JfM PAPERS

Salt fingering staircases and the three-component Phillips effect

Paul Pružina[1],[†], **David W. Hughes**[1] **and Samuel S. Pegler**[1]

[1]School of Mathematics, University of Leeds, Leeds LS2 9JT, UK

Understanding the dynamics of staircases in salt fingering convection presents a long-standing theoretical challenge to fluid dynamicists. Although there has been significant progress, particularly through numerical simulations, there are a number of conflicting theoretical explanations as to the driving mechanism underlying staircase formation. The Phillips effect proposes that layering in stirred stratified flow is due to an antidiffusive process, and it has been suggested that this mechanism may also be responsible for salt fingering staircases. However, the details of this process, as well as mathematical models to predict the evolution and merger dynamics of staircases, have yet to be established. We generalise the theory of the Phillips effect to a three-component system (e.g. temperature, salinity, energy) and demonstrate a regularised nonlinear model of layering based on mixing length parametrisations. The model predicts both the inception of layering and its long-term evolution through mergers, while generalising, and remaining consistent with, previous results for double-diffusive layering based on flux ratios. Our model of salt fingering is formulated using spatial averaging processes, and closed by a mixing length parametrised in terms of the kinetic energy and the ratio of the temperature and salt gradients. The model predicts a layering instability for a bounded range of parameter values in the salt fingering regime. Nonlinear solutions show that an initially unstable linear buoyancy gradient develops into layers, which proceed to merge through a process of stronger interfaces growing at the expense of weaker ones. Our results indicate that these mergers are responsible for the characteristic increase of buoyancy flux through thermohaline staircases.

Key words: double diffusive convection, stratified flows

---

† Email address for correspondence: mmpep@leeds.ac.uk

968 A16-1

1. Introduction

The dynamics of thermohaline staircases in double-diffusive systems has interested oceanographers since they were first reported in the 1960s. Oceanographic observations, laboratory experiments and direct numerical simulations have all shown the existence of wide, evenly mixed layers separated by sharp interfaces, lasting for long times – of the order of months – with little change. Thermohaline convection is a specific example of double-diffusive convection (DDC), in which motions are driven in a stably stratified fluid as a result of the buoyancy depending on two independent scalar quantities that diffuse at different rates. In accordance with the oceanic (thermohaline) context, we refer to the faster diffusing component as 'temperature', and the slower as 'salinity'. DDC occurs in two different regimes, depending on which component of buoyancy provides the destabilising gradient. Salt fingering refers to a configuration in which the temperature gradient is stabilising and the salinity gradient destabilising. Diffusive convection, on the other hand, occurs when the temperature gradient is destabilising and the salinity gradient stabilising. In the oceans, staircases have been found in regions susceptible to both salt fingering (Tait & Howe 1968; Schmitt *et al.* 2005) and diffusive convection (Neal, Neshyba & Denner 1969; Timmermans *et al.* 2008); their existence is also widely documented, in both regimes, in numerical studies (e.g. Radko 2003; Rosenblum *et al.* 2011; Stellmach *et al.* 2011; Hughes & Brummell 2021). Despite their prevalence, the conditions for the development of staircases, as well as the dynamics of their long-term evolution and merger, are not well understood.

Several theories have been proposed for the driving mechanism behind layering, which are well documented in reviews by Merryfield (2000) and Radko (2013). An early hypothesis was that of collective instability (Stern 1969), in which growing salt fingers excite large-scale internal waves that overturn and generate a stepped structure. It has also been suggested that staircases are the long-time state of thermohaline intrusions (Zhurbas & Ozmidov 1984; Merryfield 2000), or that they are metastable equilibria of the system, requiring a finite amplitude perturbation from an initial linearly stable state (Veronis 1965; Stern & Turner 1969). Other models rely on heating a stably stratified fluid from below, with convective layers forming sequentially from the bottom upwards (Turner & Stommel 1964; Huppert & Linden 1979).

A further idea is that of Radko (2003), who proposed that the driving factor behind staircases is the result of an instability arising from variation of the ratio of the thermal to solutal fluxes. With $T(z, t)$ and $S(z, t)$ representing the horizontally averaged temperature and salinity fields, and subscripts of t and z denoting partial derivatives with respect to time and height, Radko (2003) models the two components contributing to the density by

$$T_t = f_z, \quad S_t = c_z, \qquad (1.1a,b)$$

where $f(R)$ is the temperature flux, and $c(R)$ is the salinity flux, dependent on the density ratio $R = T_z/S_z$; the flux ratio $\gamma(R)$ is defined by $\gamma = f/c$. The growth rate of perturbations is found to be positive if and only if $d\gamma/dR$ is negative. On the basis of numerical simulations, Radko (2003) argues that $\gamma(R)$ should be non-monotonic with a single minimum, so that there is an unstable range of R, but the instability is arrested in regions where $d\gamma/dR > 0$. Radko's model provides a helpful conceptual framework for relating the condition for instability in terms of properties of the buoyancy fluxes. However, it describes only the conditions for an initial linear instability, with a growth rate that diverges at infinite wavenumber. As identified by Radko (2019), such an ultraviolet catastrophe precludes the identification of a preferred wavelength or maximal growth rate, and prevents its use to study the dynamics of larger-scale layers. A more complex

regularised physical theory is required to model the full evolution from initial perturbation to staircase. In addition, the flow velocity is absent from the model, with the fluxes depending only on the density ratio R. To regularise the high-wavenumber instability, Radko (2019) proposed a model based on an asymptotic multiscale analysis, which leads to hyperdiffusion terms in the temperature and salinity equations, giving negative growth rates at high wavenumbers. Thus the flux–gradient model can be adapted to study the evolution beyond an initial instability to a large-scale staircase structure.

The formation of layers is antidiffusive, with up-gradient transport, representing behaviour contrary to the homogenisation that one might expect. Early work on layering by Phillips (1972) and Posmentier (1977) proposed a mechanism for the development of staircases in a stirred stratified fluid with a single component of buoyancy. They appealed directly to this antidiffusive property, based on the turbulent diffusion of buoyancy with a diffusion coefficient that could be negative. The buoyancy field is modelled by a one-dimensional turbulent diffusion equation

$$b_t = \frac{\partial}{\partial z} f(b_z), \tag{1.2}$$

where $f(b_z)$ is a flux function dependent on the buoyancy gradient b_z. A linear stability analysis shows that a uniform buoyancy gradient $b_z = g_0$ is unstable to perturbations if

$$\frac{\mathrm{d}f}{\mathrm{d}b_z}(g_0) < 0. \tag{1.3}$$

If (1.3) holds for a finite range of g_0, then in this unstable range, perturbations will grow. However, this is a prediction only from linear theory and, furthermore, leads to an ill-posed high-wavenumber instability with growth rate $s \sim m^2$ as wavenumber $m \to \infty$. Nonetheless, Posmentier (1977) proposed that perturbations do not remain unstable indefinitely, but stabilise once the gradient reaches a value outside the unstable region, evolving into a stepped structure.

To regularise the dynamics, Balmforth, Llewellyn Smith & Young (1998) (hereinafter BLY) coupled the buoyancy equation (1.2) to an energy equation, considering the system

$$g_t = f_{zz}, \tag{1.4}$$
$$e_t = (\kappa e_z)_z + p, \tag{1.5}$$

where $g(z, t)$ is the buoyancy gradient, $f(z, t)$ is some flux function, $e(z, t)$ is the turbulent kinetic energy, κ is a turbulent diffusion coefficient, and p is a general source (production) of energy. In the absence of double-diffusive effects, a parametrisation for p must include an energy source to drive the layering process. For linear instability in the system (1.4)–(1.5), the equivalent of the Phillips condition (1.3) is

$$\frac{\mathrm{d}f}{\mathrm{d}g} \equiv \frac{f_g p_e - f_e p_g}{p_e} < 0. \tag{1.6}$$

The high-wavenumber instability inherent to the Phillips model (1.2) is avoided by parametrisations such that $\mathrm{d}f/\mathrm{d}g < 0$ but $\partial f/\partial g > 0$. With this regularisation, (1.4)–(1.5) provide a complete model that can be used to analyse layer formation, evolution and merger in a stirred stratified fluid. A crucial aspect of the model is the dependence of the turbulent fluxes on a turbulent mixing length $l(g, e)$, used to close the system.

Models of the general form (1.4)–(1.5) have been used to study layering in several contexts. Malkov & Diamond (2019) produced a similar style of model to describe the

formation of potential vorticity staircases. Pružina, Hughes & Pegler (2022) extended the BLY model in its original context of stirred stratified layering to examine the influence of the boundary conditions on the layering process and long-term merger trends. We will refer to these as two-component models. However, to study DDC, the buoyancy field must be split into two independent components (e.g. temperature and salt). Hence to produce a BLY-style model for double-diffusion, a third equation must be added to account for the second component of buoyancy. Paparella & von Hardenberg (2014) adapted the BLY model to a double-diffusive context by arguing, on the basis of their numerical simulations (Paparella & von Hardenberg 2012), that the flux ratio γ remained constant, and hence that the evolution of the temperature and salinity fields could be investigated with a single equation for the total buoyancy. This assumption reduces their model to a two-component model similar to that of BLY. Motivated by the results of numerical simulations (Paparella & von Hardenberg 2012), the model of Paparella & von Hardenberg (2014) is forced by a constant up-gradient salt finger flux, with an eddy diffusivity term representing stirring due to 'clusters' of salt fingers. The constant salt finger flux has no effect on the buoyancy equation, but contributes a positive source in the energy equation. As such, while the underlying physics is different, the model of Paparella & von Hardenberg (2014) takes a very similar form to that of BLY, modelling salt fingering staircases with a forced stratified system with a single component of buoyancy. The model of Paparella & von Hardenberg (2014) was studied theoretically by Coclite, Paparella & Pellegrino (2018), who proved the existence of solutions, and discussed some of their properties.

It remains an open question whether layering is also possible without such forcing, instead including double-diffusion directly. In this case, a third equation is necessary that allows temperature and salinity to be described individually. There has been some limited use of three-component models to study $E \times B$ staircases in plasma drift–wave turbulence (Ashourvan & Diamond 2016, 2017; Guo *et al.* 2019). In these systems, the instability takes place in only two of the equations, so the modelling of instability again reduces to a form similar to the two-component BLY framework.

For a uniform buoyancy gradient to develop into a more complex layered structure, some energy input is necessary. BLY included an explicit source term to represent stirring, while Paparella & von Hardenberg (2014) included a constant background salt finger flux. However, several computational studies of DDC have shown that no such external energy input is necessary for staircases to form. Instead, the double-diffusive instability provides a mechanism for the transfer of potential energy into kinetic energy. This is true in both the salt fingering regime (e.g. Stellmach *et al.* 2011) and the diffusive convection regime (e.g. Rosenblum *et al.* 2011; Hughes & Brummell 2021). As such, we seek to formulate a model with no prescribed external forcing.

Ma & Peltier (2022) note that the values of the density ratio R found in observed oceanic staircases in the diffusive convection regime ($2 < 1/R < 7$) differ significantly from the values predicted by classical linear stability theory ($1 < 1/R < 1.14$) (e.g. Turner 1973). By contrast, in the salt fingering regime, the values of R in observed staircases match well with linear theory. On this basis, Ma & Peltier (2022) suggest that the instability in diffusive convection is not actually the driver behind diffusive staircases, instead proposing that the layering instability relies on external forcing from a background flow, with double-diffusive effects being important only in the regularisation and stabilisation of layers.

In the current work, we develop the first three-component turbulent mixing-length model of double-diffusive layering in terms of temperature, salinity and energy. We demonstrate its application to the analysis of layer evolution and merger dynamics.

Our starting point is the stability analysis for a general three-component system in terms of two independent components of density and the turbulent kinetic energy. We categorise the different modes of instability that can occur, and make comparisons with the previous theories of Phillips and BLY. The γ instability of Radko (2003) is shown to be mathematically equivalent to the Phillips instability in the case where the flux functions are parametrised in terms of the background density ratio $R = T_z/S_z$.

We present a new model for thermohaline staircases in turbulent flow, which is derived from the Boussinesq equations using a horizontal averaging process. We apply this model to the salt fingering regime of DDC. We analyse the linear stability of steady states, and demonstrate that the system is susceptible to the Phillips instability for a range of parameter values within the salt fingering regime. Numerical solutions of the model to long times indicate that staircases evolve via the 'B-merger' pattern described by Radko (2007), whereby strong interfaces grow at the expense of weaker ones. Each layer merger causes the buoyancy gradient in surviving interfaces to increase, and the buoyancy flux through the layers to increase, so the staircase has significantly higher buoyancy flux than the initial unlayered state.

The paper is arranged as follows. In § 2, we present a linear stability analysis of a general three-component system, thereby giving a general characteristic equation for growth rates in terms of vertical wavenumber. We analyse the solutions to the characteristic equation in the limit of small wavenumber in § 2.1, the limit of large wavenumber in § 2.2, and the conditions for marginal stability in § 2.3. We discuss the conditions for layering in § 2.4. In § 2.5, we demonstrate the relationship between Radko's γ instability and the Phillips effect. Section 3 describes our three-component double-diffusive model and discusses the mixing length on which it depends. In § 4, we apply the results of § 2 to our model, demonstrating the expected regions of instability, and discussing the effect of changing the parameters of the model. In § 5, we present long-term numerical solutions and discuss the behaviour of the buoyancy flux through layer mergers. We end in § 6 by summarising our key conclusions.

2. The three-component Phillips effect

To develop a theory for three-component models, we first investigate the linear stability properties of a general three-component system of the form

$$g_t = f_{zz}, \tag{2.1}$$

$$d_t = c_{zz}, \tag{2.2}$$

$$e_t = (\kappa e_z)_z + p. \tag{2.3}$$

Here, $g(z, t)$ and $d(z, t)$ are the independent components of the buoyancy gradient, with $f(g, d, e)$ and $c(g, d, e)$ their corresponding turbulent fluxes. To analyse the conditions for instability in this general system, we perform a linear stability analysis. We assume that (g_0, d_0, e_0) is a uniform steady state, such that $p(g_0, d_0, e_0) = 0$ and (g', d', e') is a small perturbation. From the chain rule, we can write the z derivative as

$$\frac{\partial}{\partial z} = \frac{\partial g}{\partial z} \frac{\partial}{\partial g} + \frac{\partial d}{\partial z} \frac{\partial}{\partial d} + \frac{\partial e}{\partial z} \frac{\partial}{\partial e}. \tag{2.4}$$

On applying (2.4), expanding $p(g, d, e)$ as a Taylor series, and neglecting terms quadratic in the perturbation quantities, we obtain the linear form of the general model:

$$g'_t \approx g'_{zz}f_g + d'_{zz}f_d + e'_{zz}f_e, \tag{2.5}$$

$$d'_t \approx g'_{zz}c_g + d'_{zz}c_d + e'_{zz}c_e, \tag{2.6}$$

$$e'_t \approx e'_{zz}\kappa + g'p_g + d'p_d + e'p_e, \tag{2.7}$$

where the partial derivatives f_g etc. are evaluated in the uniform steady state. On seeking solutions of the form $(g', d', e') \propto \exp(st + \mathrm{i}mz)$, where s is the growth rate and m the vertical wavenumber, the linearised forms of (2.1)–(2.3) may be expressed in matrix form as

$$\begin{pmatrix} s + m^2 f_g & m^2 f_d & m^2 f_e \\ m^2 c_g & s + m^2 c_d & m^2 c_e \\ -p_g & -p_d & s + m^2\kappa - p_e \end{pmatrix} \begin{pmatrix} g_1 \\ d_1 \\ e_1 \end{pmatrix} = 0. \tag{2.8}$$

Equation (2.8) has a non-trivial solution only if the determinant of the matrix is zero, leading to the characteristic equation

$$\begin{aligned} s^3 &+ s^2[m^2(f_g + c_d + \kappa) - p_e] \\ &+ s[m^4(f_g c_d - f_d c_g + \kappa f_g + \kappa c_d) + m^2(f_e p_g - f_g p_e + c_e p_d - c_d p_e)] \\ &+ m^6\kappa(f_g c_d - f_d c_g) + m^4(f_g c_e p_d - f_g c_d p_e + f_e c_d p_g - f_e c_g p_d + f_d c_g p_e - f_d c_e p_g) = 0, \end{aligned} \tag{2.9}$$

forming a cubic equation relating the growth rate to the vertical wavenumber.

We now explore the conditions for the existence of unstable wavenumbers ($\mathrm{Re}(s) > 0$). To gain an analytic foothold, we consider the asymptotic limits of small and large wavenumbers m, noting that a positive growth rate $\mathrm{Re}(s) > 0$ in either limit is a sufficient condition for instability.

2.1. *Instability at small wavenumbers*

When $m = 0$, corresponding to infinitely long spatial scales, the characteristic equation (2.9) reduces to

$$s^3 - p_e s^2 = 0, \tag{2.10}$$

giving one root $s = p_e$ and two zero roots. If

$$-p_e < 0, \tag{2.11}$$

then there is growth in the energy equation alone, without requiring interaction from the temperature equations. This energy mode instability was also possible theoretically in the two-component BLY formulation, but the parametrisation adopted by BLY produces $-p_e > 0$ everywhere.

To determine the stability of the two zero roots of (2.10), it is necessary to include higher-order terms. On taking the limit $m \to 0$, the dominant balance in (2.9) results from

$s = O(m^2)$, giving

$$s^2 + (F_g + C_d)m^2 s + (F_g C_d - F_d C_g)m^4 = 0, \qquad (2.12)$$

where we have adopted the following notations for simplicity:

$$F_g \equiv \frac{\mathrm{d}f}{\mathrm{d}g} = \frac{f_g p_e - f_e p_g}{p_e}, \qquad (2.13)$$

$$C_d \equiv \frac{\mathrm{d}c}{\mathrm{d}d} = \frac{c_d p_e - c_e p_d}{p_e}, \qquad (2.14)$$

$$F_d \equiv \frac{\mathrm{d}f}{\mathrm{d}d} = \frac{f_d p_e - f_e p_d}{p_e}, \qquad (2.15)$$

$$C_g \equiv \frac{\mathrm{d}c}{\mathrm{d}g} = \frac{c_g p_e - c_e p_g}{p_e}. \qquad (2.16)$$

Equation (2.12) has at least one root with positive real part if

$$\text{either} \quad F_g C_d - F_d C_g < 0 \quad \text{or} \quad F_g + C_d < 0. \qquad (2.17a,b)$$

If (2.17a) is satisfied, then there is exactly one positive root, implying that the state is unstable. If (2.17b) is satisfied, but not (2.17a), then there are two positive roots. Condition (2.17a) represents the direct equivalent of the Phillips effect in a three-component system; (2.17b) extends this to allow for an oscillatory instability.

These conditions can also be interpreted in a vector framework, which is helpful for generalising to an N-component system. Let $\boldsymbol{F}$ be the vector function

$$\boldsymbol{F}(\boldsymbol{G}) = \begin{pmatrix} f(g, d, e(g, d)) \\ c(g, d, e(g, d)) \end{pmatrix}, \qquad (2.18)$$

where $e(g, d)$ is defined implicitly via $p = 0$. The Jacobian of $\boldsymbol{F}$ with respect to $\boldsymbol{G} = (g, d)$ is then

$$\boldsymbol{J} = \begin{pmatrix} F_g & F_d \\ C_g & C_d \end{pmatrix}. \qquad (2.19)$$

Hence the conditions (2.17a,b) can be rewritten respectively as

$$\det(\boldsymbol{J}) < 0, \quad \text{tr}(\boldsymbol{J}) < 0. \qquad (2.20a,b)$$

Together, these conditions are equivalent to the single condition that there is instability if $\boldsymbol{J}$ has at least one negative eigenvalue. The same condition can be obtained by considering the system in the general form

$$\frac{\partial \boldsymbol{G}}{\partial t} = \frac{\partial^2}{\partial z^2} \boldsymbol{F}(\boldsymbol{G}, e(\boldsymbol{G})), \quad p(\boldsymbol{G}, e(\boldsymbol{G})) = 0. \qquad (2.21a,b)$$

Equations (2.21a,b) could be extended readily into a general N-dimensional system, producing instability if the Jacobian of $\boldsymbol{F}$ with respect to $\boldsymbol{G}$ has at least one negative eigenvalue.

2.2. *Instability at high wavenumbers*

For $m \to \infty$, the characteristic equation (2.9) simplifies at leading order to

$$s^3 + s^2 m^2 (f_g + c_d + \kappa) + s m^4 (f_g c_d - f_d c_g + \kappa f_g + \kappa c_d) + m^6 \kappa (f_g c_d - f_d c_g) = 0. \tag{2.22}$$

In this limit, all three solutions obey $s = O(m^2)$. There is at least one root s with positive real part if either the s-independent term $m^6 \kappa (f_g c_d - f_d c_g)$ is negative, or the characteristic equation has a stationary point with $s > 0$. Assuming that f_g, c_d and κ are all positive, in order to avoid the high-wavenumber instability of Phillips (1972), both of these conditions reduce to

$$f_g c_d - f_d c_g < 0, \tag{2.23}$$

thereby providing us with a criterion for the existence of an unstable large wavenumber. Note that if there is no energy equation, then $p \equiv 0$ and $F_g = f_g$, etc. In this special case, conditions (2.17a) and (2.23) are identical, and the growth rate of the Phillips instability satisfies $s \to \infty$ as $m \to \infty$. This demonstrates how the inclusion of the energy equation (2.3) regularises the instability at high wavenumbers.

We note that if (2.23) is satisfied and there is a high-wavenumber instability, then one option to regularise it is by the addition of hyperdiffusion terms. It is straightforward to show that adding $-A g_{zzzz}$ and $-B d_{zzzz}$ to (2.1) and (2.2) gives growth rates $s \sim -m^4$ as $m \to \infty$. For brevity, we omit the details here, as the parametrisations that we will use do not lead to a high-wavenumber instability.

2.3. *Condition for marginal stability*

To check for instability at intermediate wavenumbers, we consider the point of marginal stability. Setting $s = 0$ in (2.9), we find the wavenumbers for marginal stability m_* to be $m_* = 0$ and

$$m_* = \sqrt{\frac{p_e (F_g C_d - F_d C_g)}{\kappa (f_g c_d - f_d c_g)}}. \tag{2.24}$$

If m_* is real, then s is of one sign for $0 < m < m_*$, and the opposite sign for $m > m_*$. There is only one positive value for m_*, so as m is varied, s can change sign no more than once. Likewise, the critical wavenumber for oscillatory instability can be found by setting $\mathrm{Re}(s) = 0$ in (2.9). In this case, m_* is given by solutions to

$$\begin{aligned}
m^4 (f_g + c_d)(f_g c_d - f_d c_g + \kappa(f_g + c_d + \kappa)) \\
+ m^2 \left(\frac{1}{-p_e} \kappa(f_g + c_d + F_g + C_d) \right. \\
+ (f_g + c_d)^2 + \frac{1}{-p_e}(f_g f_e p_g + c_d c_e p_d + f_d c_e p_g + f_e c_g p_d) \Bigg) \\
+ (-p_e)(F_g C_d - F_d C_g) = 0.
\end{aligned} \tag{2.25}$$

Assuming that $f_g > 0$ and $c_d > 0$, in order to avoid the high-wavenumber instability of Phillips (1972), the only possibility for a positive real solution m_* to exist is if one of the previous instability conditions (2.11), (2.17), (2.23) is satisfied.

968 A16-8

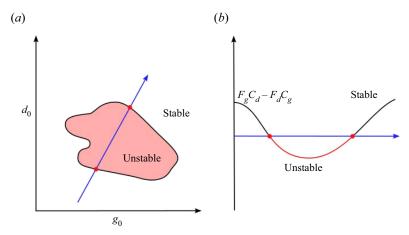(*a*)   (*b*)

Figure 1. (*a*) Sketch of a region of instability in g_0–d_0 space, shaded pink. The locus of marginal stability is shown in black. The blue line shows an arbitrary cross-section through the unstable region. (*b*) The value of $F_g C_d - F_d C_g$ along the blue path – it is negative only in the finite region between the two points shown in red, giving only a finite region where (2.17*a*) is satisfied.

Hence the only possible instabilities are via (2.11) (energy mode, $m \to 0$, m_* real), (2.17*a,b*) (small wavenumber Phillips instability, m_* real), or (2.23) (high wavenumber, m_* real). A combination of the conditions is possible, such that m_* is imaginary and there is a positive growth rate $\mathrm{Re}(s) > 0$ for all m.

2.4. *Conditions for layering*

We have determined the conditions for linear instability, but have not yet demonstrated how these conditions can lead to layering. BLY proposed that in a two-component model, layering requires an N-shaped relation between the buoyancy flux f and the buoyancy gradient g, so that $F_g < 0$ for only a finite range of gradients. The equivalent condition for a three-component model with independent contributions to the buoyancy is that there exits a bounded region of g–d space in which any one of the conditions for instability (2.11), (2.17) and (2.23) is met, as illustrated schematically in figure 1. On any path through the unstable region, only a finite range of points are unstable, with stable regions either side arresting the instability.

2.5. *Comparison with Radko's γ instability*

Radko (2003) put forward the idea that the driving factor behind layering is an instability arising from the parametric variation of the flux ratio γ as a function of the density ratio R. He modelled the two components of the density as

$$T_t = \frac{\partial}{\partial z} f(\gamma, Nu), \tag{2.26}$$

$$S_t = \frac{\partial}{\partial z} c(\gamma, Nu), \tag{2.27}$$

where the flux ratio $\gamma = f/c$ and the Nusselt number Nu is the ratio of convective to conductive heat transfer. The functions $\gamma(R)$ and $Nu(R)$ depend only on the density ratio $R = \alpha T_z / \beta S_z$, i.e. the ratio of the contributions to the density from temperature and salt.

Steady states of the flux–gradient relations (2.26)–(2.27) are found to be linearly unstable to perturbations when

$$\frac{d\gamma}{dR} < 0, \tag{2.28}$$

representing the γ instability.

To compare the criterion (2.28) with the conditions for instability that we have determined in the context of our general three-component model, we formulate our system in terms of Radko's parameters. Making Radko's assumption that the equations can be parametrised in terms of only the ratio of gradients R, we express the fluxes in terms of turbulent diffusivities $K_T(R)$ and $K_S(R)$ as

$$f = K_T(R)\, g, \tag{2.29}$$

$$c = K_S(R)\, d, \tag{2.30}$$

with $R = g/d$. Returning to the Jacobian form of the problem (2.19) with this new notation, we obtain

$$\boldsymbol{J} = \begin{pmatrix} K_T + K_T' R & -K_T' R^2 \\ K_S' & K_S - K_S' R \end{pmatrix}, \tag{2.31}$$

where primes denote differentiation with respect to R. Note that $\boldsymbol{J}$ does not depend on g or d independently, and the only parameter of importance is the density ratio R. The trace and determinant of $\boldsymbol{J}$ are given by

$$\mathrm{tr}(\boldsymbol{J}) = K_T + K_S + R(K_T' - K_S'), \tag{2.32}$$

$$\det(\boldsymbol{J}) = K_T K_S + R(K_T' K_S - K_T K_S'). \tag{2.33}$$

With fluxes of the form (2.29)–(2.30), the flux ratio $\gamma = R K_T / K_S$, so

$$\frac{d\gamma}{dR} = \frac{d}{dR}\left(R\,\frac{K_T}{K_S}\right) = \frac{K_T}{K_S} + \frac{R}{K_S^2}\left(K_T' K_S - K_T K_S'\right) = K_S^2 \det(\boldsymbol{J}). \tag{2.34}$$

The negative determinant condition (2.20) exactly recovers Radko's γ condition. The trace condition (2.20b) is new, and allows for two unstable modes.

To summarise, Radko's γ condition (2.28) is mathematically equivalent to the Phillips instability, in the specific context of double-diffusive flux–gradient relations dependent on R. The three-component model (2.1)–(2.3), with instability conditions (2.11), (2.17) and (2.23), describes a generalisation of both the Phillips and γ instabilities for a three-component system with explicit dependence on the kinetic energy e. The inclusion of e avoids the ultraviolet catastrophe inherent to the γ instability and the single-component Phillips instability, allowing the model to capture not only the initial growth of perturbations but also the possible development of layers and their long-term evolution. Note that for a system in the general form (2.1)–(2.3), condition (2.17a) can lead to instability by two different physical mechanisms. If the function p is parametrised to include an energy source term, then the system describes the forced mechanism of BLY and Paparella & von Hardenberg (2014). By contrast, with no source term in p, but appropriate parametrisations for f and c such that (2.17a) is satisfied, the instability comes from the γ-style instability analogous to that of Radko (2003).

3. A three-component model for thermohaline staircases

To develop our model for the formation and evolution of thermohaline staircases, we consider a domain of height h, with a background dimensional temperature gradient Θ_z, salinity gradient Σ_z and reference density ρ_0. The evolution of the velocity $\boldsymbol{u}(\boldsymbol{x}, t)$, temperature $T(\boldsymbol{x}, t)$ and salinity $S(\boldsymbol{x}, t)$ are governed by the Boussinesq equations

$$\boldsymbol{u}_t + \boldsymbol{u} \cdot \nabla \boldsymbol{u} = -\frac{1}{\rho_0} \nabla p + \frac{g\,(\rho - \rho_0)}{\rho_0}\, \boldsymbol{e}_z + \nu\, \nabla^2 \boldsymbol{u}, \tag{3.1}$$

$$T_t + \boldsymbol{u} \cdot \nabla T = \kappa_T\, \nabla^2 T, \tag{3.2}$$

$$S_t + \boldsymbol{u} \cdot \nabla S = \kappa_S\, \nabla^2 S, \tag{3.3}$$

$$\nabla \cdot \boldsymbol{u} = 0, \tag{3.4}$$

$$\frac{\rho - \rho_0}{\rho_0} = \beta S - \alpha T, \tag{3.5}$$

where $\rho(\boldsymbol{x}, t)$ is the density, and $p(\boldsymbol{x}, t)$ is the pressure. The equations depend on the kinematic viscosity ν, the thermal and solutal diffusivities κ_T and κ_S, gravitational acceleration g, and thermal and solutal expansion coefficients α and β.

We non-dimensionalise the system (3.1)–(3.5) via

$$\hat{t} = \frac{\kappa_T}{L^2}\, t, \quad \hat{z} = \frac{1}{L} z, \quad \hat{\boldsymbol{u}} = \frac{L}{\kappa_T}\, \boldsymbol{u}, \quad \hat{T} = \frac{\alpha g L^3}{\kappa_T \nu}\, T, \quad \hat{S} = \frac{\beta g L^3}{\kappa_T \nu}\, S, \quad \hat{p} = \frac{L^2}{\rho_0 \nu \kappa_T}\, p, \tag{3.6a–f}$$

with hats denoting dimensionless quantities. The characteristic length L is taken to be the salt finger length, corresponding to the length on which the magnitude of the local Rayleigh number is equal to unity (Stern 1960):

$$|\,\text{Ra}\,| = \frac{\alpha g\, |\Theta_z|\, L^4}{\kappa_T \nu} = 1. \tag{3.7}$$

With the choice of non-dimensionalisation (3.6a–f), the magnitudes of the dimensionless background temperature and salinity gradients are equal to the thermal and solutal Rayleigh numbers:

$$|\hat{\Theta}_z| = \frac{\alpha g\, |\Theta_z|\, L^4}{\kappa_T \nu} = |\,\text{Ra}\,| = 1, \tag{3.8}$$

$$|\hat{\Sigma}_z| = \frac{\beta g\, |\Sigma_z|\, L^4}{\kappa_T \nu} = |\,\text{Rs}\,| = \frac{1}{R_0}, \tag{3.9}$$

where R_0 is the density ratio. Dropping hats, we obtain the dimensionless form of the governing equations (3.1)–(3.5) as

$$\boldsymbol{u}_t + \boldsymbol{u} \cdot \nabla \boldsymbol{u} = -\sigma\, \nabla p + \sigma b \boldsymbol{e}_z + \sigma\, \nabla^2 \boldsymbol{u}, \tag{3.10}$$

$$T_t + \boldsymbol{u} \cdot \nabla T = \nabla^2 T, \tag{3.11}$$

$$S_t + \boldsymbol{u} \cdot \nabla S = \tau \nabla^2 S, \tag{3.12}$$

$$\nabla \cdot \boldsymbol{u} = 0, \tag{3.13}$$

$$b = T - S, \tag{3.14}$$

where $b(\boldsymbol{x}, t)$ is the non-dimensional buoyancy field. These equations depend on the diffusivity ratio $\tau = \kappa_S/\kappa_T$ and the Prandtl number $\sigma = \nu/\kappa_T$. The dimensionless height of the domain is $H = h/L$.

We now present a one-dimensional model for double-diffusive layering of the form (2.1)–(2.3), developed using a horizontal averaging process. Oceanic observations of double-diffusive staircases show a horizontal extent far greater than the thickness of the individual layers (e.g. Schmitt *et al.* 1987). As such, a horizontally averaged one-dimensional model is appropriate, providing insight to the physics of layering within a model that is relatively simple computationally. We apply the averaging process employed by Pružina *et al.* (2022), detailed in Appendix A, to obtain the following system:

$$T_t = \left(\frac{l^2 e}{le^{1/2} + 1} T_z \right)_z , \tag{3.15}$$

$$S_t = \left(\frac{l^2 e}{le^{1/2} + \tau} S_z \right)_z , \tag{3.16}$$

$$e_t = \left(\frac{l^2 e}{le^{1/2} + \sigma} e_z \right)_z - \sigma \left(\frac{l^2 e}{le^{1/2} + 1} T_z - \frac{l^2 e}{le^{1/2} + \tau} S_z \right) + \sigma e_{zz} - \epsilon \frac{e^{3/2}}{l}, \tag{3.17}$$

where T, S and e represent the horizontally averaged temperature, salinity and turbulent kinetic energy. The parameter ϵ controls the strength of the dissipation term. The system is closed by the mixing length l. The parametrisation of l is critical to the model, as it controls the form of the flux terms, and therefore the nature of any instability (cf. § 2).

It should be noted that this is a model of turbulent flow, relying on parametrisations of fluxes in terms of eddy diffusivities. As such, it should not be expected, and is not intended, to describe non-turbulent states. Numerical simulations (e.g. Stellmach *et al.* 2011) show that the salt fingering instability leads quickly to a highly turbulent state; although our mixing length model cannot capture the initial salt fingering stage of the evolution, it is nonetheless appropriate to describe all subsequent development, including the formation and evolution of staircases. Despite the fact that we are modelling a turbulent system, observations and numerical simulations show that the diffusivity ratio τ is critical to the evolution of double-diffusive staircases; hence it is important to keep it in the model. As such, the eddy diffusivities are not identical in each equation, as may be expected in a turbulence model. If $\tau = 1$, then there would be no difference between the two components of buoyancy, and layering would not be possible without an external forcing.

The aim of our model is to describe layering within a mixing-length framework, through the instability mechanism described in § 2, which we have shown to be mathematically equivalent to the γ instability of Radko (2003) in the case where the flux functions are parametrised in terms of the density ratio $R = T_z/S_z$. The γ instability theory depends on the parametrisation of fluxes specifically in terms of R, rather than in terms of the gradients individually. With this in mind, we propose to parametrise the mixing length l in terms of R. From a physical perspective, the length scale can be interpreted as a characteristic size of turbulent eddies, indicating that it should depend on e as well as R.

For a layered system, the mixing length must be parametrised such that l is small in the narrow, high-gradient interfaces, and larger in the wider, well-mixed layers. In salt fingering systems, it has been established, both numerically and experimentally, that the temperature and salinity fluxes have a decreasing dependence on the density ratio $R = T_z/S_z$ (e.g. McDougall & Taylor 1984; Kimura, Smyth & Kunze 2011). Furthermore, for the γ instability to be present, the flux ratio $\gamma = f/c$ has a decreasing

dependence on the density ratio $R = T_z/S_z$ in the equilibrium states. In our system (3.15)–(3.17), the temperature flux is $f = l^2 e T_z/(l e^{1/2} + 1)$ and the salt flux is $c = l^2 e S_z/(l e^{1/2} + \tau)$. From these forms, f, c and γ are all increasing functions of the length scale l; hence for $f(R)$, $c(R)$ and $\gamma(R)$ to be decreasing functions, $l(R)$ must also be decreasing. Mathematically, for a system of the form (2.1)–(2.3), a choice of l that depends on R alone leads to the high-wavenumber instability discussed in § 2, forming an ill-posed problem. Therefore, we parametrise l also to include a dependence on the local kinetic energy e. Based on these considerations, we adopt the parametrisation

$$l = \frac{(e^2 + \delta R^2)^{1/2}}{e^{1/2} R}, \tag{3.18}$$

where δ is a parameter chosen to be small, such that for $O(1)$ values of e and R, the mixing length $l \sim \sqrt{e}/R$. The value of δ must be non-zero, as otherwise $e = 0$ is a steady-state solution for all values of R_0, susceptible to the energy-mode instability (2.11). For values of R close to unity, the prescription (3.18) is a decreasing function of R, giving a large length when $R \approx 1$, and smaller lengths for larger values of R, as required. We note that while the model (3.15)–(3.17) was derived via averaging processes and physical arguments, by contrast there is not such a direct physical motivation for the exact form of the length scale. The prescription (3.18) is therefore not the only possible choice, but is, nonetheless, a simple parametrisation with the appropriate qualitative dependence on e and R to model layers.

It should be noted that in DDC described by the Boussinesq equations (3.1)–(3.5), the buoyancy anomalies leading to convective motions appear at or below the salt fingering length scale ($l = 1$ in our non-dimensionalisation), while turbulent mixing is expected to occur on scales larger than this. Our turbulent model cannot describe the small-scale dynamics directly, and instead assumes that turbulent motion continues down to the smallest scales. The length l measures the scale of the turbulent motion, rather than the layers themselves, so that in these strongly convective regions, we expect l to represent the size of a turbulent eddy, which may be significantly smaller than the full layer depth.

At this stage, it is helpful to review the parameter values for which the two different regimes of DDC occur. In the salt fingering regime, both temperature and salt gradients are positive. For the fluid to be statically stable, it is required that $b_z = 1 - 1/R_0 > 0$, and hence that the background density ratio

$$R_0 > 1. \tag{3.19}$$

For diffusive convection, both the temperature and salinity gradients are negative. The overall buoyancy gradient is $b_z = -1 + 1/R_0$, so for the fluid to be statically stable, it is required that

$$0 < R_0 < 1. \tag{3.20}$$

The system (3.15)–(3.18) depends on four dimensionless parameters: τ, σ, δ and ϵ. The first two are material parameters of the fluid under consideration. For example, salt water has diffusivity ratio $\tau \approx 0.01$ and Prandtl number $\sigma = O(10)$, depending on the temperature and salinity. In this study, we focus primarily on illustrating the results of our model for the case of salt water, but we note that the choice of τ and σ can be adapted to model other physical contexts. The other two parameters, δ and ϵ, are empirical modelling parameters introduced in the derivation to control the relative importance of turbulent dissipation and the form of the mixing length (3.18) (the parameter ϵ also appears in the model of BLY). We will determine values of δ and ϵ that lead to physically realistic behaviour in the solutions.

4. Steady states and their stability

To analyse the stability of the system (3.15)–(3.17), we begin by applying the general linear stability theory of three-component systems developed in § 2. Equations (3.15)–(3.17) may be expressed in the form (2.1)–(2.3) by writing

$$f = \frac{l^2 e}{l e^{1/2} + 1} g, \tag{4.1}$$

$$c = \frac{l^2 e}{l e^{1/2} + \tau} d, \tag{4.2}$$

$$\kappa = \frac{l^2 e}{l e^{1/2} + \sigma} + \sigma, \tag{4.3}$$

$$p = -\sigma \left(\frac{l^2 e}{l e^{1/2} + 1} g - \frac{l^2 e}{l e^{1/2} + \tau} d \right) - \epsilon \frac{e^{3/2}}{l}, \tag{4.4}$$

where $g = T_z$ and $d = S_z$. For a given value of R_0, the system admits the uniform steady state $(g_0, d_0, e_0) = (\pm 1, \pm 1/R_0, e_0(R_0))$.

4.1. *Steady states*

With the length scale prescribed by (3.18), the steady-state equation $p = 0$ for $e_0(R_0)$ leads to the following algebraic relation between e_0 and the parameter R_0:

$$g_0(R_0 - 1)(e_0^2 + \delta R_0^2)^2 + g_0(\tau R_0 - 1)(e_0^2 + \delta R_0^2)^{3/2} R_0$$
$$+ \frac{\epsilon}{\sigma} R_0^3 e_0^2 (e_0^2 + \delta R_0^2) + (1 + \tau) \frac{\epsilon}{\sigma} (e_0^2 + \delta R_0^2)^{1/2} e_0^2 + \frac{\epsilon}{\sigma} R_0^5 \tau e_0^2 = 0, \tag{4.5}$$

where $g_0 = \pm 1$. We interpret the uniform-gradient steady state physically as a representation of the flow resulting from salt fingers. Individual fingers cannot be distinguished, but a mean fluid motion is supported by one stable and one unstable component of the buoyancy gradient.

Figure 2(*a*) shows e_0 as a function of R_0. When there is no overall stratification ($R_0 = 1$), the energy $e_0 \approx \sigma/\epsilon - 1$ (shown by the blue dotted line in the inset plot). As R_0 increases, e_0 decreases, reaching $e_0 = 0$ at $R_0 = (1 + \sqrt{\delta})/(\tau + \sqrt{\delta})$. However, it is important to note that such an unphysical 'zero energy turbulent state' is precluded in our time-dependent model. To see this, we consider the evolution of a state that begins with $e > 0$ for all z. At a given time, let $z = z_*$ denote the position of a local minimum of e, with $e_z(z_*) = 0$, $e_{zz}(z_*) > 0$, and $e(z_*)$ near zero. Additionally, we note from (3.18) that $l e^{1/2} \to \sqrt{\delta}$ as $e \to 0$. After some algebra, the governing energy equation (3.17) at $z = z_*$ reduces to

$$e_t(z_*) \sim \left(\frac{\delta}{\sqrt{\delta} + \sigma} + \sigma \right) e_{zz} - \sigma \left(\frac{\delta T_z}{\sqrt{\delta} + 1} - \frac{\delta S_z}{\sqrt{\delta} + \tau} \right). \tag{4.6}$$

Provided that $R_0 < (\sqrt{\delta} + 1)/(\sqrt{\delta} + \tau)$ (i.e. R_0 is in the range where uniform steady states exist), every term on the right-hand side of (4.6) is positive. It follows that $e_t(z_*) > 0$, implying that the minimum energy cannot decrease, precluding the energy from ever reaching zero. Hence, while zero energy and negative energy states can exist within the equations, they will never be attained from initial conditions starting with positive energy. Nonetheless, the change of sign of e_0 when $R_0 > (\sqrt{\delta} + 1)/(\sqrt{\delta} + \tau)$ means that this model is applicable only for sufficiently low density ratios.
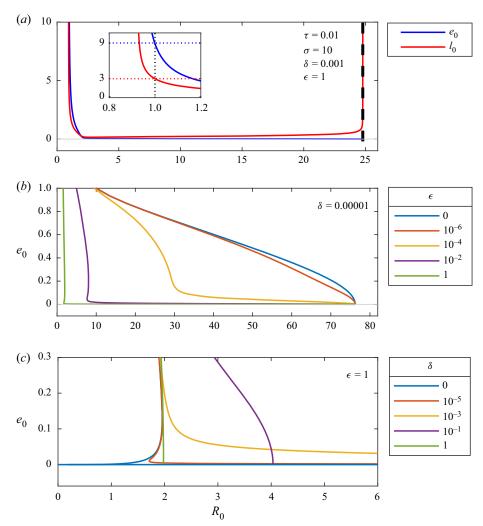
968 A16-14

Figure 2. Steady-state solutions to (4.5) with $\tau = 0.01$, $\sigma = 10$. (a) Energy e_0 and corresponding value of the length scale l_0 given by (3.18), as functions of R_0 for $\epsilon = 1$, $\delta = 0.001$. For R_0 small, e_0 and l_0 are large; as R_0 increases, e_0 decreases, with l_0 initially decreasing but $l_0 \to \infty$ as $e_0 \to 0$. Inset plot shows behaviour of e_0 and l_0 near $R_0 = 1$, with red and blue dotted lines showing the values $e_0 = \sigma/\epsilon - 1 = 9$ and $l_0 = \sqrt{\sigma/\epsilon - 1} = 3$. (b) Plot of e_0 as a function of R_0, for a range of values of δ with $\epsilon = 1$ fixed. (c) Plot of e_0 as a function of R_0, for a range of values of ϵ with $\delta = 0.001$ fixed. Sufficiently small values of δ and large values of ϵ lead to $e_0(R_0)$ being multi-valued.

Figure 2(a) also shows the value of the length scale l_0 calculated from the energy using (3.18). As $R_0 \to 1$, $l_0 \sim \sqrt{e_0} \approx \sqrt{\sigma/\epsilon - 1}$, which is shown by the red dotted line in the inset. As R_0 increases from 1, the length scale initially decreases, reaching a minimum, before increasing, with $l_0 \to \infty$ and $e_0 \to 0$ as $R_0 \to (\sqrt{\delta} + 1)/(\sqrt{\delta} + \tau)$. However, we have shown that the zero energy state is never reached, and hence this divergence in the mixing length will never occur.

The relationship between e_0 and R_0 given by (4.5) is shown for various choices of the parameters ϵ and δ in figures 2(b,c). The results illustrate a strong dependence of e_0 on both ϵ and δ, with smaller values of δ and larger values of ϵ causing e_0 to be

multi-valued for some range of R_0. For example, the solution shown in purple in figure 2(*b*) is multi-valued near $R_0 \approx 8$. In cases where there are multiple steady-state energies, one steady state is unstable to the energy mode, leading to growth in energy on the domain scale. To investigate layering processes, we wish to avoid this situation, so we set $\epsilon = 1$ and $\delta = 0.001$. These are the parameters used in figure 2(*a*), where e_0 is clearly single-valued throughout.

In the diffusive convection regime, $T_z, S_z > 0$ and $0 < R_0 < 1$. Within these ranges, all the terms in (4.5) are negative, and hence there are no positive solutions for e_0 in the diffusive convection regime. This result holds in general for any system of the form (4.1)–(4.4), no matter what parametrisations are adopted for the length scale and dissipation term. As such, it appears that an unforced system of this form is not sufficient to model layering in diffusive convection, lending weight to the proposition of Ma & Peltier (2022) that external forcing may be necessary. In this paper, we will restrict our focus to the salt fingering regime, with the modelling of layering in diffusive convection providing an interesting problem for future work.

4.2. *Linear stability*

The stability of the steady states of (3.15)–(3.17) can be analysed using the framework described in § 2. For a range of values of R_0, we first calculate $e_0(R_0)$ and substitute the value into the expressions for $-p_e$, $F_g C_d - F_d C_g$, $F_g + C_d$ and $f_g c_d - f_d c_g$; these quantities are plotted as functions of R_0 in figure 3(*a*). For a finite range of R_0 (between the red dots), $F_g C_d - F_d C_g < 0$, thereby satisfying the condition for the Phillips instability. By comparison with the schematic in figure 1(*b*), we see that our expectation of a finite unstable region is met. Note that by our choice of non-dimensionalisation, $|\text{Ra}| = 1$, so varying R_0 in figure 3(*a*) represents a single path through the unstable region of Ra–Rs space. The values of ϵ and δ are chosen so that the energy mode is stable (see (2.11)), and neither (2.17*b*) nor (2.23) is met. Figure 3(*b*) shows a plot of growth rate against wavenumber for the case $R_0 = 1.8$ – there is a single unstable mode, with a uniquely defined wavenumber of maximum growth rate, which can be used to predict the width of the fastest growing perturbations, and hence the width of the initial layers formed.

Recall from the end of § 2 that in a system of the general form (2.1)–(2.3), the layering instability (condition (2.17*a*)) may be caused by either the forcing mechanism of BLY and Paparella & von Hardenberg (2014), or the γ instability of Radko (2003). With the specific system (3.15)–(3.17), this is no longer the case. For a model with no source term in the energy equation (3.17), the γ mechanism is the only one in play.

We now investigate the effect on the stability of the system of varying the values of the material parameters, while fixing $\delta = 0.001$ and $\epsilon = 1$. Figure 4 shows the effect on the critical values of R_0 for instability of changing τ and σ independently. The black line in figure 4(*a*) shows the minimum and maximum values of R_0 for which instability occurs, as τ is varied with all other parameters kept fixed. The critical value of τ at the tip of the curve is $\tau_c = 0.1055$. This critical value is independent of σ, although larger values of σ lead to larger unstable ranges of R_0. Figure 4(*b*) shows the effect of varying σ on the critical values of R_0. For $\sigma \ll 1$, only a very narrow range of R_0 leads to instability; at larger values of R_0, this range increases significantly, saturating at approximately $2 \lesssim R_0 \lesssim 14$ for large σ. The dashed line shows the lower boundary of the fingering regime, at $R_0 = 1$. Larger values of τ reduce the size of the unstable range of R_0, but there is little qualitative change. For small σ, the entire unstable range lies below $R_0 = 1$, and is therefore not in the salt fingering regime. This result is consistent with
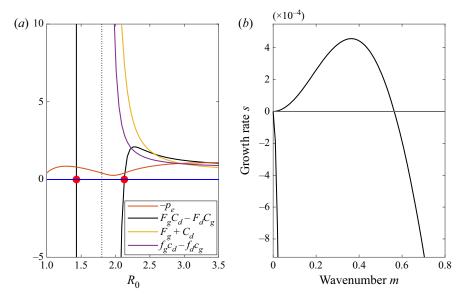
Figure 3. (a) Various stability measures of the uniform steady state as R_0 varies, with $\tau = 0.01$, $\sigma = 10$, $\epsilon = 1$, $\delta = 0.001$. The quantity $-p_e$ is shown in red, $F_g C_d - F_d C_g$ in black, $F_g + C_d$ in yellow, and $f_g c_d - f_d c_g$ in purple. The red circles mark the minimum and maximum values of R_0 for which conditions (2.17a,b) are satisfied. (b) The growth rate s against wavenumber m for the steady state with $R_0 = 1.8$ (marked with a dotted black line in a). There is a single unstable mode with maximum growth rate $s = 4.6 \times 10^{-4}$ at wavenumber $m = 0.363$.
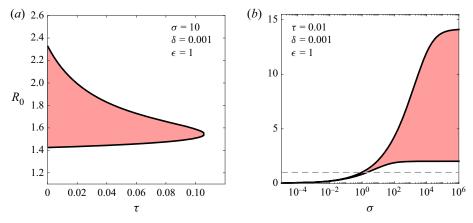

Figure 4. Effect on the layering instability of changing (a) the diffusivity ratio τ, and (b) the Prandtl number σ. Black lines show the boundary of the unstable range of R_0, as (a) τ is changed with fixed $\sigma = 10$, and (b) σ is changed for fixed $\tau = 0.01$. In each case, changing the value of the other parameter leads to no qualitative differences. The other model parameter values are $\delta = 0.001$ and $\epsilon = 1$; for these choices of δ and ϵ, there is no energy mode instability. The dashed line in (b) shows $R_0 = 1$, the lower boundary of the salt fingering regime.

those of Traxler, Garaud & Stellmach (2011), who studied salt fingering at low Prandtl number using three-dimensional numerical simulations. Traxler *et al.* (2011) found that the empirical flux ratio γ increased monotonically with density ratio R, so layering by the

968 A16-17

γ instability was not expected at small σ. Instead, it was suggested that any layering was due to the collective instability of Stern (1969).

5. Nonlinear evolution and long-term merger behaviour

A key advantage of the regularisation inherent in our mixing-length formulation (3.15)–(3.17) is that it allows the investigation of long-term dynamics of staircase evolution and merger beyond any initial instability. To investigate this long-term behaviour, we solve the system (3.15)–(3.17) with length scale (3.18) subject to the boundary conditions

$$T(0) = 0, \quad T(H) = H, \tag{5.1a,b}$$

$$S(0) = 0, \quad S(H) = H/R_0, \tag{5.2a,b}$$

$$e_z(0) = 0, \quad e_z(H) = 0, \tag{5.3a,b}$$

and initial conditions

$$T = z - g' \sin\left(\frac{2n\pi z}{H}\right), \tag{5.4}$$

$$S = \frac{z}{R_0} - d' \sin\left(\frac{2n\pi z}{H}\right), \tag{5.5}$$

$$e = e_0 - e' \frac{2n\pi}{H} \cos\left(\frac{2n\pi z}{H}\right), \tag{5.6}$$

where the perturbation amplitudes (g', d', e') are chosen such that the initial condition is an eigenstate of the linear stability problem, with $2n\pi/H$ chosen to be the maximally unstable wavenumber based on the linear theory of § 4. We solve the model numerically using the MATLAB pdepe solver, with 4000 spatial mesh points across a domain of depth 500, which is sufficient for well-resolved solutions. The solver varies time steps dynamically to ensure adequate resolution.

5.1. *Nonlinear evolution*

Figure 5 shows the results of a numerical solution of (3.15)–(3.17), for the same parameters as used in figure 3(*b*). Figure 5(*a*) shows the buoyancy field $b = T - S$ plotted over a range of times. The plot illustrates the evolution from an initially uniform gradient to a layered staircase; the layers then proceed to undergo mergers over time. Eventually, at $t \approx 2 \times 10^6$, only a single interface remains, located at $z \approx 350$. Figure 5(*b*) shows the normalised buoyancy gradient at the same time points. Interfaces between layers are represented by sharp spikes in the gradient profile; these profiles reveal the fine structure at early times that is not visible in the overall temperature field in figure 5(*a*). The range in the buoyancy gradient, i.e. $\max(b_z) - \min(b_z)$, is shown in figure 5(*c*), measuring the difference between the gradients in the interfaces and the layers. There is a clear gradual increase in the maximum gradient, beginning at $t \approx 6 \times 10^5$ and ending at $t \approx 2 \times 10^6$, defining the range of times over which mergers occur. During a layer merger, the overall buoyancy variation across a region must be conserved, resulting in sharper interfaces with higher gradients after each merger. Referring to the unstable region shown in figure 3(*a*), there is no constraint individually on T_z and S_z, provided that R_0 stays within the bounds of the unstable region. Hence b_z can become arbitrarily large, resulting in ever steeper interfaces as successive merger events take place. This contrasts with results obtained
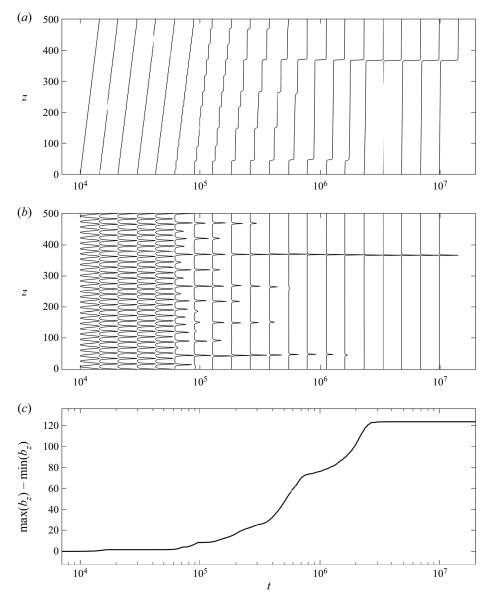
968 A16-18

Figure 5. Nonlinear evolution of the system (3.15)–(3.17), with length scale (3.18), subject to boundary conditions (5.1a,b)–(5.3a,b) and initial conditions (5.4)–(5.6), for parameter values $\tau = 0.01$, $\sigma = 10$, $\delta = 0.001$, $\epsilon = 1$, $R_0 = 1.8$. (a) Depth profiles of the overall buoyancy field $b = T - S$ at a range of times distributed logarithmically between $t = 10^4$ and $t = 10^7$. (b) Profiles of the buoyancy gradient $b_z = T_z - S_z$ scaled by the range of its values at each time, plotted at the same times as in (a). (c) Range of gradients, i.e. $\max(b_z) - \min(b_z)$. The solution evolves from the initial condition into a dense stack of layers (seen as the first solution presented in b). At $t \approx 6 \times 10^5$, the layers begin to undergo mergers, which cause the maximum gradient to increase, until by $t \approx 2 \times 10^6$, only a single interface remains at $z \approx 350$, with $b_z \approx 120$.

for related models of stirred, one-component layering (BLY; Pružina *et al.* 2022), where a well-defined maximum unstable gradient is determined, such that subsequent mergers produce thicker interfaces of fixed gradient. The long-term evolution of merger behaviour follows the same inverse logarithmic trend identified by Pružina *et al.* (2022), with the number of remaining interfaces N obeying the scaling $1/N \sim \log(t)$ as $t \to \infty$.

5.2. *Merger dynamics*

To understand the merging behaviour in more detail, we consider the results in the context of the analysis of Radko (2007), who identified two types of mergers: the B-merger, in which relatively strong interfaces grow at the expense of weaker neighbouring interfaces, and the H-merger, where neighbouring interfaces drift and collide. By considering a one-dimensional buoyancy conservation equation in a stepped basic state, and analysing the variation of the buoyancy flux across a step, Radko demonstrated the so-called merger theorem, showing that the B-merger has growth rate $\lambda_B \propto -\partial \tilde{F}/\partial \tilde{B}$, and the H-merger has growth rate $\lambda_H \propto \partial \tilde{F}/\partial \tilde{H}$, where $\tilde{F}(\tilde{B}, \tilde{H})$ is the buoyancy flux across a step, $\tilde{B}$ is the buoyancy jump, and $\tilde{H}$ is the height of the step.

To apply the merger theorem to the BLY model, Radko (2007) considered constant flux solutions, for which the model can be reduced to a nonlinear oscillator equation for $e(b)$. To apply the same analysis to our model, we adopt the same approach. Thus we seek steady-state solutions to the system (3.15)–(3.17) with uniform temperature and salinity fluxes f_0 and c_0. The system can then be reduced to a single equation (see Appendix B), given by

$$\frac{1}{2} e_T^2 + U(D) = E \left(\frac{D^2}{\kappa f_0 \, (D+1)} \right)^2, \tag{5.7}$$

where $D(e) = le^{1/2}$, the potential $U(D)$ is defined by

$$U(D) = \left(\frac{D^2}{\kappa f_0 \, (D+1)} \right)^2 \int \left(-\sigma \, (f_0 - c_0) - \epsilon \, \frac{e(D)^2}{D} \right) \kappa \, \mathrm{d}D, \tag{5.8}$$

and $e(D)$ is defined by

$$e(D) = \frac{\sqrt{D^2 - \delta} \, (D+1)}{D + \tau} \gamma_0. \tag{5.9}$$

Equation (5.7) represents a nonlinear oscillator for e as a function of temperature T, with variable weight. By inverting (5.9) for $D(e)$, (5.7) is transformed into an equation for e_T in terms of e. Analytically, this requires writing $D(e)$ as the root of a quartic, but the inversion is simple to do numerically by coupling (5.9) with (5.8).

The potential $U(e)$ is plotted in figure 6(*a*). For a narrow range of values of f_0 and γ_0, $U(e)$ has two peaks, at e_1 and e_2. With this two-peak shape for $U(e)$, the oscillator equation (5.7) has two stable steady states corresponding to the peaks of the potential: one with a smaller energy, and the other with larger energy. These correspond to values of the energy in interfaces and layers, respectively. The profile of $U(e)$ depends sensitively on γ_0 and f_0, but for each f_0, there is a precise value of γ_0 such that $U(e_1) = U(e_2)$; this value of U is shown by the red dashed line in figure 6(*a*). For this critical value of γ_0, the oscillator $e(T)$ has a special kink solution linking the two maxima (BLY). By analogy with similar kink solutions to the Cahn–Hilliard equation, more complex solutions with gradient spikes can also be constructed (e.g. Fraerman *et al.* 1997).
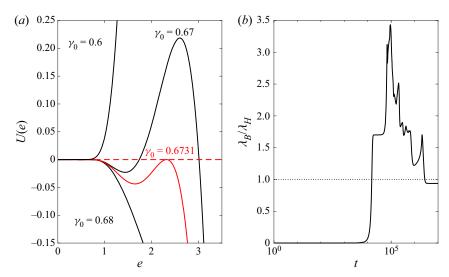
Figure 6. (*a*) Potential $U(e)$ found by integrating (5.8) with respect to D, and coupling with (5.9) to find corresponding values of e. The red line has two peaks, at e_1 and e_2, where $U(e_1) = U(e_2)$. Parameter values are $f_0 = 0.45$, $\tau = 0.01$, $\sigma = 10$, $\epsilon = 1$ and $\delta = 0.001$. (*b*) Plot of the ratio λ_B/λ_H calculated in the small-gradient case given by (5.10), for the solutions presented in figure 5. The dotted line marks $\lambda_B/\lambda_H = 1$: above this line, the B-merger dominates; below it, the H-merger dominates.

With the potential taking this two-peak form, our three-component system maps directly to the Radko (2007) analysis of the merging behaviour of the BLY model, which is susceptible to both B- and H-mergers. Radko shows that, in this situation, the ratio of their growth rates is

$$\frac{\lambda_B}{\lambda_H} = \frac{\bar{g} - g_{min}}{g_{min}}, \qquad (5.10)$$

where g_{min} is the minimum gradient (in layers), and $\bar{g}$ is the background gradient (averaged across the whole layer–interface system). If the ratio λ_B/λ_H is greater than unity, then B-mergers occur, and if the ratio is less than unity, then H-mergers will dominate instead. In the solutions shown in figure 5, it appears that the mergers occur via the B-merger pattern, with weaker interfaces shrinking without significant drifting. To show consistency with this condition on λ_B/λ_H, we calculate λ_B/λ_H at each time, using (5.10). We take g_{min} to be the global minimum gradient at each time, and $\bar{g} = 1$ as the background gradient. The ratio λ_B/λ_H is shown in figure 6(*b*), where we see that for times $10^4 \lesssim t \lesssim 2 \times 10^6$, the ratio $\lambda_B/\lambda_H > 1$, implying consistency with the numerical results, in which B-mergers dominate.

5.3. *Increase of the buoyancy flux*

In the full thermohaline system, the existence of staircases leads to greater turbulent transport of both heat and salt through the fluid in comparison with an unlayered state (e.g. Rosenblum *et al.* 2011; Hughes & Brummell 2021). To investigate this effect using our model, figure 7 shows the evolution of the mean of the upward buoyancy flux

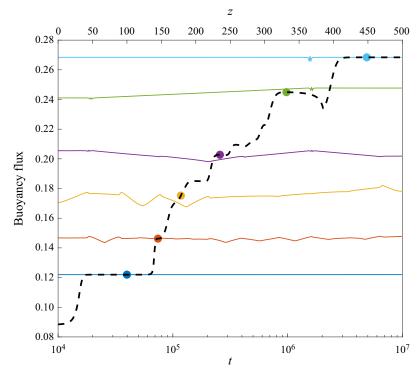$$\frac{1}{H} \int_0^H (c - f) \, \mathrm{d}z, \qquad (5.11)$$

Figure 7. Evolution of the buoyancy flux field for the solution shown in figure 5(*a*). The black dashed line shows the vertical mean of the upward buoyancy flux $(c - f)$ defined by (5.11), plotted against time on the lower horizontal axis. The spatial profile of the buoyancy flux (with z on the upper horizontal axis) is also shown at a range of times, with the corresponding mean flux at each time marked with a dot of the same colour.

calculated at each time for the solution shown previously in figure 5. There is little change in the flux at early times, until the initial linear perturbation grows into a stack of layers at $t \approx 10^4$. As layers undergo mergers, the magnitude of the flux increases, with thicker mixed layers producing larger fluxes, consistent with simulations of salt fingering (e.g. Rosenblum *et al.* 2011). Also plotted are the spatial profiles of the upward buoyancy flux $(c - f)$ at a range of times, showing that, in general, the flux is reasonably consistent across the different layers at each time, but with small perturbations at the interfaces.

The increase in flux seen in figure 7 can be explained using the merger theorem of Radko (2007). We saw in figure 5 that layers undergo B-mergers, where weak interfaces decay with little vertical drift. To explain this, we consider a region of fluid initially containing two layers and one such weak interface. Initially, there is a buoyancy jump of B_1 across the interface. When a B-merger takes place, the resultant state is a single mixed layer, with the buoyancy variation from the top to the bottom now $B_2 \ll B_1$. The merger theorem states that the system is unstable to B-mergers if the buoyancy flux decreases as the buoyancy variation increases, so this decrease of B during the merger must increase the flux in the region.

5.4. *Variation of parameter values*

An exploration of parameter space reveals that the solutions shown in figure 5 are representative of a large range of parameter values. To demonstrate this, we illustrate the
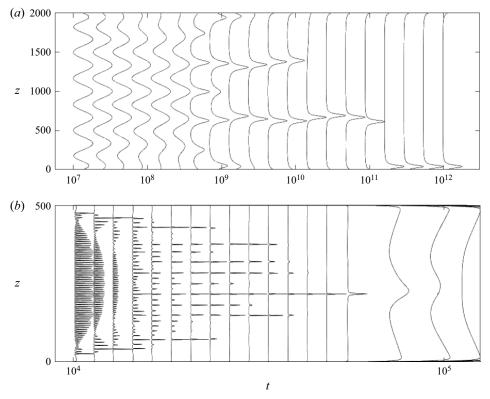
968 A16-22

Figure 8. Evolution of solutions to (3.15)–(3.17) with length scale (3.18), subject to boundary conditions (5.1a,b)–(5.3a,b) and initial conditions (5.4)–(5.6). (a) Very high Prandtl number case, with $\tau = 0.0001$, $\sigma = 10^4$, $\delta = 0.001$, $\epsilon = 1$ and $R_0 = 3.64$, showing wide, smooth interfaces and layers. (b) No overall background buoyancy gradient, with $\tau = 0.01$, $\sigma = 1$, $\delta = 0.01$, $\epsilon = 1$ and $R_0 = 1$, in which the time scale for mergers is similar to that for layer growth. Layers merge quickly, eventually giving way to a single convective state across the entire domain.

solution for some extreme choices of parameters, shown in figure 8. Figure 8(a) takes a very large value of the Prandtl number ($\sigma = 10^4$), for which the scale of the most unstable mode is very large, resulting in layers and interfaces that are wider and smoother in comparison with the results of figure 5. Figure 8(b) shows the case with $R_0 = 1$, on the boundary between the salt fingering and statically unstable regimes. In this case, layers do form, but mergers occur on such a fast time scale that a regular staircase never develops, and the system transitions very quickly from the linear growth phase to a single interface in the middle, which then decays, forming a single layer across the entire domain depth. In both of the cases shown, while there are some quantitative differences from the results in figure 5, the qualitative behaviour is the same.

We selected the parameter values $\delta = 0.001$ and $\epsilon = 1$ to prevent the energy mode from being unstable. If instead we choose parameters such that there is an energy mode instability, then there is energy growth on the domain scale, producing a wide interior region with a large energy and constant temperature and salt gradients, with narrow boundary regions on either side to satisfy the boundary conditions (5.1a,b)–(5.3a,b).

968 A16-23

6. Discussion

We have demonstrated the development of models for staircases in three-component systems and formulated the first regularised mixing-length theory of salt fingering staircases that includes double-diffusive physics directly. We have shown the possibility of instability without the need for external forcing. There are four possible types of instability. First, there is the energy mode, which is due to instability in the energy equation alone; this mode is most unstable at wavenumber $m = 0$, leading to growth in energy across the whole domain. Next, there are three possible instabilities attributable to either the Phillips effect or a γ-style instability, depending on the parametrisations of the fluxes. At low wavenumbers, there can be one or two positive growth rates, corresponding to positive eigenvalues of the Jacobian of the temperature and salinity fluxes with respect to their gradients. There is also the possibility of a high-wavenumber instability, which can be avoided by suitable parametrisation of the flux terms.

We compare these results with those of two important previous studies. The system of BLY is unstable to a single Phillips mode, and also allows for a stable energy mode. We have shown the possibility of an extra Phillips mode, allowing for an oscillatory instability, as may be expected in the diffusive convection regime. While BLY's use of the energy equation prevented the high-wavenumber instability inherent to Phillips (1972) and Posmentier (1977), a careful choice of fluxes is necessary to avoid instability at high wavenumbers in our three-component system. The model of Radko (2003) also displays a high-wavenumber instability; as in BLY, this is regularised by our inclusion of an energy equation. Further, we have demonstrated that Radko's γ instability is mathematically equivalent to the Phillips effect in the case where turbulent fluxes depend on the density ratio alone. The multiscale analysis of Radko (2019) provides an alternative method of regularising the high-wavenumber instability by the inclusion of hyperdiffusion terms. This leads to a model that produces the key dynamics of double-diffusive layering in terms of only the temperature and salinity fields (in comparison with our three-component model). Radko's model depends on the empirical calibration of several coefficients, requiring direct numerical simulations *a priori* to inform the choices of coefficients. By contrast, our model is derived using a spatial averaging process and the choice of a simple mixing length dependent on only one parameter. Both our model and that of Radko (2019) provide similar numerical results, with staircases appearing and gradually reducing via the B-merger pattern of Radko (2007).

We have applied the general results for the linear stability of three-component systems in § 2 to the model for DDC presented in § 3. Our model describes the evolution of horizontally averaged temperature, salinity and turbulent kinetic energy fields, based on turbulent flux terms expressed in terms of a mixing length parametrised as a function of the dependent variables. There is no externally imposed stirring or energy source, in contrast to the model of Paparella & von Hardenberg (2014), which also employed a BLY-like framework. The parametrisation of the length scale is a key ingredient of the model, and the form chosen, (3.18), is not the only possibility. It is interesting to note that neither the length scale adopted by BLY and Pružina *et al.* (2022) in a model of stirred convection, nor any similar simple length scale parametrised in terms of b_z, provides the appropriate release of potential energy from the flux terms to generate layering in (non-stirred) DDC. A prescription based on the density ratio R alone leads to a high-wavenumber instability. However, a relatively simple parametrisation in terms of both R and e, with a form chosen based on the γ instability theory, captures the essential physics of the layering process.

We note here that a slightly simpler mathematical model of layering could be produced by a reduced form of the model in which $e_t \equiv 0$. In this case, the energy equation becomes

diagnostic, and e is determined directly at all times as a functional dependent on the global temperature and salinity fields. This modification does not affect the conditions for the layering instability (2.17*a,b*), but removes the possibility of the energy mode instability (2.11). In this case, we are effectively left with a two-component system for T and S, closed via a functional parametrisation of e in the form of the steady energy equation. The layering instability, and its regularisation, are dependent entirely on the specific forms of f, c and p.

In § 4, we showed that the model (3.15)–(3.17) admits uniform-gradient steady states $(T_z, S_z, e) = (z, z/R_0, e_0)$ in the salt fingering regime, with $e_0(R_0)$ being single-valued for appropriate choices of δ and ϵ. These steady states are unstable to perturbations for a finite range of R_0, with a well-defined wavenumber of maximum growth rate. Increasing the value of τ decreases the range of R_0 susceptible to instability, with no instability possible for $\tau \gtrsim 0.105$ (a value dependent on δ and ϵ). Larger values of σ give wider ranges of R_0 for instability. For values of the Prandtl number $\sigma \lesssim 1$, the unstable range of R_0 is very narrow, and exists only for $R_0 < 1$, i.e. outside of the salt fingering regime. Hence, for sufficiently small Prandtl number, there is no possibility for salt fingering staircases to form via the γ-style instability, in agreement with the results of three-dimensional numerical simulations of the full thermohaline system (Traxler *et al.* 2011).

We have presented numerical solutions to the model in § 5, showing the initial development of a staircase and its subsequent evolution to late times. The staircase evolves via the B-merger process described by Radko (2007), eventually leaving a single strong interface with well-mixed layers on either side. During a merger event, the buoyancy gradient in the remaining interface increases, so that the gradient in the final interface is significantly higher than in the first interfaces to develop. The buoyancy flux through the system also shows sharp increases during merger events, which can be explained by the condition responsible for the B-merger that flux increases as buoyancy jump decreases. This increase in flux agrees with previous results (e.g. Rosenblum *et al.* 2011), and is an important piece in the puzzle needed to understand transport processes through staircases.

One important direction for further study is a comparison of this model with direct numerical simulations. The model may be tuned simply by choice of parameters, but also by reconsidering some of the assumptions on scalings in the original derivation. Through such fine-tuning, we hope to produce a more quantitative model that can be used to make accurate predictions for real staircase structures.

We saw in § 4 that the model in its current form does not admit steady states in the diffusive convection regime. Furthermore, numerical solutions of our model in this regime do not produce layers. It is therefore of significant interest to seek to adapt the model to diffusive convection. Ma & Peltier (2022) suggest that an energy source is necessary for diffusive layering to occur in a reduced model, with double-diffusive effects acting to regularise the layering process, rather than being the driving factor behind it. Adding an energy source to (3.17) does indeed allow the development of layers. However, three-dimensional numerical simulations (e.g. Stellmach *et al.* 2011) have shown the formation of clear layers with no such source term. As such, further study is required to understand the diffusive convective regime using a mixing-length framework.

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

Author ORCIDs.

Paul Pružina https://orcid.org/0000-0003-0458-2856;

David W. Hughes https://orcid.org/0000-0002-8004-8631;

Samuel S. Pegler https://orcid.org/0000-0001-8373-2693.

Appendix A. Derivation of the model

This appendix contains a derivation of the system (3.15)–(3.17) from the Boussinesq equations (3.10)–(3.14) using the horizontal averaging process of Pružina *et al.* (2022).

We define the horizontal average of a quantity $q(x, t)$ over a spatial area A at a height z as

$$\langle q \rangle \equiv \frac{1}{A} \int_A q(x, t) \, dA. \tag{A1}$$

Let $u_h = (u, v)$ represent the horizontal velocity, and ∇_h the horizontal gradient operator. The variables will be considered in terms of the sum of their horizontal mean and fluctuation components: $T = \langle T \rangle + T'$, $S = \langle S \rangle + S'$, $u_h = \langle u_h \rangle + u_h'$, $w = \langle w \rangle + w'$.

Taking the average of the incompressibility condition (3.13) and applying the condition of either impermeability or periodicity on the horizontal components of the velocity, we obtain

$$\langle \nabla_h \cdot u_h \rangle + \langle w_z \rangle = 0, \quad \text{and hence} \quad \langle w \rangle_z = 0. \tag{A2}$$

The horizontally averaged vertical velocity is therefore uniform across the height of the domain. Assuming impermeability conditions on the top and bottom boundaries, it follows that $\langle w \rangle = 0$. Thus there is no mean vertical velocity and $w = w'$.

We demonstrate the averaging process for the temperature equation. Beginning with (3.11), we take the horizontal average and apply (A2) to give

$$\langle T \rangle_t + \langle wT' \rangle_z = \langle T \rangle_{zz}. \tag{A3}$$

We subtract (A3) from the full temperature equation (3.11), and apply a quasi-linear approximation to neglect both the perturbation-perturbation term $u_h' \cdot \nabla_h T'$ and the perturbation temperature flux $wT_z' - \langle wT' \rangle_z$, giving

$$T_t' + w\langle T \rangle_z = \nabla^2 T'. \tag{A4}$$

We now use a scaling argument to parametrise the term $\langle wT' \rangle$ in terms of mean quantities. We assume that the mean of the square of the vertical velocity is a constant multiple of the mean total kinetic energy $\langle e \rangle = \langle u \cdot u \rangle / 2$, i.e.

$$\langle w^2 \rangle = \beta^2 \langle u \cdot u \rangle / 2 = \beta^2 \langle e \rangle, \tag{A5}$$

for some dimensionless constant β. While, in general, the value of β may vary, we assume that it is constant here to avoid overcomplicating the system. We assume that the turbulence varies on a mixing-length scale l, and on the dynamical time scale $\tau_d \sim l/\beta\langle e \rangle^{1/2}$, i.e. the characteristic time to move a vertical distance l. The length scale will be parametrised later in terms of the dependent variables, as discussed in § 3. With these assumptions, we approximate the time and space derivatives as $\partial_t \sim 1/\tau_d = \beta\langle e \rangle^{1/2}/l$ and $\nabla^2 \sim -1/l^2$.

Multiplying (A4) by w, averaging, using these scalings, rearranging, and applying (A5), gives the turbulent temperature flux as

$$\langle wT' \rangle = -\beta^2 \frac{l^2 \langle e \rangle}{\beta l \langle e \rangle^{1/2} + 1} \langle T \rangle_z. \tag{A6}$$

Combining (A3) and (A6) then gives the mean temperature equation as

$$\frac{1}{\beta} \langle T \rangle_t = \left(\frac{l^2 \langle e \rangle}{l \langle e \rangle^{1/2} + \beta^{-1}} \langle T \rangle_z \right)_z + \frac{1}{\beta} \langle T \rangle_{zz}. \tag{A7}$$

In this form, the dimensionless parameter β acts simply as a scale factor on the time derivative, effectively setting a new dimensionless time variable $\tilde{t} = \beta t$. This can be incorporated into the non-dimensionalisation using a scaled thermal diffusivity $\tilde{\kappa}_T = \beta \kappa_T$. With this transformation, and dropping angle brackets, the temperature equation becomes

$$T_t = \left(\frac{l^2 e}{l e^{1/2} + 1} T_z \right)_z, \tag{A8}$$

where we assume that the molecular diffusion term is small in comparison with the turbulent diffusion, and hence can be neglected. The salinity equation (3.16) is obtained by exactly the same process.

To obtain an equation for the mean kinetic energy, we begin by formulating the three-dimensional energy equation by taking the scalar product of the momentum equation (3.10) with $\boldsymbol{u}$, giving

$$e_t + \boldsymbol{\nabla} \cdot (\boldsymbol{u} e) = -\sigma \boldsymbol{\nabla} \cdot (\boldsymbol{u} p) + wb + \sigma (\nabla^2 e - |\boldsymbol{\nabla} \boldsymbol{u}|^2). \tag{A9}$$

The terms on the right-hand side represent, in order, the effect of pressure, conversion from potential to kinetic energy, diffusion of kinetic energy, and the viscous dissipation $D = -\sigma |\boldsymbol{\nabla} \boldsymbol{u}|^2$. The horizontal average of (A9) gives

$$\langle e \rangle_t + \langle we' \rangle_z - \sigma \langle e \rangle_{zz} = \langle wb' \rangle - \langle D \rangle. \tag{A10}$$

The turbulent energy flux $\langle we' \rangle$ is parametrised using the same method as for the temperature flux; the buoyancy flux is calculated as the temperature flux minus the salinity flux. Following BLY, the dissipation term is parametrised by $D = \epsilon e^{3/2}/l$. This form is commonly used in k–ϵ models of turbulence (e.g. Jones & Launder 1972). We finally obtain the energy equation

$$e_t = \left(\frac{l^2 e}{l e^{1/2} + \sigma} e_z \right)_z - \sigma \left(\frac{l^2 e}{l e^{1/2} + 1} T_z - \frac{l^2 e}{l e^{1/2} + \tau} S_z \right) + \sigma e_{zz} - \epsilon \frac{e^{3/2}}{l}, \tag{A11}$$

leading to the full system (3.15)–(3.17).

In the derivation above, we have assumed that $\nabla^2 \sim -1/l^2$ for each of the temperature, salinity and kinetic energy fields. A more accurate representation could possibly be obtained by using different scalings in each equation to reflect the different molecular diffusivities of each field. Here, however, we use the same scalings to produce a relatively simple model that admits staircase solutions.

Appendix B. Merger behaviour

This appendix contains the derivation of the nonlinear oscillator equation (5.7) for the energy, by considering constant flux steady states. The process outlined below mirrors that of BLY for their two-component system. Our derivation here includes the third equation in our three-component model.

Assuming a steady state, the system (3.15)–(3.17) can be written as

$$f_0 = \frac{D^2}{D+1} T_z, \tag{B1}$$

$$c_0 = \frac{D^2}{D+\tau} S_z, \tag{B2}$$

$$0 = (\kappa e_z)_z - \sigma (f_0 - c_0) - \epsilon \frac{e^2}{D}, \tag{B3}$$

where $D = le^{1/2}$ and $\kappa = D^2/(D+\sigma) + \sigma$. Note that (B1) defines $D(T_z)$ uniquely, and (B2) defines $S_z(D)$ uniquely. The salt gradient is therefore tied to the temperature gradient, rather than the two fields being independent.

Dividing (B1) by (B2), we obtain

$$\gamma_0 = \frac{f_0}{c_0} = \frac{D+\tau}{D+1} R, \tag{B4}$$

thus defining R in terms of D, where $\gamma_0 = f_0/c_0$ is the steady-state flux ratio. The prescription for the length scale (3.18), coupled with the definition $D = le^{1/2}$, can be rearranged to give an expression for $e(D, R)$, namely,

$$e^2 = (D^2 - \delta)R^2. \tag{B5}$$

Combining (B5) with (B4), we define the energy solely in terms of D by

$$e(D) = \frac{\sqrt{D^2 - \delta}\,(D+1)}{D+\tau} \gamma_0. \tag{B6}$$

Multiplying the energy equation (B3) by κe_z, integrating with respect to z, and changing variables such that $e_z \, dz = e_D \, dD$, we obtain

$$\frac{1}{2} (\kappa e_z)^2 + \int \left(-\sigma (f_0 - c_0) - \epsilon \frac{e(D)^2}{D} \right) \kappa e_D \, dD = E, \tag{B7}$$

where E is a constant.

To transform (B7) to an equation for e_b, we could write $e_z = e_b b_z$ and divide all terms in (B7) by $(\kappa b_z)^2$. However, when $b_z = 0$, this leads to a singularity in the potential. To avoid this, we note that $T_z \neq 0$ when $f_0 \neq 0$, and write $e_z = e_T T_z$ in terms of the temperature instead. This use of T instead of b as the 'time' variable is valid because D and S_z are defined as functions of T_z via (B1)–(B2). On using (B1) to rewrite T_z as a function of D, we obtain

$$\frac{1}{2} e_T^2 + U(D) = E \left(\frac{D^2}{\kappa f_0 (D+1)} \right)^2, \tag{B8}$$

where the potential $U(D)$ is defined by

$$U(D) = \left(\frac{D^2}{\kappa f_0 (D+1)} \right)^2 \int \left(-\sigma (f_0 - c_0) - \epsilon \frac{e(D)^2}{D} \right) \kappa \, dD. \tag{B9}$$

REFERENCES

ASHOURVAN, A. & DIAMOND, P.H. 2016 How mesoscopic staircases condense to macroscopic barriers in confined plasma turbulence. *Phys. Rev.* E **94**, 051202.

ASHOURVAN, A. & DIAMOND, P.H. 2017 On the emergence of macroscopic transport barriers from staircase structures. *Phys. Plasmas* **24**, 012305.

BALMFORTH, N.J., LLEWELLYN SMITH, S.G. & YOUNG, W.R. 1998 Dynamics of interfaces and layers in a stratified turbulent fluid. *J. Fluid Mech.* **355**, 329–358.

COCLITE, G.M., PAPARELLA, F. & PELLEGRINO, S.F. 2018 On a salt fingers model. *Nonlinear Anal.* **176**, 100–116.

FRAERMAN, A.A., MEL'NIKOV, A.S., NEFEDOV, I.M., SHERESHEVSKII, I.A. & SHPIRO, A.V. 1997 Nonlinear relaxation dynamics in decomposing alloys: one-dimensional Cahn–Hilliard model. *Phys. Rev.* B **55**, 6316–6323.

GUO, W., DIAMOND, P.H., HUGHES, D.W., WANG, L. & ASHOURVAN, A. 2019 Scale selection and feedback loops for patterns in drift wave–zonal flow turbulence. *Plasma Phys. Control. Fusion* **61**, 105002.

HUGHES, D.W. & BRUMMELL, N.H. 2021 Double-diffusive magnetic layering. *Astrophys. J.* **922**, 195.

HUPPERT, H.E. & LINDEN, P.F. 1979 On heating a stable salinity gradient from below. *J. Fluid Mech.* **95**, 431–464.

JONES, W.P. & LAUNDER, B.E. 1972 The prediction of laminarization with a two-equation model of turbulence. *Intl J. Heat Mass Transfer* **15**, 301–314.

KIMURA, S., SMYTH, W. & KUNZE, E. 2011 Turbulence in a sheared, salt-fingering-favorable environment: anisotropy and effective diffusivities. *J. Phys. Oceanogr.* **41**, 1144–1159.

MA, Y. & PELTIER, W.R. 2022 Thermohaline staircase formation in the diffusive convection regime: a theory based upon stratified turbulence asymptotics. *J. Fluid Mech.* **931**, R4.

MALKOV, M.A. & DIAMOND, P.H. 2019 Dynamics of potential vorticity staircase evolution and step mergers in a reduced model of beta-plane turbulence. *Phys. Rev. Fluids* **4**, 044503.

MCDOUGALL, T.J. & TAYLOR, J.R. 1984 Flux measurements across a finger interface at low values of the stability ratio. *J. Mar. Res.* **42**, 1–14.

MERRYFIELD, W.J. 2000 Origin of thermohaline staircases. *J. Phys. Oceanogr.* **30**, 1046–1068.

NEAL, V.T., NESHYBA, S. & DENNER, W. 1969 Thermal stratification in the Arctic Ocean. *Science* **166**, 373–374.

PAPARELLA, F. & VON HARDENBERG, J. 2012 Clustering of salt fingers in double-diffusive convection leads to staircaselike stratification. *Phys. Rev. Lett.* **109**, 014502.

PAPARELLA, F. & VON HARDENBERG, J. 2014 A model for staircase formation in fingering convection. *Acta Appl. Maths* **132**, 457–467.

PHILLIPS, O.M. 1972 Turbulence in a strongly stratified fluid – is it unstable? *Deep-Sea Res.* **19**, 79–81.

POSMENTIER, E.S. 1977 The generation of salinity finestructure by vertical diffusion. *J. Phys. Oceanogr.* **7**, 298–300.

PRUŽINA, P., HUGHES, D.W. & PEGLER, S.S. 2022 Development and long-term evolution of density staircases in stirred stratified turbulence. *Phys. Rev. Fluids* **7**, 104801.

RADKO, T. 2003 A mechanism for layer formation in a double-diffusive fluid. *J. Fluid Mech.* **497**, 365–380.

RADKO, T. 2007 Mechanics of merging events for a series of layers in a stratified turbulent fluid. *J. Fluid Mech.* **577**, 251–273.

RADKO, T. 2013 *Double-Diffusive Convection*. Cambridge University Press.

RADKO, T. 2019 Thermohaline layering on the microscale. *J. Fluid Mech.* **862**, 672–695.

ROSENBLUM, E., GARAUD, P., TRAXLER, A. & STELLMACH, S. 2011 Turbulent mixing and layer formation in double-diffusive convection: three-dimensional numerical simulations and theory. *Astrophys. J.* **731**, 66.

SCHMITT, R.W., LEDWELL, J.R., MONTGOMERY, E.T., POLZIN, K.L. & TOOLE, J.M. 2005 Enhanced diapycnal mixing by salt fingers in the thermocline of the tropical Atlantic. *Science* **308**, 685–688.

SCHMITT, R.W., PERKINS, H., BOYD, J.D. & STALCUP, M.C. 1987 C-salt: an investigation of the thermohaline staircase in the western tropical North Atlantic. *Deep-Sea Res.* I **34**, 1655–1665.

STELLMACH, S., TRAXLER, A., GARAUD, P., BRUMMELL, N. & RADKO, T. 2011 Dynamics of fingering convection. Part 2. The formation of thermohaline staircases. *J. Fluid Mech.* **677**, 554–571.

STERN, M.E. 1960 The 'salt-fountain' and thermohaline convection. *Tellus* **12**, 172–175.

STERN, M.E. 1969 Collective instability of salt fingers. *J. Fluid Mech.* **35**, 209–218.

STERN, M.E. & TURNER, J.S. 1969 Salt fingers and convecting layers. *Deep-Sea Res.* **16**, 497–511.

TAIT, R.I. & HOWE, M.R. 1968 Some observations of thermo-haline stratification in the deep ocean. *Deep-Sea Res.* **15**, 275–280.

TIMMERMANS, M.L., TOOLE, J., PROSHUTINSKY, A., KRISHFIELD, R. & PLUEDDEMANN, A. 2008 Eddies in the Canada Basin, Arctic Ocean, observed from ice-tethered profilers. *J. Phys. Oceanogr.* **38**, 133–145.

TRAXLER, A., GARAUD, P. & STELLMACH, S. 2011 Numerically determined transport laws for fingering ('thermohaline') convection in astrophysics. *Astrophys. J.* **728**, L29.

TURNER, J.S. 1973 *Buoyancy Effects in Fluids*. Cambridge University Press.

TURNER, J.S. & STOMMEL, H. 1964 A new case of convection in the presence of combined vertical salinity and temperature gradients. *Proc. Natl Acad. Sci.* USA **52**, 49–53.

VERONIS, G. 1965 On finite amplitude instability in thermohaline convection. *J. Mar. Res.* **23**, 1–17.

ZHURBAS, V.M. & OZMIDOV, R.V. 1984 Step-like structure forms of the oceanic thermocline and their generation mechanisms. *Okeanologiya* **24**, 197–203.