# Nonlinear dynamics of forced baroclinic critical layers

## Chen Wang[1]† and Neil J. Balmforth[1]

[1]Department of Mathematics, University of British Columbia, Vancouver, BC V6T 1Z2, Canada

In this paper, we study the forcing of baroclinic critical levels, which arise in stratified fluids with horizontal shear flow along the surfaces where the phase speed of a wave relative to the mean flow matches a natural internal wavespeed. Linear theory predicts the baroclinic critical layer dynamics is similar to that of a classical critical layer, characterized by the secular growth of flow perturbations over a region of decreasing width. By using matched asymptotic expansions, we construct a nonlinear baroclinic critical layer theory to study how the flow perturbation evolves once they enter the nonlinear regime. A key feature of the theory is that, because the location of the baroclinic critical layer is determined by the streamwise wavenumber, the nonlinear dynamics filters out harmonics and the modification to the mean flow controls the evolution. At late times, we show that the vorticity begins to focus into yet smaller regions whose width decreases exponentially with time, and that the addition of dissipative effects can arrest this focussing to create a drifting coherent structure. Jet-like defects in the mean horizontal velocity are the main outcome of the critical-layer dynamics.

## 1. Introduction

A centrepiece in the theory of inviscid shear flow is the classical critical level, where the phase speed $c$ of a steady wave matches the local mean flow speed $U$. In linear theory, the levels where $c = U$ become singular, demanding the inclusion of the weak effects of unsteadyness, nonlinearity or viscosity (Maslowe 1986). Although these inclusions can remove the singularity of the linear inviscid theory, perturbations to the flow can still develop strongly in the neighbourhood of the critical levels, creating distinctive flow structures and rearrangements within the so-called critical layers that may subsequently break down to generate mixing and turbulence. In this vein, Stewartson (1978) and Warn & Warn (1976, 1978) studied the nonlinear dynamics of the critical layers of forced Rossby waves. They found that steady waves developed over the bulk of the shear flow, but that the critical layer remained unsteady, exciting mean-flow corrections and all the harmonics of the original wavenumber, and twisting up the background vorticity into Kelvin cat's eye pattern. A similar scenario exists for the critical layers of internal gravity waves travelling vertically through stratified shear flow, with important repercussions on wave breaking, momentum transport and mixing in the atmosphere (Booker & Bretherton 1967; Brown & Stewartson 1980, 1982a,b).

If the flow is stratified vertically but sheared horizontally, then a new type of critical level appears in the linear inviscid wave theory. The new critical levels arise along the surfaces where the phase speed relative to the background shear flow matches a characteristic velocity of gravity waves; *i.e.* $c - U = \pm N/k$, where $N$ is the buoyancy frequency

† Email address for correspondence: chenwang@math.ubc.ca

and $k$ is the streamwise wave number. Existing literature on these 'baroclinic critical levels', has mainly focused on the propagation of linear wave packets. Using ray-tracing theory, Olbers (1981), Basovich & Tsimring (1984) and Badulin, Shrira & Tsimring (1985) found that wave packets slow down as they approach the baroclinic critical level, never reaching it. Simultaneously, the wave amplitude and cross-stream wavenumber grow indefinitely, indicating that linear theory eventually fails in a wave-trapping process like that found earlier for classical critical levels (Bretherton 1966). Staquet & Huerre (2002) and Edwards & Staquet (2005) performed numerical simulations to study the nonlinear evolution during trapping, concluding that the trapped waves may either break into small-scale turbulence or be dissipated by dispersion, viscosity and diffusion. More related to the current work is the study by Boulanger, Meunier & Le Dizès (2007), who explored the analogues of baroclinic critical levels in stratified, titled vortices, and resolved the resulting singularities by introducing viscosity.

Baroclinic critical layers have also featured heavily in recently reported computations of three dimensional rotating stratified shear flows with self-replicating vortices (Marcus *et al.* 2015, 2016; Barranco, Pei & Marcus 2018). The replication process involves the forcing of baroclinic critical layers by internal waves excited by an initial vortex; large-amplitude re-arrangements forced in these layers then roll up to create new votices, which in turn shed more internal waves to repeat a cycle. The self-replication eventually filled the computational domain with localized vortical structures, which was suggested to be trigger for the angular momentum transport required to drive accretion in astrophysical disks that are too cool to suffer the magneto-rotational instability.

The aim of the present paper is to theoretically study the evolution of forced baroclinic critical layers, following the paradigm of Stewartson (1978) and Warn & Warn (1976, 1978) for Rossby waves, or Booker & Bretherton (1967) and Brown & Stewartson (1980, 1982a,b) for internal waves in stratified shear flow. The linear dynamics of a forced baroclinic critical layer is expected to be similar to that of a classical critical layer, owing to the similarity of the singularities in the linear wave equations. However, the subsequent nonlinear evolution is likely to be very different because the location of the baroclinic critical level itself is dictated by the streamwise wave number, which is different among all the harmonics of the original wave. This suggests that they cannot feature in the nonlinear dynamics within the baroclinic critical layer, unlike in classical critical layer theory.

The layout of the paper is as follows: in §2, we give the model and governing equations of the problem. In §3, we solve the linear problem explicitly and draw out structure that first develops within the baroclinic critical level. In §4, we extend the analysis by considering weakly nonlinear perturbations, which allows us to determine the time and length scales that characterize the nonlinear critical layer. This leads us, in §5, to derive a reduced model of nonlinear dynamics via a matched asymptotic expansion. We then present numerical solutions of the reduced model and a further asymptotic analysis of them. We explore the effects of dissipation in the baroclinic critical layer in §6, and then discuss the implications of the results and the relation to previous and future work in §7.

## 2. Model and governing equations

We consider forced disturbances to an unbounded horizontal shear flow, orientated in the $x-$direction with a constant shear rate $\Lambda > 0$ in the $y-$direction. The domain rotates around the vertical axis at angular velocity $\Omega$, and the fluid is stratified in $z$ with constant buoyancy frequency $N$. Waves are driven into the shear flow by a wavemaker that we locate along $y = 0$. This forcing has the streamwise and vertical wavenumbers,
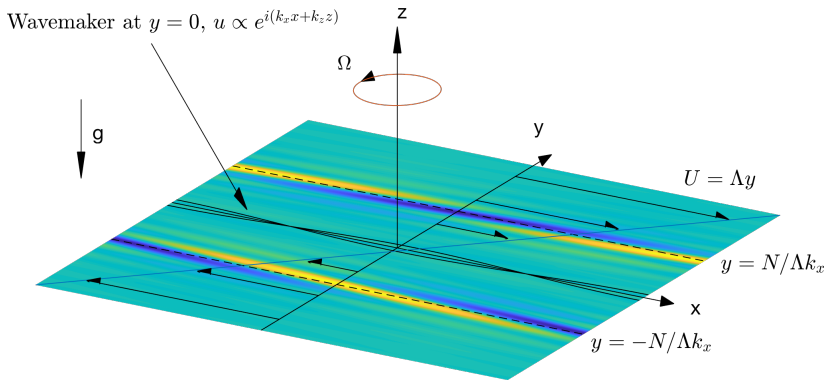
FIGURE 1. Sketch of the model. A wavemaker with wavenumber $k_x$ and $k_z$ is imposed at $y = 0$, and baroclinic critical levels are forced at $y = \pm N/(\Lambda k_x)$, corresponding to dimensionless locations $\pm \mathcal{N}$, where $\mathcal{N} = N \Lambda^{-1}$. The shading represents a rendering of the density perturbation based on the linear theory of §3.

$k_x$ and $k_z$, respectively. The baroclinic critical levels are located at $y = \pm N/(\Lambda k_x)$. The sketch of the model is shown in figure 1.

We work with a dimensionless version of the governing fluid equations in which length, time, velocity, pressure and density perturbations are scaled by $k_x^{-1}$, $\Lambda^{-1}$, $\Lambda k_x^{-1}$, $\rho_0 \Lambda^2 k_x^{-2}$ and $\rho_0 \Lambda^2/(k_x g)$, respectively. Here, $\rho_0$ is a reference density and $g$ is gravity. We employ the Boussinesq approximation and, for the most part of our study, neglect viscosity and diffusion in view of the large spatial scales that characterize geophysical and astrophysical flows. At the end of the work, we briefly explore the effect of diffusion. The perturbations to the velocity $(u, v, w)$, pressure $p$ and perturbation density $\rho$ then satisfy

$$u_t + yu_x + (1 - f)v + uu_x + vu_y + wu_z = -p_x, \tag{2.1}$$

$$v_t + yv_x + fu + uv_x + vv_y + wv_z = -p_y, \tag{2.2}$$

$$w_t + yw_x + uw_x + vw_y + ww_z = -p_z - \rho, \tag{2.3}$$

$$\rho_t + y\rho_x - \mathcal{N}^2 w + u\rho_x + v\rho_y + w\rho_z = 0, \tag{2.4}$$

$$u_x + v_y + w_z = 0, \tag{2.5}$$

where subscripts represent partial derivatives and we have introduced the dimensionless Coriolis parameter $f = 2\Omega/\Lambda$ and buoyancy frequency $\mathcal{N} = N\Lambda^{-1}$. Because our interest lies in the forcing of the baroclinic critical layers of an internal wave, we consider basic flows that are linearly stable to prevent unstable modes from dominating the dynamics. Centrifugal instabilities arise when $0 < f < 1$ (Emanuel 1994), so we set $f > 1$ or $f < 0$ to eliminate them; strato-rotational instability is not present because it requires reflective boundaries (Yavneh, McWilliams & Molemaker 2001; Wang & Balmforth 2018) which are absent here.

Initially, there is no disturbance, implying $u = v = w = \rho = p = 0$ at $t = 0$. The wavemaker is then switched on to excite waves with baroclinic critical levels. To idealize the forcing and formulate a concise mathematical problem, we assume that the wavemaker introduces a time-independent jump in the tangential horizontal velocity at $y = 0$, but not in the normal velocity. That is, we impose the jump conditions,

$$u|_{x=0+} - u|_{x=0-} = \varepsilon_0 \exp(\mathrm{i}x + \mathrm{i}mz) + \text{c.c.}, \quad v|_{x=0+} = v|_{x=0-}, \tag{2.6}$$

where $\varepsilon_0$ represents the strength, $m = k_z/k_x$, c.c. represents the complex conjugate, and the $\pm$ superscripts indicate the limits from either side. This forcing approximates

a thin, spatially periodic vortex sheet. In the numerical simulation of Marcus *et al.* (2013), waves were forced by a periodic array of localized Gaussian vortices. Our forcing therefore represents an idealization of their model in that we consider the leading-order Fourier component while neglecting the evolution and cross-stream thickness of the forcing. The configuration is slightly different to that in the studies of Stewartson (1978) and Booker & Bretherton (1967), where a wavy boundary forced the normal velocity. The current configuration implies that waves are generated at $y = 0$ and develop with baroclinic critical levels to either side (although simplifications are afforded by the symmetry described presently). Had we placed the wavemaker along a boundary at $y = 0$, only one critical level would have featured, but the wall may also make the basic flow linearly unstable (Wang & Balmforth 2018). Other idealizations include wavemakers that gradually switch on (Béland 1976), that generates disturbances with finite phase speed (displacing the baroclinic critical levels), or that with finite thickness (as for the vortices of Marcus *et al.*). Nevertheless, the precise form of forcing of the wave is not expected to affect the qualitative dynamics of the baroclinic critical layers, a feature on which we elaborate further later.

Note that the system in (2.1)-(2.6) is invariant under the transformation,

$$(u, v, w, \rho) \rightarrow -(u, v, w, \rho) \quad \text{and} \quad p \rightarrow p, \quad \text{for} \quad (x, y, z) \rightarrow -(x, y, z). \tag{2.7}$$

This observation permits us to solve the problem only in $y > 0$, and therefore consider only one baroclinic crtical layer, then generate the solution in $y < 0$ using the implied symmetry conditions.

Also, combining (2.1)-(2.5), we may derive an equation for the vertical component of vorticity:

$$\frac{D}{Dt}(v_x - u_y) - \mathcal{N}^{-2}(f - 1 + v_x - u_y)\frac{\partial}{\partial z}\frac{D\rho}{Dt} + w_x v_z - w_y u_z = 0, \tag{2.8}$$

where

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + (y + u)\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + w\frac{\partial}{\partial z}. \tag{2.9}$$

## 3. Linear theory

The linearized governing equations are

$$u_t + yu_x + (1 - f)v = -p_x, \tag{3.1}$$

$$v_t + yv_x + fu = -p_y, \tag{3.2}$$

$$w_t + yw_x + \rho = -p_z, \tag{3.3}$$

$$\rho_t + y\rho_x - \mathcal{N}^2 w = 0, \tag{3.4}$$

$$u_x + v_y + w_z = 0. \tag{3.5}$$

The linearized equation of (2.8) reduces to a conservation law of potential vorticity, $q_t + yq_x = 0$, or, given that $q = 0$ everywhere at $t = 0$,

$$q = (f - 1)\rho_z - \mathcal{N}^2(v_x - u_y) = 0. \tag{3.6}$$

In the absence of linear instability, the forcing (2.6) drives a steady wave response throughout the bulk of the flow (as can be established by solving the initial-value problem using Laplace transforms, and then performing a large-time asymptotic analysis, following Warn & Warn (1976) and Booker & Bretherton (1967)). Near the baroclinic

critical levels, however, the flow remains unsteady, requiring a finer analysis of those regions similar to that used by Stewartson (1978).

### 3.1. *The steady wave response outside the baroclinic critical layers*

The steady wave solution outside the critical layers takes the form:

$$(u, v, w, p, \rho) = [\hat{u}(y), \hat{v}(y), \hat{w}(y), \hat{p}(y), \hat{\rho}(y)] \exp{(ix + imz)} + \text{c.c.} \qquad (3.7)$$

Substituting (3.7) into (3.1)-(3.5), one can derive an equation for $\hat{p}(y)$,

$$\hat{p}'' - \frac{2y}{y^2 - f(f-1)}\hat{p}' - \left[\frac{y^2 - f(f+1)}{y^2 - f(f-1)} + m^2\frac{y^2 - f(f-1)}{y^2 - \mathcal{N}^2}\right]\hat{p} = 0, \qquad (3.8)$$

with

$$\hat{u} = \frac{(f-1)\hat{p}' - y\hat{p}}{y^2 - f(f-1)}, \quad \hat{v} = \frac{\mathrm{i}(y\hat{p}' - f\hat{p})}{y^2 - f(f-1)}, \quad \hat{w} = -\frac{my\hat{p}}{y^2 - \mathcal{N}^2}, \quad \hat{\rho} = \frac{\mathrm{i}m\mathcal{N}^2\hat{p}}{y^2 - \mathcal{N}^2}.$$
$$(3.9a, b, c, d)$$

(*cf.* Vanneste & Yavneh, 2007). Note that the singularities at $y^2 = f(f-1)$ in (3.8) and (3.9) are removable. The baroclinic critical levels $y = \pm\mathcal{N}$, however, are true singular points. The Frobenius solutions near $y = \mathcal{N}$ are,

$$\hat{p}_A = 1 - \frac{m^2[\mathcal{N}^2 - f(f-1)]}{2\mathcal{N}}(\mathcal{N} - y)\log|\mathcal{N} - y| - \alpha(\mathcal{N} - y) + \dots \qquad (3.10)$$

$$\hat{p}_B = y - \mathcal{N} + \dots \qquad (3.11)$$

where $\alpha$ is determined by the condition that $\hat{p}_A \to 0$ as $y \to \infty$. In terms of these Frobenius solutions, we express $\hat{p}$ for $y > 0$ by

$$\hat{p} = \begin{cases} A_L\hat{p}_A, & y > \mathcal{N}, \\ A_L\hat{p}_A + B_L\hat{p}_B, & 0 < y < \mathcal{N}, \end{cases} \qquad (3.13)$$

where $A_L$ and $B_L$ are constants.

Although $\hat{p}$ is bounded for $y \to \mathcal{N}$, the amplitudes of the velocity, $(\hat{u}, \hat{v}, \hat{w})$, and density, $\hat{\rho}$, all diverge, signifying that the steady wave solution fails at the critical levels. In particular, we observe that

$$\hat{p} \to A_L, \qquad \hat{\rho} \to \frac{\mathrm{i}m\mathcal{N}A_L}{2(y - \mathcal{N})} \qquad (3.14)$$

and

$$\hat{u} \to \left[\frac{m^2(f-1)}{2\mathcal{N}}(\log|\mathcal{N} - y| + 1) + \frac{\alpha(f-1) - \mathcal{N}}{\mathcal{N}^2 - f(f-1)}\right]A_L + \begin{cases} 0 & y > \mathcal{N}, \\ \frac{f-1}{\mathcal{N}^2 - f(f-1)}B_L & y < \mathcal{N}, \end{cases}$$
$$(3.15)$$

for $y \to \mathcal{N}$.

### 3.2. *The linear critical layers*

We now focus on the baroclinic critical layer at $y = \mathcal{N}$. Here, we search for an unsteady solution depending on the long timescale $T = \delta t$ and with the short spatial scale $Y = (y - \mathcal{N})/\delta$, where $\delta \ll 1$ is a small parameter organizing an asymptotic expansion. We then set

$$(u, v, w, p, \rho) = \left[\tilde{u}(Y, T), \tilde{v}(Y, T), \delta^{-1}\tilde{w}(Y, T), A_L, \delta^{-1}\tilde{\rho}(Y, T)\right] \exp{(ix + imz)} + \text{c.c.},$$
$$(3.16)$$

in view of the limits in (3.14)-(3.15).

Combining (3.3) and (3.4) to eliminate $w$, then substituting in (3.16) now gives, to leading order in $\delta$,

$$\left(\frac{\partial}{\partial T} + \mathrm{i}Y\right)\widetilde{\rho} = -\frac{1}{2}m\mathcal{N}A_L. \tag{3.17}$$

In the early stage of linear evolution, $t \sim O(1)$, $\rho \sim O(1)$, so we have the initial condition $\widetilde{\rho} \to 0$ as $T \to 0$, which yields

$$\widetilde{\rho} = -\frac{1}{2}\mathrm{i}m\mathcal{N}A_L\frac{e^{-\mathrm{i}YT} - 1}{Y}, \tag{3.18}$$

Hence

$$\rho = -\frac{1}{2}\mathrm{i}m\mathcal{N}A_L t\left[\frac{e^{-\mathrm{i}(y-\mathcal{N})t} - 1}{(y-\mathcal{N})t}\right]e^{\mathrm{i}x+\mathrm{i}mz} + \text{c.c.} \tag{3.19}$$

This solution has a spatial structure dependent on the self-similar combination $t(y - \mathcal{N})$. Hence, the amplitude grows linearly and the width of the critical layer shrinks with time.

Next, the main balance in (3.6) implies that $\widetilde{u}_Y \sim -\mathrm{i}m(f-1)\mathcal{N}^{-2}\widetilde{\rho}$, or

$$\widetilde{u}_Y = -\frac{m^2(f-1)A_L}{2\mathcal{N}}\frac{e^{-\mathrm{i}YT} - 1}{Y}. \tag{3.20}$$

But the limits of the steady wave response in (3.15) imply that $\widetilde{u}$ jumps by an amount $(f-1)B_L/[\mathcal{N}^2 - f(f-1)]$ across the baroclinic critical layer. Hence,

$$B_L = -\frac{m^2 A_L[f(f-1) - \mathcal{N}^2]}{2\mathcal{N}}\lim_{L\to\infty}\int_{-L}^{L}(e^{-\mathrm{i}YT} - 1)\frac{\mathrm{d}Y}{Y} = \mathrm{i}\pi\frac{m^2[f(f-1) - \mathcal{N}^2]}{2\mathcal{N}}A_L. \tag{3.21}$$

(*cf.* Stewartson 1978).

### 3.3. *Closure*

We can now apply the forcing condition to close the problem. The symmetry property (2.7) applied to the steady wave (3.7) indicates that

$$[\hat{u}(y), \hat{v}(y), \hat{w}(y), \hat{\rho}(y)] = -[\hat{u}(-y), \hat{v}(-y), \hat{w}(-y), \hat{\rho}(-y)]^*, \quad \hat{p}(y) = \hat{p}(-y)^*, \tag{3.22}$$

where the superscript $^*$ represents the complex conjugate. Hence, substituting the steady wave solution into the jump condition (2.6) representing the forcing, we arrive at

$$(A_L - A_L^*)\hat{p}_A(0) + (B_L - B_L^*)\hat{p}_B(0) = 0,$$

$$(A_L + A_L^*)\hat{p}_A'(0) + (B_L + B_L^*)\hat{p}_B'(0) = -f\varepsilon_0. \tag{3.23a,b}$$

Exploiting (3.21), we obtain

$$A_L = -\frac{f\varepsilon_0(\hat{p}_A - \mathrm{i}\beta\hat{p}_B)}{2(\hat{p}_A\hat{p}_A' + \beta^2\hat{p}_B\hat{p}_B')}\bigg|_{y=0}, \quad \beta = \frac{\pi m^2[f(f-1) - \mathcal{N}^2]}{2\mathcal{N}}. \tag{3.24}$$

The amplitude of the pressure perturbation at the critical layer is therefore

$$\varepsilon = |A_L| = \frac{|f\varepsilon_0|\sqrt{\hat{p}_A^2 + \beta^2\hat{p}_B^2}}{2\,|\hat{p}_A\hat{p}_A' + \beta^2\hat{p}_B\hat{p}_B'|}\bigg|_{y=0}. \tag{3.25}$$

A sample steady wave solution is plotted in figure 2.

Note that equations (3.23)-(3.25) appear to become trivial if $f = 0$, suggesting that rotation is essential to the forcing of the baroclinic critical layer. In fact, a deeper analysis of the Frobenius solutions demonstrates that this is not the case, because $\hat{p}_A'(0)$ and
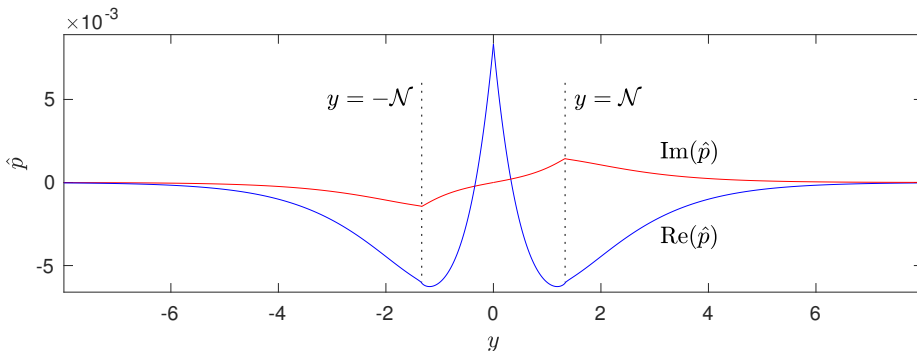
FIGURE 2. Steady-wave solution $\hat{p}$, with $m = 0.5$, $\mathcal{N} = 4/3$, $f = 4/3$, $\varepsilon_0 = 0.05$ (*cf.* Marcus *et al.* 2013). The critical levels $y = \pm\mathcal{N}$ are indicated.

$\hat{p}'_B(0)$ become $O(f)$ in this limit, and the closure relation in (3.23) remains non-trivial. Consequently, in the model, we may take the limit $f \to 0$, highlighting how rotation is not an essential ingredient to the dynamics.

The same feature does not apply to the vertical wavenumber or stratification, which control the secular growth inside the critical layer, as seen in (3.18) and (3.20); without either a vertical dependence in the forcing or stratification, there is no baroclinic critical-layer dynamics. Note that, despite appearances, the limit $\mathcal{N} \to 0$ in (3.20) is not problematic: further analysis of $\hat{p}_A$ and $\hat{p}_B$ indicates that $|A_L| \sim \mathcal{N}/\log\mathcal{N}$ for $\mathcal{N} \to 0$, and so the secular growth in the critical layer is eliminated in this limit.

It is also noteworthy that, in the limit that any of the parameters $m$, $f$, or $\mathcal{N}$ are large, the disturbance decays exponentially from the forcing to the baroclinic critical levels (*cf.* Vanneste & Yavneh (2007) and Wang & Balmforth (2018)). The amplitude ratio $\varepsilon/\varepsilon_0$ then becomes exponentially small, and the secular growth in the critical layer is much weakened.

## 4. The weakly nonlinear critical layer

We now advance beyond linear theory and perform a weakly nonlinear expansion by setting

$$
\begin{aligned}
(u, v, w, \rho, p) = \varepsilon &\left\{ [u_1(Y,T), v_1(Y,T), \delta^{-1}w_1(Y,T), \delta^{-1}\rho_1(Y,T), p_1(Y,T)]e^{ix+imz} + \text{c.c.} \right\} \\
&+ \varepsilon^2 [u_0(Y,T), v_0(Y,T), w_0(Y,T), \rho_0(Y,T), p_0(Y,T)] \\
&+ \varepsilon^2 \left\{ [u_2(Y,T), v_2(Y,T), w_2(Y,T), \rho_2(Y,T), p_2(Y,T)]e^{2(ix+imz)} + \text{c.c.} \right\},
\end{aligned}
\tag{4.1}
$$

focussing upon the critical layer with $y = \mathcal{N} + \delta Y$. The scaling of the fundamental Fourier component follows the linear critical layer theory outlined above, and we have $\varepsilon[u_1, v_1, w_1, \rho_1, p_1] \to [\tilde{u}, \tilde{v}, \tilde{w}, \tilde{\rho}, A_L]$ at early times ($T \ll 1$). The goal of the current section is to identify the timescale and width of the critical layer (as dictated by the small parameter $\delta$) for which the mean flow correction and first harmonic reach sufficient strength to modify the evolution of fundamental mode. This connects $\delta$ to the amplitude parameter $\varepsilon$, establishing the scalings of the nonlinear critical layer.

### 4.1. *Mean-flow response*

The mean-flow component of (2.5) gives $v_{0Y} = 0$, which implies $v_0 = 0$ since the mean flow response decays outside the critical layer. The streamwise mean-flow velocity $u_0$ is

described by the $j = 0$ component of (2.1), which is

$$\frac{\partial u_0}{\partial T} = \delta^{-2}(\mathrm{i}mw_1u_1^* - v_1^*u_{1Y}) + \text{c.c.} \tag{4.2}$$

To leading order in $\delta$, the mean-flow components of (2.3) and (2.4) are,

$$\rho_0 = -\delta^{-2}v_1^*w_{1Y} + \text{c.c.}, \tag{4.3}$$

$$-\mathcal{N}^2 w_0 = -\delta^{-2}(v_1^*\rho_{1Y} + \mathrm{i}mw_1^*\rho_1) + \text{c.c.} \tag{4.4}$$

Thus, $u_0$, $w_0$, and $\rho_0$ are all $O(\delta^{-2})$.

## 4.2. *First harmonic*

The largest first harmonic components of (2.3), (2.4), (2.5) and (2.8) indicate that

$$2\mathrm{i}\mathcal{N}w_2 + \rho_2 + 2\mathrm{i}mp_2 = -\delta^{-2}(v_1w_{1Y} + \mathrm{i}mw_1^2), \tag{4.5}$$

$$2\mathrm{i}\mathcal{N}\rho_2 - \mathcal{N}^2 w_2 = -\delta^{-2}(v_1\rho_{1Y} - \mathrm{i}mw_1\rho_1). \tag{4.6}$$

$$\delta^{-1}v_{2Y} + 2\mathrm{i}mw_2 = 0. \tag{4.7}$$

$$\frac{u_{2Y}}{\delta} + \frac{2\mathrm{i}m(f-1)}{\mathcal{N}^2}\rho_2 = \frac{\mathrm{i}\delta^{-2}}{2\mathcal{N}}\left[\frac{2\mathrm{i}m(f-1)}{\mathcal{N}^2}(v_1\rho_{1Y} + \mathrm{i}mw_1\rho_1) + (v_1u_{1Y} + \mathrm{i}mw_1u_1)_Y\right]. \tag{4.8}$$

However, (2.2) demands that $p_2 = O(\delta u_2, \delta v_2)$ and so $p_2$ is much smaller than $w_2$ or $\rho_2$. Hence,

$$w_2 = \frac{\delta^{-2}}{3\mathcal{N}^2}[2\mathrm{i}\mathcal{N}(v_1w_{1Y} + \mathrm{i}mw_1^2) - v_1\rho_{1Y} - \mathrm{i}mw_1\rho_1], \tag{4.9}$$

$$\rho_2 = \frac{\delta^{-2}}{3\mathcal{N}}[\mathcal{N}(v_1w_{1Y} + \mathrm{i}mw_1^2) + 2\mathrm{i}(v_1\rho_{1Y} + \mathrm{i}mw_1\rho_1)], \tag{4.10}$$

which are $O(\delta^{-2})$, whereas $u_2$ and $v_2$ are $O(\delta^{-1})$.

## 4.3. *Weakly nonlinear feedback*

On again combining (2.3) and (2.4), we find the fundamental components,

$$\left(\frac{\partial}{\partial T} + \mathrm{i}Y\right)\rho_1 + \frac{1}{2}m\mathcal{N}p_1 = -\varepsilon^2\delta^{-1}\mathrm{i}u_0\rho_1, \tag{4.11}$$

with the leading-order nonlinear terms included on the right, and after a considerable number of cancellations stemming from the use of (4.3), (4.4), (4.9) and (4.10) and the leading-order relations $\rho_1 = -\mathrm{i}\mathcal{N}w_1$ and $v_{1Y} = -\mathrm{i}mw_1$. Note that the nonlinear terms generated by the first harmonic and mean-flow components $w_0$ and $\rho_0$ completely cancel out at this stage, leaving only the effect of the modification to the streamwise mean flow $u_0$. But the scaling established for the mean flow correction implies that the right-hand side of (4.11) is $O(\delta^{-3}\varepsilon^2)$. Thus, the mean flow feedbacks on the fundamental mode when $\delta = \varepsilon^{2/3}$. That is, for

$$t = O(\varepsilon^{-\frac{2}{3}}), \quad y = \mathcal{N} + O(\varepsilon^{\frac{2}{3}}). \tag{4.12}$$

These are the scalings for the nonlinear critical layer theory outlined in the next section.

Note that we may extend the analysis to consider the higher harmomics. One finds that when $\delta = \varepsilon^{2/3}$, the Fourier component $e^{\mathrm{i}j(x+mz)}$ with $j > 1$ is $O(\varepsilon^{j/3})$, which signifies that the higher-order harmonics $j \geqslant 3$ are still weak when the mean flow correction begins to feedback on the fundamental. Thus, they play no role in the nonlinear theory.

## 5. Nonlinear critical-layer theory

### 5.1. *The reduction*

Motivated by the weakly nonlinear analysis, we now introduce the rescalings,

$$T = \varepsilon^{\frac{2}{3}} t, \quad Y = \frac{y - \mathcal{N}}{\varepsilon^{\frac{2}{3}}}. \tag{5.1}$$

The outer solution for the pressure is

$$p = \varepsilon p_1 e^{i(x+mz)} + \text{c.c.}, \qquad p_1 = \begin{cases} A(T)\hat{p}_A(y), & y > \mathcal{N}, \\ A(T)\hat{p}_A(y) + B(T)\hat{p}_B(y), & 0 < y < \mathcal{N}, \end{cases} \tag{5.2}$$

which is a single dominant Fourier mode characterized by the steady wave solution. However, the amplitudes $A$ and $B$ now evolve with the slow time $T$, because the nonlinear evolution of critical layer can affect the outer flow. Initially, $A$ and $B$ are given by the linear analysis:

$$A(0) = \frac{A_L}{\varepsilon}, \quad B(0) = \frac{B_L}{\varepsilon}. \tag{5.3}$$

Inside the critical layers, we set

$$p = \varepsilon A(T) e^{i(x+mz)} + \text{c.c.} + ..., \quad [w, \rho] = \varepsilon^{\frac{1}{3}} [w_1(Y,T), \rho_1(Y,T)] e^{i(x+mz)} + \text{c.c.} + ...$$

$$[u, v] = \varepsilon [u_1(Y,T), v_1(Y,T)] e^{i(x+mz)} + \text{c.c.} + \varepsilon^{\frac{2}{3}} [U_0(Y,T), 0] + ... \tag{5.4}$$

Equation (4.11) and the leading-order fundamental-mode components of (2.1), (2.3)-(2.5) and (2.8) now become

$$\frac{\partial \rho_1}{\partial T} + iY\rho_1 + \frac{m\mathcal{N}}{2} A = -iU_0 \rho_1. \tag{5.5}$$

$$i\mathcal{N} u_1 - (f-1) v_1 + iA = -v_1 U_{0Y}, \tag{5.6}$$

$$w_1 = \frac{i}{\mathcal{N}} \rho_1, \qquad v_{1Y} = -imw_1, \tag{5.7}$$

$$\mathcal{N}^2 u_{1Y} + im(f - 1 - U_{0Y})\rho_1 = i\mathcal{N} v_1 U_{0YY}. \tag{5.8}$$

The initial condition of $\rho_1$ is given by the linear result

$$\rho_1 \to -\frac{im\mathcal{N}A(0)}{2} \frac{e^{-iYT} - 1}{Y}, \quad T \to 0. \tag{5.9}$$

Similar to (4.2), the mean-flow velocity $U_0$ is governed by

$$\frac{\partial U_0}{\partial T} = -v_1^* u_{1Y} + imw_1 u_1^* + \text{c.c.} \tag{5.10}$$

The initial condition is $U_0 \to 0$ as $T \to 0$, as in early linear evolution the mean-flow modification is minimal.

It is possible to algebraically manipulate (5.5)-(5.8) and then integrate in $T$ to show that

$$U_0 = -\frac{2}{\mathcal{N}^3} |\rho_1|^2, \tag{5.11}$$

a result that can be traced back to the fact that the change to the mean flow is given by the Eulerian pseudo-momentum (Bühler 2014), which is the right-hand side of (5.11) to leading order. Hence

$$\frac{\partial \rho_1}{\partial T} + iY\rho_1 + \frac{1}{2} m\mathcal{N}A = i\frac{2}{\mathcal{N}^3} |\rho_1|^2 \rho_1. \tag{5.12}$$

To match the inner and outer solutions, we first note, from (5.7), that $v_{1Y} = m\mathcal{N}^{-1}\rho_1$.

Integrating this relation in $Y$ over the critical layer then provides the jump of the outer solution $v_1 = \mathrm{i}(yp_{1,y} - fp_1)/[y^2 - f(f-1)]$ for the limit of $y \to \mathcal{N}$ (cf. (3.9b)), which yields

$$B = -\mathrm{i}m\frac{f(f-1) - \mathcal{N}^2}{\mathcal{N}^2}\int_{-\infty}^{\infty}\rho_1\mathrm{d}Y, \tag{5.13}$$

in a similar manner to §3.2 and (3.21).

Last, we again use the forcing condition at $y = 0$ to close the problem:

$$(A - A^*)\hat{p}_A(0) + (B - B^*)\hat{p}_B(0) = 0,$$
$$(A + A^*)\hat{p}_A'(0) + (B + B^*)\hat{p}_B'(0) = -f\frac{\varepsilon_0}{\varepsilon} \tag{5.14}$$

(cf. §3.3 and (3.23)). Note that the form of the forcing impacts the reduced model only through the closure relations in (5.14). Had we used a different idealization of the forcing here, there would be a different algebraic relation between $A$, $B$ and $\varepsilon_0/\varepsilon$. However, this relation still connects $A$ with the forcing amplitude and the integral of $\rho_1$ over the critical layer, and in the scaled, canonical system presented below, all that would change would be how the parameters of that system (denoted $c_0$, $c_1$ and $c_2$ in §5.2) depend on the original physical constants. In this sense, the reduced model is independent of the choice of forcing.

### 5.2. Canonical system

The final rescalings

$$\rho_1 = \left(\frac{m\mathcal{N}^4}{4}\right)^{\frac{1}{3}}\gamma(\eta,\tau), \quad T = \left(\frac{2\mathcal{N}}{m^2}\right)^{\frac{1}{3}}\tau, \quad Y = \left(\frac{m^2}{2\mathcal{N}}\right)^{\frac{1}{3}}\eta, \tag{5.15}$$

lead to the canonical form,

$$\frac{\partial\gamma}{\partial\tau} + \mathrm{i}\eta\gamma + A = \mathrm{i}|\gamma|^2\gamma, \tag{5.16}$$

$$A(\tau) = c_0 + \frac{\mathrm{i}c_1}{\pi}\int_{-\infty}^{\infty}\gamma_r\mathrm{d}\eta - \frac{c_2}{\pi}\int_{-\infty}^{\infty}\gamma_i\mathrm{d}\eta, \tag{5.17}$$

where $\gamma = \gamma_r + \mathrm{i}\gamma_i$,

$$c_0 = -\mathrm{sgn}\left(\frac{f}{\hat{p}_A'(0)}\right)\frac{|1 + c_1c_2|}{\sqrt{1 + c_1^2}}, \tag{5.18}$$

and

$$\left\{\begin{array}{c} c_1 \\ c_2 \end{array}\right\} = \frac{\pi m^2[f(f-1) - \mathcal{N}^2]}{2\mathcal{N}}\left\{\begin{array}{c} \hat{p}_B(0)/\hat{p}_A(0) \\ \hat{p}_B'(0)/\hat{p}_A'(0) \end{array}\right\}. \tag{5.19}$$

For $\tau \ll 1$, we must match $\gamma(\eta,\tau)$ to the corresponding solution of the linear problem, given by

$$\gamma = \mathrm{i}A\frac{1 - e^{-\mathrm{i}\eta\tau}}{\eta}, \qquad A = \frac{c_0(1 - \mathrm{i}c_1)}{1 + c_1c_2}, \tag{5.20}$$

which provides the initial condition for (5.16).

The reduced model equations in (5.16)-(5.20) are solved numerically in the next section. The system is integro-differential in the sense that (5.16) is an equation of motion in time, solved at each level of $\eta$, with the integral constraint in (5.17). There is no dependence on either $x$ or $z$, because the leading-order dynamics involves only the fundamental mode of the forcing wave pattern and the mean-flow response (which is then prescribed by the pseudo-momentum). The only nonlinearity is the cubic term on the right of (5.16),
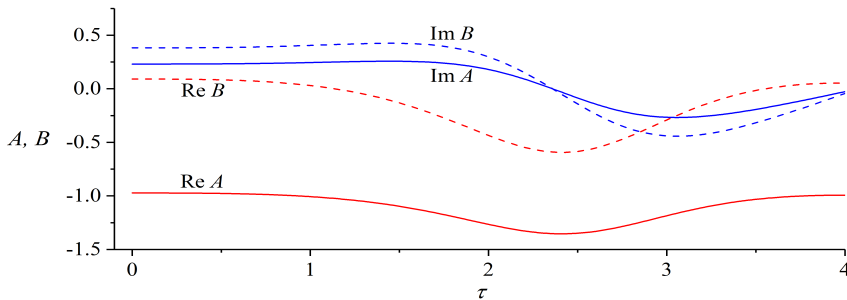
FIGURE 3. Evolution of $A$ and $B$ with $\tau$; $m = 1/2$, $f = 4/3$, $\mathcal{N} = 4/3$.

which is generic in weakly nonlinear theories of non-dissipative systems with few degrees of freedom. The model is therefore rather different from those that emerge for classical forced critical layers, which usually take the form of partial differential equations in all the spatial variables. The reduced model has the two parameters, $c_1$ and $c_2$, and the choice of sign for $f\hat{p}'_A(0)$ in $c_0$. In most situations $\hat{p}_A$ and $\hat{p}_B$ are characterized by a similar exponential away from $y = 0$, implying $c_1 \approx c_2$.

From (5.16)-(5.17), one can establish that the quantity,

$$\mathcal{H} = \int_{-\infty}^{\infty} \left[ \frac{1}{2}|\gamma|^4 - \eta|\gamma|^2 + 2\mathrm{Im}(A^*\gamma) \right] \mathrm{d}\eta + \frac{c_1}{\pi} \left[ \int_{-\infty}^{\infty} \gamma_r \mathrm{d}\eta \right]^2 + \frac{c_2}{\pi} \left[ \int_{-\infty}^{\infty} \gamma_i \mathrm{d}\eta \right]^2 , \quad (5.21)$$

must be conserved, and therefore equal to $\pi c_1(1 + c_1 c_2)/(1 + c_1^2)$ in view of the initial conditions. This constraint implies that the linear-in-time growth of $\gamma(\eta, \tau)$ predicted by linear theory must eventually become arrested, as otherwise the quartic first term in (5.21) cannot be counter balanced by the remaining quadratic and constant terms. To determine the manner in which the arrest takes place, we turn to a numerical solution of the reduced model.

### 5.3. *Numerical solutions*

To solve the canonical system of equations numerically, we first select a grid in $\eta$ spanning a finite domain (we use 1501 equally spaced gridpoints over the interval $1.5 < \eta < 3$ where $\gamma$ has large gradients, then 1544 gridpoints distributed evenly over $-25 < \eta < 1.5$ and $3 < \eta < 25$). We then integrate (5.16) forward in time numerically using a 4th-order Runge-Kutta method at each of the grid points. To evaluate the integrals in (5.17), we use an approach similar to Warn & Warn (1978) to extrapolate the limits to infinity. We use parameter settings guided by the computations of of Marcus *et al.* (2013): $m = 1/2$, $f = 4/3$, $\mathcal{N} = 4/3$, which yield $c_1 = 0.238$, $c_2 = 0.219$.

Figure 3 displays the evolution in $\tau$ of the forced wave amplitudes, $A$ and $B$, which is relatively mild with $\mathrm{Re}(A) \approx c_0 \approx -1$ and $\mathrm{Im}(A)$, $\mathrm{Re}(B)$ and $\mathrm{Im}(B)$ all remaining small. This mild behaviour results because, in (5.17), $|c_1|$ and $|c_2|$ are fairly small. Thus, the forced wave evolves slowly over the bulk of the shear flow (*i.e.* the outer region), maintaining a profile similar to the linear distribution in figure 2.

The density perturbation $\gamma(\eta, \tau)$, shown in figure 4, exhibits a richer behaviour: for $\tau < 1$, the numerical solution follows the linear prediction in (5.20), with its characteristically developing undulations and linear growth near $\eta = 0$ (see figure 4(a,b)). Once $|\zeta|$ reaches order-one values there, however, the growth of the numerical solution saturates, as demanded by the constraint in (5.21). Despite this, the solution continues to undulate over increasingly shorter spatial scales. Moreover, nonlinear effects

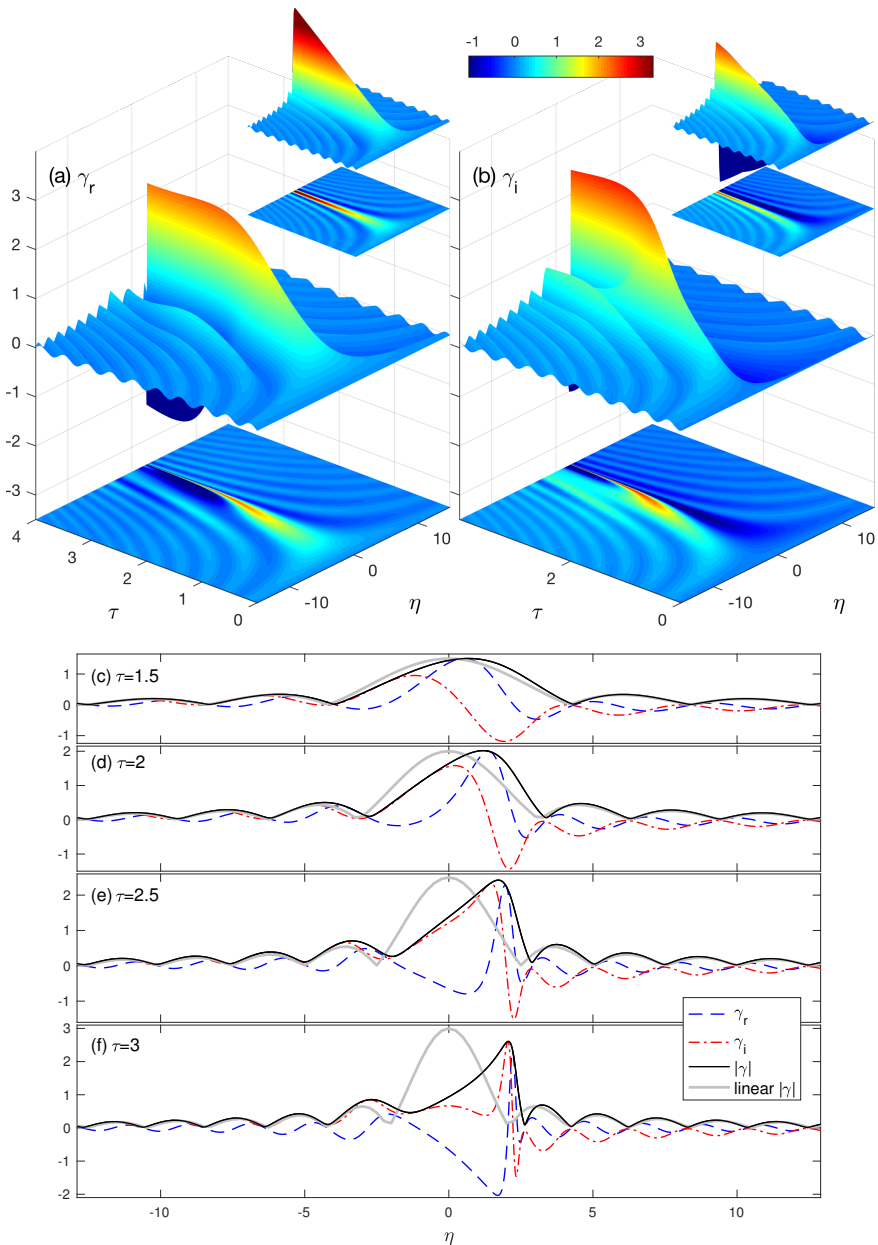FIGURE 4. (a) Real and (b) imaginary parts of the critical-layer density perturbation $\gamma(\eta, \tau)$, shown as surfaces above the $(\eta, \tau)-$plane. To prevent the viewing perspective from obscuring parts of the solution, we also show density maps of the solutions underneath. The insets show corresponding plots of the linear solution in (5.20). Panels (c)–(f) plot snapshots of $\gamma(\eta, \tau)$ at the times indicated; the linear result for $|\gamma(\eta, \tau)|$ is also included. ($m = 1/2$, $f = 4/3$, $\mathcal{N} = 4/3$.)

distort the density profile further, shifting the maximum magnitude from $\eta = 0$ to a small, positive level in $\eta$ and generating pronounced fine structure over a narrow region nearby.

The rapid spatial variation in $\gamma(\eta, \tau)$ significantly impacts the critical-layer vorticity, which depends on the $\eta-$derivatives of $\gamma(\eta, \tau)$. In particular, the leading-order vertical
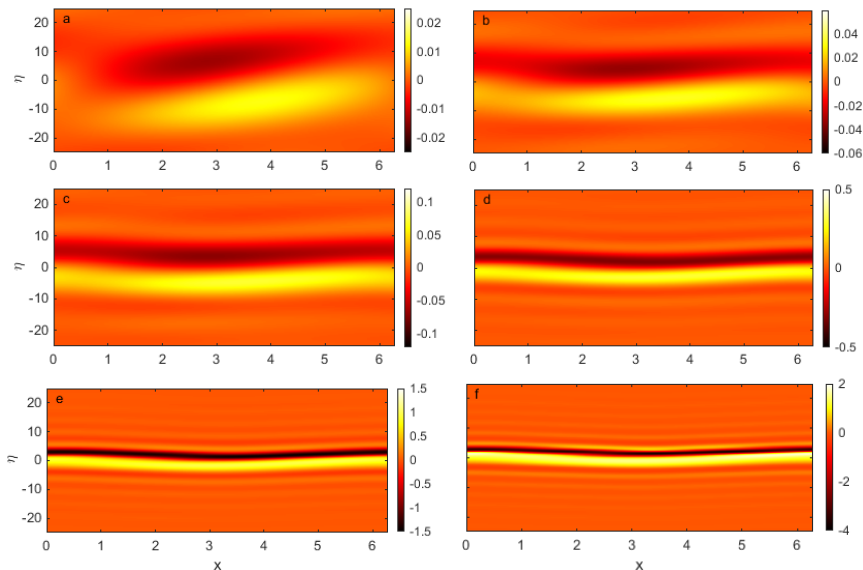
FIGURE 5. Snapshots of vertical vorticity $\zeta$ within the baroclinic critical layer near $y = \mathcal{N} = 4/3$, plotted as a colormap on the $(x, \eta)$−plane for (a) $\tau = 0.3$, (b) $\tau = 0.45$, (c) $\tau = 0.6$, (d) $\tau = 1$, (e) $\tau = 1.5$, and (f) $\tau = 1.8$. The domain plotted is $|\eta| < 25$, corresponding to $|y - \mathcal{N}| < 0.39$, at cross-section $z = 0$ and over one streamwise wavelengths of the forcing pattern. ($\varepsilon_0 = 0.05$, $\varepsilon = 0.0062$, $m = 1/2$, $f = 4/3$, $\mathcal{N} = 4/3$.)

vorticity is given by the mean-flow vorticity $\zeta_0$:

$$\zeta \sim \zeta_0 \equiv \frac{\partial}{\partial \eta} |\gamma|^2. \tag{5.22}$$

However, from the matched asymptotic expansion, we may reconstruct $\zeta(x, \eta, z, t)$ to higher orders, incorporating the fundamental Fourier mode $\zeta_1$ and first harmonic $\zeta_2$, as summarized in Appendix A. The evolution of the reconstructed vertical vorticity field is plotted in figure 5. For early times, $\zeta_0 \ll 1$, and the vertical vorticity is actually given by the higher-order linear solution (as in figure 5a, *cf* (3.20)). With the increase of $\tau$, the vorticity distribution tilts over and $\zeta_0$ grows to dominate $\zeta$, as seen in figure 5b,c. This growth leads to the distinctive dipolar stripe seen in figure 5d. In the later stages of evolution (figure 5e,f), the stripe becomes stronger and more focussed, shifting slightly above $\eta = 0$, and corresponding to the sharpening oscillations in $\gamma$ seen in figure 4.

The behaviour of the numerical solution seen in figures 3-5 is generic for most parameter settings; for moderate $m$, $f$ (either $f > 1$ or $f < 0$) and $\mathcal{N}$, the parameters $c_1$ and $c_2$ of the reduced model are relatively small in magnitude, prompting similar dynamics. Even when $|c_1|$ and $|c_2|$ become order one, the evolution still bears qualitative similarities. However, more complicated behaviour can occur in the reduced model when these parameters take higher values. Such parameter settings can be achieved at special combinations of $m$, $f$ and $\mathcal{N}$ for which $\hat{p}_A(0)$ becomes small, or perhaps for other types of forcing. We avoid consideration of special situations of this sort, and instead turn to a deeper analysis of the focussing dynamics observed in the reduced model.

### 5.4. *Long-time focussing*

In view of the result that $A$ changes slowly, we now use the approximation of $A =$ constant to gain further analytical insights to the focussing phenomenon. This device was

used previously by Stewartson (1978) to obtain an analytical solution to the nonlinear evolution of Rossby wave critical layers. In our model, constant $A$ in (5.17) requires $c_1 = c_2 = 0$, hence $A = -\mathrm{sgn}(f\hat{p}'_A(0))$, which is $-1$ for the current parameter setting. The evolution equation (5.16) can then be written as the one-degree-of-freedom Hamiltonian system,

$$\frac{\partial \gamma_r}{\partial \tau} = \frac{\partial H}{\partial \gamma_i} = 1 + \eta \gamma_i - \gamma_r^2 \gamma_i - \gamma_i^3,$$
$$\frac{\partial \gamma_i}{\partial \tau} = -\frac{\partial H}{\partial \gamma_r} = -\eta \gamma_r + \gamma_r^3 + \gamma_r \gamma_i^2, \qquad (5.23)$$

with Hamiltonian,

$$H = -\frac{1}{4}\left(\gamma_r^2 + \gamma_i^2\right)^2 + \frac{1}{2}\eta\left(\gamma_r^2 + \gamma_i^2\right) + \gamma_i \qquad (5.24)$$

(the point-wise version of the conserved quantity $\mathcal{H}$ in (5.21) for $c_1 = c_2 = 0$). For the specific initial condition of our critical-layer problem, $H = 0$ for all values of $\eta$.

Figure 6(a) illustrates the phase portrait of the system (5.23) for the special choice $\eta = \eta_c = 3/\sqrt[3]{2}$. In this case, the orbit from $(\gamma_r, \gamma_i) = (0,0)$ lies along a separatrix that converges to a saddle point at $(\gamma_r, \gamma_i) = (0, -\gamma_e)$, for $\tau \to \infty$, with $\gamma_e = \sqrt[3]{2} \approx 1.26$. Trajectories from $(\gamma_r, \gamma_i) = (0,0)$ for a spread of values of $\eta$ around $\eta_c$ are illustrated in figure 6(b); the presence of the separatrix at $\eta = \eta_c$ implies that these trajectories bifurcate in direction on the phase plane on passing through that special level. Thus, a small variation in $\eta$ about $\eta_c$ can result in a large change of $\gamma$ at later times, implying high values of $\gamma_\eta$ to feed into $\zeta$.

For the numerical solutions of §5.3, although $c_1$ and $c_2$ do not vanish, the forced-wave amplitude does remain slowly varying in $\tau$, leading to a qualitatively similar dynamics: figure 6(c) plots the phase portrait of $\gamma$ for five values of $\eta$ within the region where the dipolar stripe is focussed. As $\eta$ varies from 2.38 to 2.48, the trajectories for different levels abruptly switch in direction near the point $(\gamma_r, \gamma_i) = (0, -1.2)$. Although the slow variation of $A(\tau)$ precludes any trajectory from reaching a steady value, the numerical solution for $\eta = 2.43$ slows down, lingers and hesitates before selecting one of the two possible directions, much like the orbits for $c_1 = c_2 = 0$ near the separatix in figure 6(a,b). The level of this trajectory is slightly shifted from $3/\sqrt[3]{2} \approx 2.38$ because $c_1$ and $c_2$ are non-zero and $A(\tau) \neq -1$. Nevertheless, we conclude that the close passage to an effective saddle point on the $(\gamma_r, \gamma_i)$ phase plane is responsible for the focussing effect. For the numerical solution, we therefore define $\eta = \eta_c \approx 2.43$ to be the level for which $\gamma$ evolves slowest near the effective saddle, and refer to this location as the nonlinear critical level.

Continuing the analysis for $c_1 = c_2 = 0$, we may linearize the system (5.23) about $\gamma = -\mathrm{i}\gamma_e$ to find that

$$\frac{\partial}{\partial \tau}\left(\begin{array}{c} \gamma_r \\ \gamma_i + \gamma_e \end{array}\right) = \left(\begin{array}{cc} 0 & \eta - 3\gamma_e^2 \\ -\eta + \gamma_e^2 & 0 \end{array}\right)\left(\begin{array}{c} \gamma_r \\ \gamma_i + \gamma_e \end{array}\right). \qquad (5.25)$$

The two eigenvalues of the matrix are $\pm\sigma$, with corresponding eigenvectors $\mathbf{v}_+$ and $\mathbf{v}_-$, where

$$\sigma = \sqrt{[(\gamma_e^2 - \eta)][\eta - 3\gamma_e^2]} \approx \sigma_c = \frac{\sqrt{3}}{\sqrt[3]{2}} \quad \text{if} \quad \eta \approx \eta_c. \qquad (5.26)$$

The solution of (5.25) is then

$$\left(\begin{array}{c} \gamma_r \\ \gamma_i + \gamma_e \end{array}\right) = r_+ \mathbf{v}_+ e^{\sigma(\tau - \tau_0)} + r_- \mathbf{v}_- e^{-\sigma(\tau - \tau_0)}, \qquad (5.27)$$
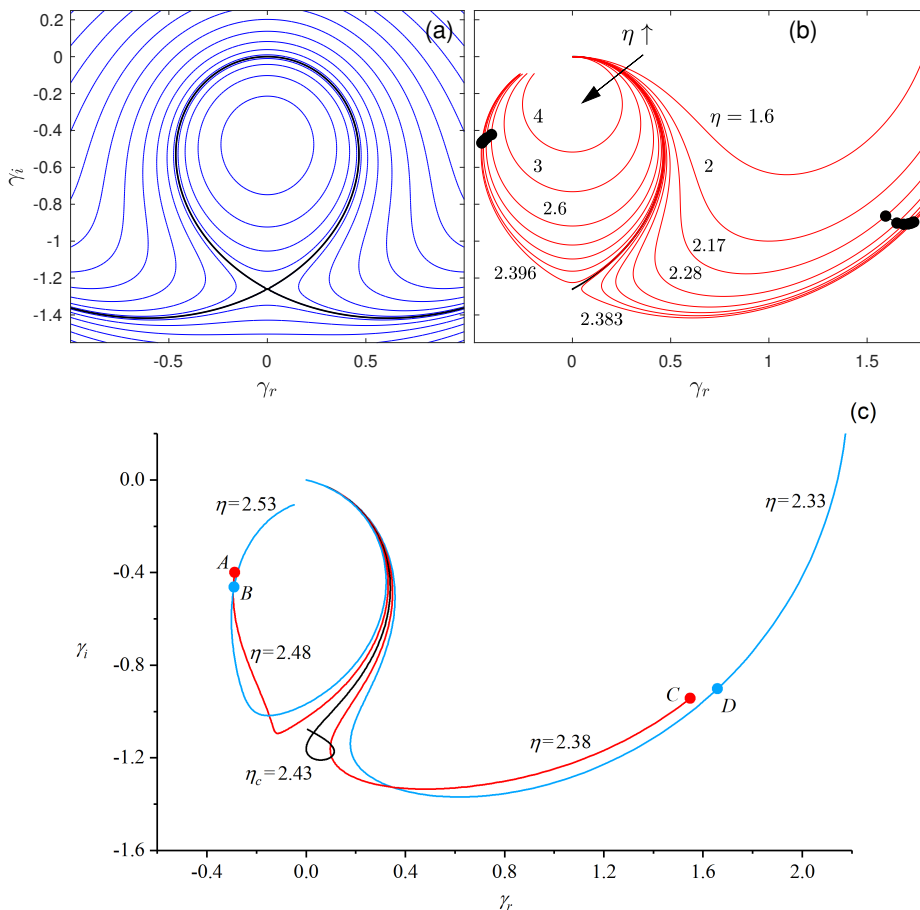
FIGURE 6. Phase portraits of $(\gamma_r, \gamma_i)$ for (a) the Hamiltonian system (5.23) with $\eta = 3/\sqrt[3]{2}$ and various $H$, with the thicker line indicating $H = 0$, (b) trajectories from the point $(\gamma_r, \gamma_i) = (0,0)$ for a selection of values of $\eta$, and (c) the numerical solution of §5.3, at the five values of $\eta$ indicated. The black points in (b) and the (red and blue) pairs marked $(A, B)$ and $(C, D)$ in (c) have the same values of $|\eta - \eta_c|e^{\sigma\tau}$.

for some constants $r_\pm$ and a time constant $\tau_0$ indicating when the orbit reaches the neighbourhood of the saddle point.

Now, along the separatrix converging to $\gamma = -\mathrm{i}\gamma_e$ for $\eta = \eta_c$, the constant $r_+$ must vanish. But when $\eta$ is close to, but not at $\eta_c$, this factor is small but finite, hence a local linearization of $r_+(\eta)$ near $\eta = \eta_c$ leads us to set $r_+ \approx C(\eta - \eta_c)$, for some constant $C$. Therefore,

$$\begin{pmatrix} \gamma_r \\ \gamma_i + \gamma_e \end{pmatrix} \sim C(\eta - \eta_c)\mathbf{v}_+ e^{\sigma_c(\tau - \tau_0)}, \tag{5.28}$$

at large times. That is, for $\eta$ near $\eta_c$, those pairs of $(\eta, \tau)$ with the same $(\eta - \eta_c)e^{\sigma_c\tau}$ should have the same $\gamma$. Although this property is derived from the local linearization about the fixed point, it still holds when trajectories have progressed further along the unstable manifolds of that saddle because the trajectories shadow that curve. This is illustrated in figure 6 for both the Hamiltonian system and the numerical solution, where the pairs of points plotted along sample orbits have the same values for $(\eta - \eta_c)e^{\sigma_c\tau}$, and therefore similar $\gamma$, even though they correspond to different choices of $(\eta, \tau)$. We can
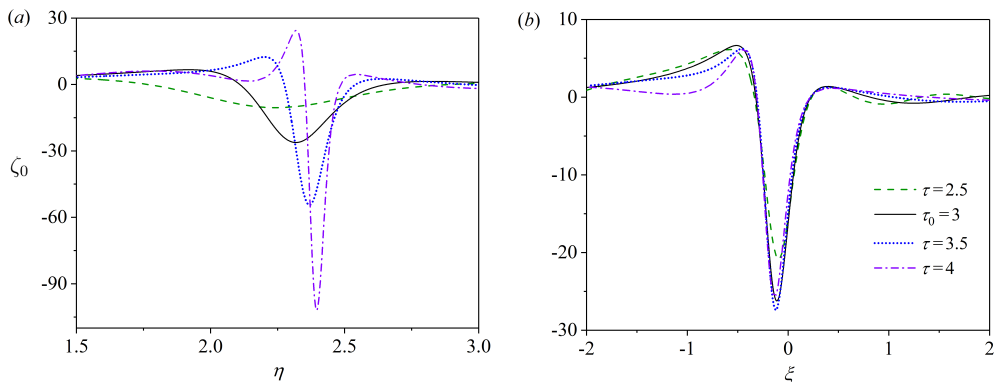
FIGURE 7. (a) Evolution of $\zeta_0$ near $\eta_c = 2.43$ at the times indicated. (b) Scaled profiles, $\zeta_0 e^{-\sigma_c(\tau-\tau_0)}$ against $\xi = (\eta - \eta_c)e^{\sigma_c(\tau-\tau_0)}$, choosing $\tau_0 = 3$; $(m = 1/2,\ f = 4/3,\ \mathcal{N} = 4/3)$.

express the property mathematically by writing the solutions in the self-similar form,

$$\gamma \approx F(\xi) \quad \text{and} \quad \zeta_0 \approx e^{\sigma_c(\tau-\tau_0)}\frac{\mathrm{d}}{\mathrm{d}\xi}|F(\xi)|^2, \quad \text{with} \quad \xi = (\eta-\eta_c)e^{\sigma_c(\tau-\tau_0)}, \quad (5.29)$$

for some function $F(\xi)$ related to the shape of the unstable manifolds of the saddle point. Thus, the lengthscale of the nonlinear critical layer at $\eta = \eta_c$ decreases exponentially in time, accounting for the relatively rapid focussing of sharp spatial variations in $\gamma$ at later times in figure 4, and the amplitude of the vertical vorticity grows exponentially. Figure 7 presents four snapshots of $\zeta_0(\eta, \tau)$ for the numerical solution, then replots them against $\xi$ and scaled by $e^{\sigma_c(\tau-\tau_0)}$, adopting $\tau_0 = 3$; while the profile of $\zeta_0$ keeps sharpening and strengthening, the rescaled profile remains nearly unchanged, confirming the self-similar structure in (5.29).

The exponential focussing towards the nonlinear critical level is problematic as it implies that the higher-order harmonics of the forcing pattern, which are neglected in our nonlinear critical layer model, grow faster than the re-arrangments of the mean flow. In particular, one can deduce that the vertical vorticity of the $j^{th}$ Fourier component, $\exp[j(\mathrm{i}x+\mathrm{i}mz)]$, grows like $e^{(j+1)\sigma_c\tau}$. The model therefore fails once the solution becomes overly focussed, heralding the onset of a further, more complicated, stage of evolution.

## 6. Effects of diffusion

The increasingly fine scales encountered in the critical layer due to the exponential focussing suggest that dissipation may also become prominent over later times, even if small initially. To explore this possibility in more detail, we return to the governing equations and include the viscous terms $\nu\nabla^2(u, v, w)$ in (2.1)–(2.3) and diffusive term $\kappa\nabla^2\rho$ in (2.4). We then take the distinguished limit $(\nu, \kappa) = O(\varepsilon^2)$, which corresponds to when dissipation first becomes important. In particular, with this sacling of $\nu$ and $\kappa$, the dissipative terms are too small to affect the quasi-steady wave in the bulk of the flow, but enter the analysis of the baroclinic critical layers owing to the reduced spatial scale in $y$. Equation (5.5) is now replaced with

$$\frac{\partial\rho_1}{\partial T} + \mathrm{i}Y\rho_1 + \frac{m\mathcal{N}}{2}A = -\mathrm{i}U_0\rho_1 + \frac{(\nu+\kappa)}{2\varepsilon^2}\frac{\partial^2\rho_1}{\partial Y^2}. \quad (6.1)$$

The Eulerian pseudomomentum is no longer equal to the mean-flow response, as in (5.11), and we have to return to the mean-flow evolution equation:

$$\frac{\partial U_0}{\partial T} = \frac{m}{\mathcal{N}^2}(A^*\rho_1 + A\rho_1^*) + \frac{\nu}{\varepsilon^2}\frac{\partial^2 U_0}{\partial Y^2}. \tag{6.2}$$

(following from the substitution of (5.6)-(5.8) into the modified version of (5.10)). The initial condition is still given by (5.20), the dissipative terms being negligible at early times when the spatial scales are larger. The closure relations given by the match to the outer solution remain (5.13) and (5.14). Equations (6.1) and (6.2) can be combined to furnish the integral relation,

$$\frac{\mathrm{d}}{\mathrm{d}T}\int_{-\infty}^{\infty}\left(|\rho_1|^2 + \tfrac{1}{2}\mathcal{N}^3 U_0\right)\mathrm{d}Y = -\frac{(\nu+\kappa)}{\varepsilon^2}\int_{-\infty}^{\infty}\left|\frac{\partial\rho_1}{\partial Y}\right|^2\mathrm{d}Y, \tag{6.3}$$

provided that $\rho_1$ and $U_0$ decay sufficiently quickly for $|Y| \to \infty$. We now briefly discuss the dynamics captured by this dissipative version of the model, focussing on the astrophysically relevant limit $\nu \ll \kappa$.

### 6.1. *Modified canonical system*

A scaling similar to that in §5.2, now furnishes the modified canonical system,

$$\frac{\partial\gamma}{\partial\tau} + \mathrm{i}\eta\gamma + A = -\mathrm{i}\gamma\mathcal{U} + \lambda\frac{\partial^2\gamma}{\partial\eta^2}, \qquad \frac{\partial\mathcal{U}}{\partial\tau} = A^*\gamma + A\gamma^* \tag{6.4}$$

and (5.17), where

$$\mathcal{U}(\eta,\tau) = \left(\frac{2\mathcal{N}}{m^2}\right)^{\frac{1}{3}}U_0 \qquad \text{and} \qquad \lambda = \frac{\kappa\mathcal{N}}{m^2\varepsilon^2}. \tag{6.5}$$

This system may be solved numerically. For the task, we now use a Crank-Nicolson method to evolve the system in time and centred finite differences method to evaluate spatial derivatives, exploiting Newton iteration at each time step to solve the nonlinear equations.

Before characterizing the features of the numerical solutions, we first pause to examine the dynamics in the limit that diffusion is relatively strong, $\lambda \gg 1$. In this limit, the large diffusive term $\lambda\gamma_{\eta\eta}$ in (6.4) must be balanced by introducing the rescalings, $(\gamma,\tau) = O(\lambda^{-1/3})$, $\eta = O(\lambda^{1/3})$ and $\mathcal{U} = O(\lambda^{-2/3})$. The advection of the density perturbation by the mean-flow correction, $\mathrm{i}\gamma\mathcal{U}$, is then small in the first equation in (6.4), and if we again make the approximation that $A$ is contant, we find

$$\gamma \approx -A\int_0^{\tau} e^{-\lambda q^3/3 - \mathrm{i}q\eta}\mathrm{d}q, \tag{6.6}$$

which is plotted in figure 8. At $\tau \ll 1$, (6.6) recovers the secular growth of the linear non-dissipative critical layer (*cf* (5.20)), but over longer times, this solution approaches a steady state, illustrating how diffusion is able to saturate that growth before nonlinearity (and the advective term $\mathrm{i}\gamma\mathcal{U}$) enters the fray. Figure 8 also illustrates how this dynamics does indeed characterize the full modified model for larger values of the diffusivity, demonstrating how the analytical solution in (6.6) agrees satisfyingly with numerical results computed with $\lambda = 5.3$. The steady state prediction from (6.6) corresponds to the result of viscous critical-layer theory presented by Boulanger, Meunier & Le Dizès (2007) for stratified tilted vortices (in which case, $\tau \to \infty$ in (6.6) and the solution can be related to the Scorer function).
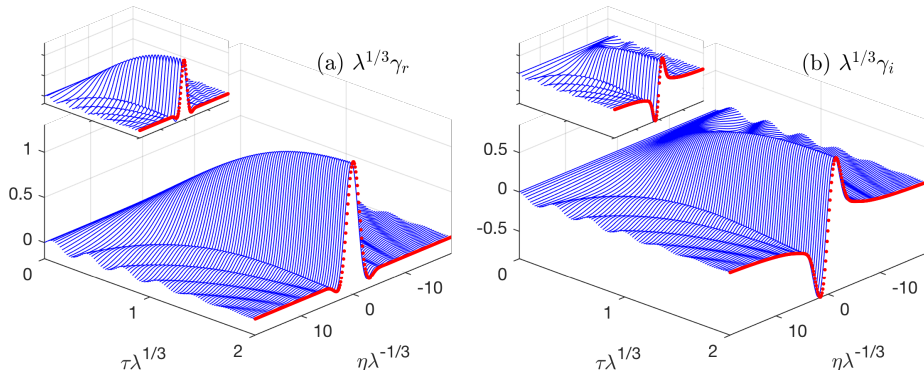
FIGURE 8. The analytical solution (6.6) for strong diffusion and $A = -1$, showing (a) $\lambda^{1/3}\gamma_r$ and (b) $\lambda^{1/3}\gamma_i$ against the scaled space and time variables $\lambda^{-1/3}\eta$ and $\lambda^{1/3}\tau$. The (red) dots show the final steady-state solution. The insets show corresponding numerical solutions to the reduced model, computed for $\lambda = 5.3$.

Nevertheless, the establishment of a steady state with spatial structure in the density perturbation is inconsistent with the integral relation in (6.3). Indeed, if $\gamma$ approaches a steady state, $\mathcal{U}$ continues to grow linearly with $\tau$, and for times of order $\lambda^{1/3}$, the advective term $\mathrm{i}\gamma\mathcal{U}$ can no longer be neglected in (6.4), heralding the onset of a different, more complicated phase of evolution. Figure 9 shows a suite of numerical solutions, illustrating this later evolutionary stage for cases with stronger diffusion (right-hand panels), and other examples with smaller $\lambda$ (left-hand panels). For the latter, diffusion is too weak to arrest the linear growth in the critical layer and nonlinear focussing begin to occur; only when the spatial scale has reduced sufficiently does the dissipative effect take hold to limit the exponential amplification found for $\lambda = 0$. At that stage, a new phase of evolution again emerges, much like that found for stronger diffusion. In particular, the oscillations of the non-dissipative dynamics begin to fade with time, and a localized coherent structure emerges that drifts to larger $\eta$ under the advective effect of the mean-flow correction. The structure leaves in its wake an increasingly strong deficit in $\mathcal{U}$, which is permitted by the constraint in (6.3) because diffusion may continually lower $\mathcal{U}$ as long as the gradients of $\gamma$ remain finite.

### 6.2. Dissipative coherent structures

The drifting coherent structure can be analyzed further owing to its fine spatial scale and the relatively slow timescale over which the system develops once the larger-scale transients have subsided: assuming that $\lambda \ll 1$ and $A$ is real and constant, we search for a quasi-steady travelling wave solution in which

$$\gamma \approx \gamma(\xi) \qquad \text{and} \qquad \xi = \frac{\eta}{\sqrt{\lambda}} - \int c\, \mathrm{d}\tau, \tag{6.7}$$

which characterizes a coherent structure with a length scale of $\sqrt{\lambda} \ll 1$ and a drift velocity given by $c$. Hence,

$$-c\gamma' + \mathrm{i}\eta_*\gamma + A \approx -\mathrm{i}\gamma\mathcal{U} + \gamma'' \qquad \text{and} \qquad -c\mathcal{U}' \approx A^*\gamma + A\gamma^* \approx 2A\gamma_r, \tag{6.8}$$

where $\xi = 0$, or $\eta_* = \sqrt{\lambda}\int c\, \mathrm{d}\tau$, prescribes the center of the coherent structure. This fifth-order system may be solved subject to the far-field constraints that $\gamma$ and $\mathcal{U}$ approach constant values as $|\xi| \to \infty$. In particular, since the coherent structure invades a region to the right in which $\gamma_r = \mathcal{U} = 0$, but $\mathcal{U}$ remains finite to the left (see figure 9), we
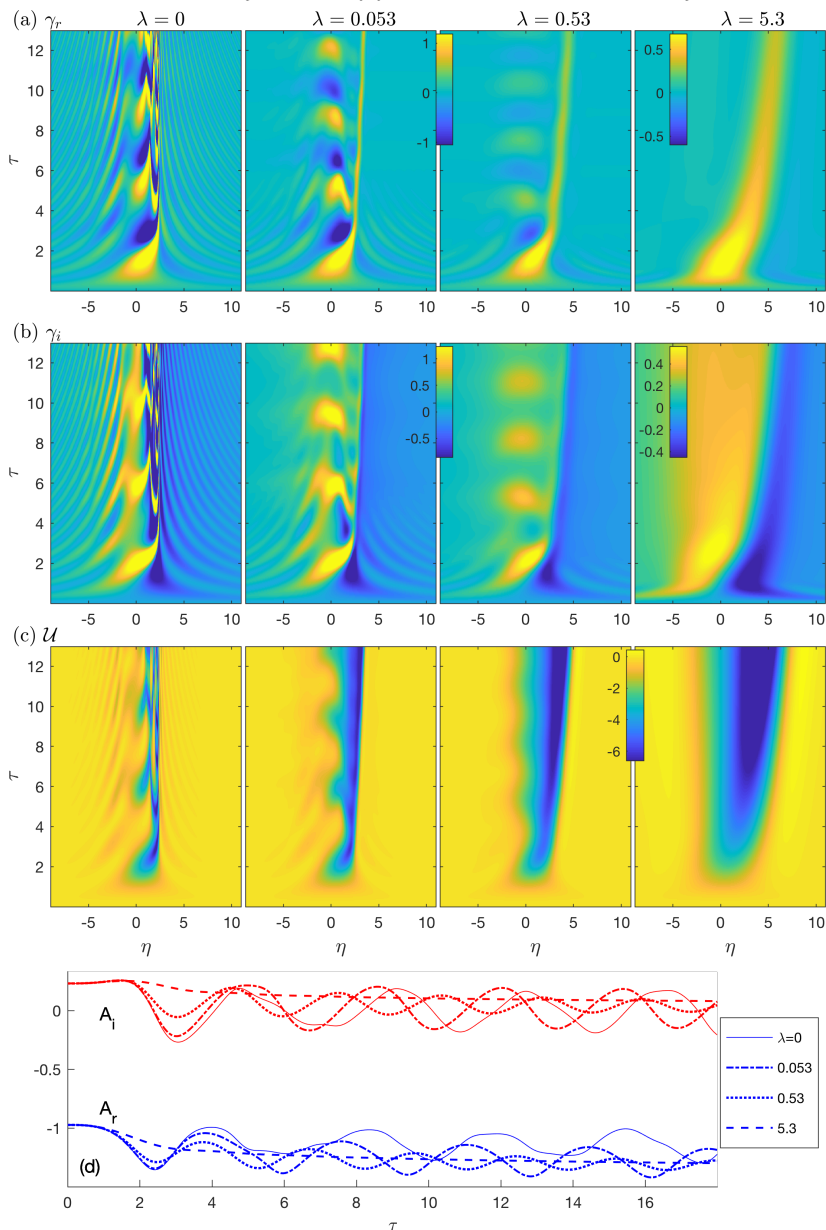
FIGURE 9. Solutions of the modified canonical model, showing (a) $\gamma_r$, (b) $\gamma_i$ and (c) $\mathcal{U}$, for $m = 1/2$, $f = 4/3$ and $\mathcal{N} = 4/3$, $c_1 = 0.238$, $c_2 = 0.219$ with the values of $\lambda$ indicated (and corresponding to the three columns). The colormap is the same in the first three panels of (a) and (b), but not the rightmost panel. The quasi-steady wave amplitude $A = A_r + \mathrm{i}A_i$ for the four computations is shown in (d).

demand the limits

$$(\gamma_r, \gamma_i, \mathcal{U}) \to \begin{cases} (0, G_+, 0) & \text{for } \xi \to \infty, \\ (0, G_-, \Delta\mathcal{U}) & \text{for } \xi \to -\infty, \end{cases} \tag{6.9}$$

where $G_+ = A\eta_*^{-1}$, $G_- = A(\eta_* + \Delta\mathcal{U})^{-1}$ and $\Delta\mathcal{U}$ is the jump in the mean flow across the structure. (6.9) imposes six boundary conditions to (6.8). One must also remove
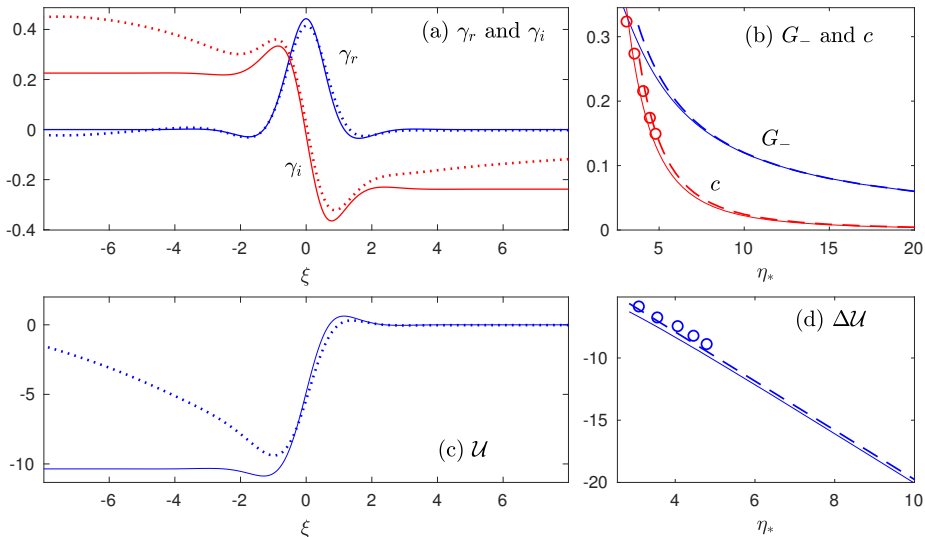
FIGURE 10. A coherent structure computed from (6.8) with $\eta_* = 5.04$ and $A = -1.2$, showing (a) $\gamma_r$ and $\gamma_i$, and then (c) $\mathcal{U}$ (solid lines). The dotted lines show the numerical solution of the modified canonical model (6.4), computed for $\lambda = 0.53$ at $\tau = 17.8$ (at which moment the residual oscillations near $\eta = 0$ are less pronounced). In (b) and (d) we show $G_-$, $c$ and the jump $\Delta\mathcal{U} = -[\mathcal{U}]_{-\infty}^{\infty}$ against $\eta_*$ from the solutions to (6.8) for $A = -1.2$ (solid lines). The dashed lines show the limiting behaviour for $\eta_* \gg 1$ given in (6.10). The circles show data for $c$ and $\Delta\mathcal{U}$ measured from the numerical solution of (6.4) with $\lambda = 0.53$ from $\tau = 5$ to $\tau = 15.4$.

the translational invariance of the system by imposing an additional constraint. Thus, given $\eta_*$, we solve (6.8) subject to those seven conditions, treating $G_-$ and $c$ as unknown parameters (eigenvalues). This furnishes localized structures taking the form of "pulses" in $\gamma_r$ and "fronts" in $\gamma_i$ and $\mathcal{U}$. Note that, as the coherent structure drifts to the right, $\eta_*$ increases, corresponding to an evolution of the coherent structure, which is treated parametrically in the quasi-steady approximation of (6.7) and (6.8).

Figure 10 shows a sample solution to (6.8) for $(A, \eta_*) = (-1.2, 5.04)$, giving $G_+ = -0.24$. These choices for $A$ and $\eta_*$ correspond to the numerical solution of the modified canonical model for $\lambda = 0.53$ shown in figure 9 at $\tau \approx 18$, and they are also plotted in figure 10. The solution to (6.8) compares satisfyingly with the snapshot of the simulations near the core of the coherent structure, although there are discrepancies further away arising from the influence of the far-field flow.

Figure 10 also includes data computed from (6.8) for $G_-$, $c$ and $\Delta\mathcal{U}$, as functions of $\eta_*$. In the limit of large $\eta_*$, a simple rescaling of (6.8) and (6.9)) indicates the limiting behaviour,

$$G_- \to G_+ = O(\eta_*^{-1}), \quad c = O(\eta_*^{-5/2}), \quad \Delta\mathcal{U} = O(\eta_*). \qquad (6.10)$$

The solution of (6.8) is compared to (6.10) together with measurements from the numerical simulation in the figure. Similarly, the characteristic strength and width of the structure are $\gamma = O(\eta_*^{-1})$, $\mathcal{U} = O(\eta_*)$ and $\xi = O(\eta_*^{-1/2})$. Thus, as the coherent structure drifts to the right, and $\eta_*$ slowly increases, the drift velocity declines, and the peak in $\gamma_r$ and jump in $\gamma_i$ must decrease and narrow. However, the jump in $\Delta\mathcal{U}$ continues to build up, predicting that the deficit in the mean flow grows linearly with $\eta$ for $\eta < \eta_*$.

This behaviour of the coherent structure rationalizes the dynamics of the modified canonical model seen in figure 9: once the linear dynamics and nonlinear focussing have subsided, the two features that remain are the decaying oscillations near $\eta = 0$

and the drifting coherent structure. The structure leaves in its wake a slowly diffusing density perturbation $\gamma \approx iG_-$ (see the right-hand plots in figure 9(b)) and a gradually strengthening mean flow correction $\Delta \mathcal{U}$, as seen on the right of figure 9(c). Thus, with diffusion, all growth in the density perturbation becomes arrested, leaving a widening and strengthening, jet-like defect in the mean flow.

One final concern is the impact of viscosity on the dynamics of the coherent structure: it is clear from (6.2) that the growth of the mean flow correction may be halted when $\nu = O(\varepsilon^2)$. Indeed, in the limit of stronger diffusion, the viscous term may allow $\mathcal{U}$ to also reach a steady state within the critical layer. However, as for the classical critical layers of Rossby waves (Brown & Stewartson 1978) and clear from the constraint (6.3), a genuine steady state is not possible with dissipation. Instead, the mean-flow correction must inevitably spread viscously out of the critical layer, even if a quasi-steady state is reached locally. Such considerations suggest that viscosity, if sufficiently strong, may prevent the creation of the drifting coherent structure, although a widening jet-like defect might still appear in the mean flow.

## 7. Discussion

In this paper, we have studied the non-dissipative, nonlinear dynamics of forced baroclinic critical layers using matched asymptotic expansion. In the linear regime, the forcing establishes a steady wave response outside the critical layers, but disturbances grow secularly inside the critical layer, which thins with time. The behavior is very similar to the forced critical layers of both Rossby and internal gravity waves (Stewartson 1978; Warn & Warn 1976, 1978; Booker & Bretherton 1967; Brown & Stewartson 1980). Continuing the analysis, we then studied the weakly nonlinear dynamics of the critical layer, finding that the adjustment of the mean flow provides the most important feedback on the growing disturbance there. Guided by the critical-layer scalings exposed by the weakly nonlinear analysis, we then derived a reduced model for the nonlinear critical layer. The numerical solution of the reduced model reveals a continued growth of the vertical vorticity as the disturbance is focussed exponentially quickly into a finer region within the critical layer. The focussing progresses uninterrupted until the reduced model breaks down.

Such pathological behaviour is quite different to that of the forced critical layer of a Rossby wave, where nonlinearity halts the secular linear growth and the mean vorticity distribution overturns into a distinctive cat's eye structure (Stewartson 1978; Warn & Warn 1978; Killworth & Mclntyre 1985). In that process, all the harmonics of the forcing pattern are excited to the same strength of the fundamental component. By contrast, in our nonlinear theory of the forced baroclinic critical layer, the adjustment to the mean flow arrests the linear growth and prompts the focussing of the vorticity before any of the higher harmonics become important. It is only once the strength and lengthscale of the focussed vorticity pass out of the asymptotic regime of our theory that the harmonics will appear. One important contributor to this feature is that the position of the baroclinic critical level itself is dictated the streamwise wavenumber. The critical level of the forcing does not therefore coincide with those of the harmonics. This filtering action weakens the impact of those harmonics within the baroclinic critical layer, leaving the adjustment the mean flow as the main nonlinearity.

The nonlinear structures developed in our forced baroclinic critical layers (jet-like defects in the mean velocity and dipolar stripes in the vorticity) may well be the analogues of features seen in the simulations of Marcus *et al.* (2013) and Wang (2016). Unlike in the reduced model, however, where these structures continue to focus, the mean flow

structures spawned in the simulations roll up into new vortices, providing part of the chain of events leading to self replication. Thus, our model likely misses important secondary instabilities. Indeed, Killworth & McIntyre (1985) and Haynes (1989) have shown that the nonlinear evolution of a forced Rossby wave can be susceptible to shorter-wavelength shear instabilities and generate "critical layer turbulence" along the filaments of vorticity wrapped around the main cat's eye (see also Balmforth & Korycansky (2001)). A roll up of the jet-like defects into new vortices seems plausible in the present case, and may arrest the uninterrupted focussing effect within the nonlinear critical layer. However, an extension of the matched asymptotic analysis is required to capture such dynamics.

Marcus *et al.* (2016) further argued that self replication is a finite-amplitude instability, requiring the amplitude of the initial disturbance to exceed a certain threshold. By contrast, the secular growth and nonlinear focussing of the disturbance inside the critical layer is triggered for an arbitrary small forcing amplitude in our analysis. Nevertheless, we have idealized the driving as a steady wavemaker, and ignored any possible evolution of that forcing. If the wavemaker cannot be sustained indefinitely, a threshold likely emerges that demands that the forcing act for sufficient time and strength to drive the baroclinic critical layers to the point where secondary instability can arise.

The continued focussing of the mean vorticity layer also indicates that dissipative effects are likely to become important in the later stages of evolution inside the baroclinic critical. Including the diffusion of density (*i.e.* heat or salt) in the theory leads to a modification of the reduced model, which now takes a partial differential form. A brief exploration of the modified model demonstrates that weak diffusion can arrest the focussing to the nonlinear critical level. Interestingly, a drifting solitary-wave like object then emerges, with a structure that can be analyzed analytically. The solitary wave leaves in its wake another jet-like defect in the mean flow, but this time the defect gradually widens and deepens as the object drifts.

In summary, when a steady forcing drives waves with baroclinic critical levels into a horizontally sheared flow with vertical stratification, the growing density perturbations predicted by linear theory become saturated by nonlinear effects. Although this saturation is demanded by the conservation laws of the governing equations, those constraints still permit the density perturbation to develop finer spatial structure over a region within the baroclinic critical layer. This nonlinear focussing effect takes place exponentially quicky, developing sharp jet-like defects in the mean flow, which can survive even in the presence of weak dissipation. This dynamics of the baroclinic critical layers is more destructuve than that for the classical critical layers of Rossby and internal waves, and plausibly rationalizes part of the cycle of vortex self replication observed by Marcus *et al.* in numerical simulations.

## Appendix A. The critical-layer vorticity distribution

The reconstruction of the critical-layer vorticity from the matched asymptotics is:

$$\zeta = \zeta_0 + \left[\varepsilon^{\frac{1}{3}}\zeta_1 e^{ix+imz} + \varepsilon^{\frac{2}{3}}\zeta_2 e^{2ix+2imz} + \text{c.c.}\right], \tag{A 1}$$

where $\zeta_0$ is given by (5.22),

$$\zeta_1 = \mathrm{i} \left( \frac{m^2}{2\mathcal{N}} \right)^{\frac{2}{3}} \left[ (f-1)\gamma + \gamma \frac{\partial |\gamma|^2}{\partial \eta} + \frac{2}{m^2} v_1 \frac{\partial^2 |\gamma|^2}{\partial \eta^2} \right], \qquad (\mathrm{A}\,2)$$

$$\zeta_2 = (f-1) \left( \frac{m^2}{2\mathcal{N}^4} \right)^{\frac{1}{3}} \left( \frac{1}{2} m^2 \gamma^2 - v_1 \gamma_\eta \right) + \frac{\mathrm{i}}{(2m\mathcal{N})^{\frac{2}{3}}} \frac{\partial}{\partial \eta} \left[ v_1 \zeta_1 + \left( \frac{m^4 \mathcal{N}}{4} \right)^{\frac{1}{3}} u_1 \gamma \right], \ (\mathrm{A}\,3)$$

and the leading-order fundamental components of the critical-layer horizontal velocity are

$$v_1 = \frac{\mathrm{i} m^2 A}{2} \left[ \log \left| \left( \frac{\varepsilon^2 m^2}{2\mathcal{N}} \right)^{\frac{1}{3}} \eta \right| + 1 \right] - \frac{m^2}{2} \fint_\eta^\infty \left( \gamma - \frac{\mathrm{i} A}{\eta'} \right) \mathrm{d}\eta' + \frac{\mathrm{i} A (\alpha \mathcal{N} - f)}{\mathcal{N}^2 - f(f-1)}, \quad (\mathrm{A}\,4)$$

$$u_1 = \frac{(f-1)v_1 - \mathrm{i} A + v_1 \zeta_0}{\mathrm{i}\mathcal{N}}, \qquad (\mathrm{A}\,5)$$

where the decoration on the integral sign implies principal value.

## REFERENCES

BADULIN, S. I., SHRIRA, V. I. & L. SH. TSIMRING 1985 The trapping and vertical focusing of internal waves in a pycnocline due to the horizontal inhomogeneities of density and currents. *J. Fluid Mech.* **158**, 199–218.

BALMFORTH, N. J. & KORYCANSKY, D. G. 2001 Non-linear dynamics of the corotation torque. *Mon. Not. Roy. Astron. Soc.* **326**, 833–851.

BARRANCO, J. A., PEI, S. & MARCUS, P. S. 2018 Zombie vortex instability. III. Persistence with nonuniform stratification and radiative damping. *Astrophys. J.* **869**, 127.

BASOVICH, A. YA. & TSIMRING, L. SH. 1984 Internal waves in a horizontally inhomogeneous flow. *J. Fluid Mech.* **142**, 233–249.

BÉLAND, M. 1976 Numerical study of the nonlinear Rossby wave critical level development in a barotropic zonal flow. *J. Atmos. Sci.* **33**, 2066–2078.

BOOKER, J. R. & BRETHERTON, F. P. 1967 The critical layer for internal gravity waves in a shear flow. *J. Fluid Mech.* **27**, 513–539.

BOULANGER, N., MEUNIER, P. & LE DIZÈS, S. 2007 Structure of a stratified tilted vortex. *J. Fluid Mech.* **583**, 443–458.

BRETHERTON F. P. 1966 The propagation of groups of internal gravity waves in a shear flow. *Quart. J. Roy. Met. Soc.* **92**, 466–480.

BROWN, S. N. & STEWARTSON, K. 1978 The evolution of the critical layer of a Rossby wave. Part II. *Geophys. Astrophys. Fluid Dyn.* **10**, 1–24.

BROWN, S. N. & STEWARTSON, K. 1980 On the nonlinear reflexion of a gravity wave at a critical level. Part 1. *J. Fluid Mech.* **100**, 577–595.

BROWN, S. N. & STEWARTSON, K. 1982a On the nonlinear reflection of a gravity wave at a critical level. Part 2. *J. Fluid Mech.* **115**, 217–230.

BROWN, S. N. & STEWARTSON, K. 1982b On the nonlinear reflection of a gravity wave at a critical level. Part 3. *J. Fluid Mech.* **115**, 231–250.

BÜHLER, O. 2014. *Waves and Mean Flows.* Cambridge University Press.

CASE, K. M. 1960 Stability of inviscid plane Couette flow. *Phys. Fluids* **3**, 143–148.

EDWARDS, N. R., & STAQUET, C. 2005 Focusing of an inertia-gravity wave packet by a baroclinic shear flow. *Dyn. Atmos. Oceans* **40**, 91–113.

EMANUEL, K. A. 1994 *Atmospheric convection.* Cambridge University Press.

HAYNES, P. H. 1989 The effect of barotropic instability on the nonlinear evolution of a Rossby-wave critical layer. *J. Fluid Mech.* **207**, 231–266.

KILLWORTH, P. D. & MCLNTYRE, M. E 1985 Do Rossby-wave critical layers absorb, reflect, or over-reflect? *J. Fluid Mech.* **161**, 449–492.

Lesur, G.R.J. & Latter, H. 2016 On the survival of zombie vortices in protoplanetary discs. *Mon. Not. Roy. Astron. Soc.* **462**, 4549–4554.

Marcus, P. S., Pei, S., Jiang, C-H. & Hassanzadeh P. 2013 Three-Dimensional Vortices Generated by Self-Replication in Stably Stratified Rotating Shear Flows. *Phys. Rev. Lett.* **111**, 084501.

Marcus, P. S., Pei, S., Jiang, C-H. & Barranco J. A. 2015 Zombie vortex instability. I. A purely hydrodynamic instability to resurrect the dead zones of protoplanetary disks. *Astrophys. J.* **808**, 87.

Marcus, P. S., Pei, S., Jiang, C-H. & Barranco J. A. 2016 Zombie Vortex Instability. II. Thresholds to Trigger Instability and the Properties of Zombie Turbulence in the Dead Zones of Protoplanetary Disks. *Astrophys. J.* **883**, 2.

Maslowe, S. A. 1986 Crtical layers in shear flows. *Ann. Rev. Fluid Mech.* **18**, 405–432.

Olbers, D. J. 1981 The propagation of internal waves in a geostrophic current. *J. Phys. Oceanogr* **11**, 1224–1233.

Staquet, C. & Huerre, G. 2002 On transport across a barotropic shear flow by breaking inertia-gravity waves. *Phys. Fluids* **14**, 1993–2006.

Stewartson, K. 1978 The evolution of the critical layer of a Rossby wave. *Geophys. Astrophys. Fluid Dyn.* **9**, 185–200.

Vanneste, J. & Yavneh, I. 2007 Unbalanced instabilities of rapidly rotating stratified shear flows. *J. Fluid Mech.* **584**, 373–396.

Wang, C. & Balmforth, N. J 2018 Strato-rotational instability without resonance. *J. Fluid Mech.* **846**, 815–833.

Wang, M. 2016 Baroclinic Critical Layers and Zombie Vortex Instability in Stratified Rotational Shear Flow. PhD thesis, University of California, Berkeley.

Warn, T. & Warn, H. 1978 The evolution of a nonlinear critical level. *Stud. Appl. Math* **59**, 37–71.

Warn, T. & Warn, H. 1976 On the development of a Rossby wave critical level. *J. Atmospheric Sci.* **33**, 2021–2024.

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