eta} = -\frac{\partial h_2}{\partial \xi} \sin\theta - h_2 \cos\theta \frac{\partial \theta}{\partial \xi}, \tag{A2a}$$

$$\frac{\partial^2 z}{\partial \eta\, \partial \xi} = \frac{\partial^2 z}{\partial \xi\, \partial \eta} \quad \Longrightarrow \quad \frac{\partial h_1}{\partial \eta} \sin\theta + h_1 \cos\theta \frac{\partial \theta}{\partial \eta} = \frac{\partial h_2}{\partial \xi} \cos\theta - h_2 \sin\theta \frac{\partial \theta}{\partial \xi}. \tag{A2b}$$

These maybe reduced to the relations

$$\frac{\partial \theta}{\partial \eta} = \frac{1}{h_1} \frac{\partial h_2}{\partial \xi}, \qquad \frac{\partial \theta}{\partial \xi} = -\frac{1}{h_2} \frac{\partial h_1}{\partial \eta}. \tag{A3a,b}$$

A.2. *Coordinate curves and scale factors*

The coordinate curves of ξ are defined by

$$\frac{\mathrm{d}\xi}{\mathrm{d}s} := \frac{\partial \xi}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}s} + \frac{\partial \xi}{\partial z}\frac{\mathrm{d}z}{\mathrm{d}s} = h_1^{-1}\left(\cos\theta \frac{\mathrm{d}x}{\mathrm{d}s} + \sin\theta \frac{\mathrm{d}z}{\mathrm{d}s}\right) = 0 \Longrightarrow \left.\frac{\mathrm{d}x}{\mathrm{d}z}\right|_{\xi} = -\tan\theta, \tag{A4a}$$

and similarly the coordinate curves of η are defined by

$$\frac{\mathrm{d}\eta}{\mathrm{d}s} := \frac{\partial \eta}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}s} + \frac{\partial \eta}{\partial z}\frac{\mathrm{d}z}{\mathrm{d}s} = h_2^{-1}\left(-\sin\theta \frac{\mathrm{d}x}{\mathrm{d}s} + \cos\theta \frac{\mathrm{d}z}{\mathrm{d}s}\right) = 0 \Longrightarrow \left.\frac{\mathrm{d}z}{\mathrm{d}x}\right|_{\eta} = \tan\theta. \tag{A4b}$$

These expressions may be used to integrate the coordinate curves across the domain. Having computed the coordinate curves, it is straightforward to compute the scale factors.

For an orthogonal coordinate system, the scale factors are the lengths of the covariant basis vectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ (see (3.3)), or one over the length of the contravariant basis vectors, $\boldsymbol{e}^1 := \boldsymbol{\nabla}\xi$ and $\boldsymbol{e}^2 := \boldsymbol{\nabla}\eta$, so

$$h_1 := \left[\left(\frac{\partial x}{\partial \xi}\right)^2 + \left(\frac{\partial z}{\partial \xi}\right)^2\right]^{1/2} \equiv \left[\left(\frac{\partial \xi}{\partial x}\right)^2 + \left(\frac{\partial \xi}{\partial z}\right)^2\right]^{-1/2}, \tag{A5a}$$

$$h_2 := \left[\left(\frac{\partial x}{\partial \eta}\right)^2 + \left(\frac{\partial z}{\partial \eta}\right)^2\right]^{1/2} \equiv \left[\left(\frac{\partial \eta}{\partial x}\right)^2 + \left(\frac{\partial \eta}{\partial z}\right)^2\right]^{-1/2}. \tag{A5b}$$

A.3. *Differential operators*

Using the normalised basis vectors, the gradient operator is given by

$$\boldsymbol{\nabla} := \hat{\boldsymbol{x}}\frac{\partial}{\partial x} + \hat{\boldsymbol{z}}\frac{\partial}{\partial z} \equiv \hat{\boldsymbol{\xi}}\frac{1}{h_1}\frac{\partial}{\partial \xi} + \hat{\boldsymbol{\eta}}\frac{1}{h_2}\frac{\partial}{\partial \eta}. \tag{A6}$$

If $\boldsymbol{a} = a_\xi \hat{\boldsymbol{\xi}} + a_\eta \hat{\boldsymbol{\eta}}$ is a vector field, then its divergence is

$$\boldsymbol{\nabla}\cdot\boldsymbol{a} = \frac{1}{h_1 h_2}\left(\frac{\partial(h_2 a_\xi)}{\partial \xi} + \frac{\partial(h_1 a_\eta)}{\partial \eta}\right). \tag{A7}$$

References

Alford, M.H., MacKinnon, J.A., Simmons, H.L. & Nash, J.D. 2016 Near-inertial internal gravity waves in the ocean. *Annu. Rev. Mar. Sci.* **8**, 95–123.

Asselin, O., Thomas, L.N., Young, W.R. & Rainville, L. 2020 Refraction and straining of near-inertial waves by barotropic eddies. *J. Phys. Oceanogr.* **50** (12), 3439–3454.

Balmforth, N.J., Llewellyn Smith, S.G. & Young, W.R. 1998 Enhanced dispersion of near-inertial waves in an idealized geostrophic flow. *J. Mar. Res.* **56**, 1–40.

Booker, J.R. & Bretherton, F.P. 1967 The critical layer for internal gravity waves in a shear flow. *J. Fluid Mech.* **27** (3), 513–539.

Bretherton, F.P. 1968 Propagation in slowly varying waveguides. *Proc. R. Soc. Lond. A: Math. Phys. Sci.* **302** (1471), 555–576.

Danioux, E., Vanneste, J. & Bühler, O. 2015 On the concentration of near-inertial waves in anticyclones. *J. Fluid Mech.* **773**, R2.

Eliassen, A. 1962 On the vertical circulation in frontal zones. *Geofysiske Publikasjoner* **24** (4), 147–160.

Fer, I., Bosse, A., Ferron, B. & Bouruet-Aubertot, P. 2018 The dissipation of kinetic energy in the Lofoten Basin Eddy. *J. Phys. Oceanogr.* **48** (6), 1299–1316.

Grisouard, N. 2018 Extraction of potential energy from geostrophic fronts by inertial–symmetric instabilities. *J. Phys. Oceanogr.* **48** (5), 1033–1051.

Grisouard, N. & Thomas, L.N. 2015 Critical and near-critical reflections of near-inertial waves off the sea surface at ocean fronts. *J. Fluid Mech.* **765**, 273–302.

van Haren, H. & Millot, C. 2005 Gyroscopic waves in the Mediterranean Sea. *Geophys. Res. Lett.* **32** (24), L24614.

Hayes, W.D. 1970 Conservation of action and modal wave action. *Proc. R. Soc. Lond. A: Math. Phys. Sci.* **320** (1541), 187–208.

Haynes, P.H., North, G.R. & Pyle, J.A. 2015 Critical layers. In *Encyclopedia of Atmospheric Sciences* (ed. G.R. North, J.A. Pyle & F. Zhang), 2nd edn, vol. 2, pp. 317–323. Elsevier.

Hilditch, J.P. 2025 SawyerEliassenSolver.jl. v1.0.0. Zenodo. Available at: https://doi.org/10.5281/zenodo.17072454.

Hilditch, J.P., Taylor, J.R. & Thomas, L.N. 2025 Refraction of near-inertial waves by submesoscale vorticity filaments. *J. Fluid Mech.* **1020**, A9.

Hilditch, J.P. & Thomas, L.N. 2023 Parametric subharmonic instability of inertial shear at ocean fronts. *J. Fluid Mech.* **966**, A34.

Hoskins, B.J. 1974 The role of potential vorticity in symmetric stability and instability. *Q. J. R. Meteorol. Soc.* **100** (425), 480–482.

HOSKINS, B.J. 1982 The mathematical theory of frontogenesis. *Annu. Rev. Fluid Mech.* **14** (1), 131–151.

INOUE, R., GREGG, M.C. & HARCOURT, R.R. 2010 Mixing rates across the Gulf Stream, part 1: on the formation of eighteen degree water. *J. Mar. Res.* **68** (5), 643–671.

JOYCE, T.M., TOOLE, J.M., KLEIN, P. & THOMAS, L.N. 2013 A near-inertial mode observed within a Gulf Stream warm-core ring. *J. Geophys. Res.: Oceans* **118** (4), 1797–1806.

KORTEWEG, D.J. & DE VRIES, G. 1895 XLI. On the change of form of long waves advancing in a rectangular canal, and on a new type of long stationary waves. *Lond. Edin. Dublin Phil. Mag. J. Sci.* **39** (240), 422–443.

KRUSE, C.G., RICHTER, J.H., ALEXANDER, M.J., BACMEISTER, J.T., HEALE, C. & WEI, J. 2023 Gravity wave drag parameterizations for Earth's atmosphere. In *Fast Processes in Large-Scale Atmospheric Models* (ed. L. Yangang & K. Pavlos), pp. 229–256. American Geophysical Union.

KUNZE, E. 1985 Near-inertial wave propagation in geostrophic shear. *J. Phys. Oceanogr.* **15** (5), 544–565.

KUNZE, E. 1986 The mean and near-inertial velocity fields in a warm-core ring. *J. Phys. Oceanogr.* **16** (8), 1444–1461.

KUNZE, E., SCHMITT, R.W. & TOOLE, J.M. 1995 The energy balance in a warm-core ring's near-inertial critical layer. *J. Phys. Oceanogr.* **25** (5), 942–957.

LAZANEO, C.Z., THOMAS, L., SZUTS, Z.B., CUSACK, J.M. & CHANG, K.-F. 2024 Interaction of typhoon-driven near-inertial waves with an anticyclone in the Philippine Sea. *Oceanography* **37** (4), 68–81.

LELONG, M.-P., CUYPERS, Y. & BOURUET-AUBERTOT, P. 2020 Near-inertial energy propagation inside a Mediterranean anticyclonic eddy. *J. Phys. Oceanogr.* **50** (8), 2271–2288.

MCWILLIAMS, J.C. 2017 Submesoscale surface fronts and filaments: secondary circulation, buoyancy flux, and frontogenesis. *J. Fluid Mech.* **823**, 391–432.

MOOERS, C.N.K. 1975 Several effects of a baroclinic current on the cross-stream propagation of inertial-internal waves. *Geophys. Fluid Dyn.* **6** (3), 245–275.

PRYCE, J.D. 1993 Shooting and the scaled Prüfer method. In *Numerical Solution of Sturm–Liouville Problems*, pp. 88–116. Clarendon Press.

QU, L., THOMAS, L. & GULA, J. 2021*a* Bottom mixing enhanced by tropical storm-generated near-inertial waves entering critical layers in the Straits of Florida. *Geophys. Res. Lett.* **48** (15), e2021GL093773.

QU, L., THOMAS, L.N. & HETLAND, R.D. 2021*b* Near-inertial-wave critical layers over sloping bathymetry. *J. Phys. Oceanogr.* **51** (6), 1737–1756.

ROCHA, C.B., WAGNER, G.L. & YOUNG, W.R. 2018 Stimulated generation: extraction of energy from balanced flow by near-inertial waves. *J. Fluid Mech.* **847**, 417–451.

SAWYER, J.S. 1956 The vertical circulation at meteorological fronts and its relation to frontogenesis. *Proc. R. Soc. Lond. A: Math. Phys. Sci.* **234** (1198), 346–362.

SHCHERBINA, A.Y., D'ASARO, E.A., LEE, C.M., KLYMAK, J.M., MOLEMAKER, M.J. & MCWILLIAMS, J.C. 2013 Statistics of vertical vorticity, divergence, and strain in a developed submesoscale turbulence field. *Geophys. Res. Lett.* **40** (17), 4706–4711.

TAYLOR, J.R. & FERRARI, R. 2009 On the equilibration of a symmetrically unstable front via a secondary shear instability. *J. Fluid Mech.* **622**, 103–113.

THOMAS, L.N., MOUM, J.N., QU, L., HILDITCH, J.P., KUNZE, E., RAINVILLE, L. & LEE, C.M. 2024 Blocked drainpipes and smoking chimneys: discovery of new near-inertial wave phenomena in anticyclones. *Oceanography* **37** (4), 3.

THOMAS, L.N. & TAYLOR, J.R. 2014 Damping of inertial motions by parametric subharmonic instability in baroclinic currents. *J. Fluid Mech.* **743**, 280–294.

WELLER, R.A. 1982 The relation of near-inertial motions observed in the mixed layer during the JASIN (1978) experiment to the local wind stress and to the quasi-geostrophic flow field. *J. Phys. Oceanogr.* **12** (10), 1122–1136.

WHITHAM, G.B. 1965 A general approach to linear and non-linear dispersive waves using a Lagrangian. *J. Fluid Mech.* **22** (2), 273–283.

WHITT, D.B. & THOMAS, L.N. 2013 Near-inertial waves in strongly baroclinic currents. *J. Phys. Oceanogr.* **43** (4), 706–725.

WHITTAKER, E.T. & WATSON, G.N. 2000 *A Course of Modern Analysis*, 4th edn. Cambridge Mathematical Library 8. Cambridge University Press.