Negative available potential energy dissipation as the fundamental criterion for double diffusive instabilities

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1017/jfm.2024.647

www.reading.ac.uk/centaur

CentAUR

JM PAPERS

Negative available potential energy dissipation as the fundamental criterion for double diffusive instabilities

R. Tailleux†

Department of Meteorology, University of Reading, Reading RG6 6ET, UK

The background potential energy (BPE) is the only reservoir that double diffusive instabilities can tap their energy from when developing from an unforced motionless state with no available potential energy (APE). Recently, Middleton and Taylor linked the extraction of BPE into APE to the sign of the diapycnal component of the buoyancy flux, but their criterion can predict only diffusive convection instability, not salt finger instability. Here, we show that the problem can be corrected if the sign of the APE dissipation rate is used instead, making it emerge as the most fundamental criterion for double diffusive instabilities. A theory for the APE dissipation rate for a two-component fluid relative to its single-component counterpart is developed as a function of three parameters: the diffusivity ratio, the density ratio, and a spiciness parameter. The theory correctly predicts the occurrence of both salt finger and diffusive convection instabilities in the laminar unforced regime, while more generally predicting that the APE dissipation rate for a two-component fluid can be enhanced, suppressed, or even have the opposite sign compared to that for a single-component fluid, with important implications for the study of ocean mixing. Because negative APE dissipation can also occur in stably stratified single-component and doubly stable two-component stratified fluids, we speculate that only the thermodynamic theory of exergy can explain its physics; however, this necessitates accepting that APE dissipation is a conversion between APE and the internal energy component of BPE, in contrast to prevailing assumptions.

Key words: double diffusive convection, turbulent mixing

† Email address for correspondence: R.G.J.Tailleux@reading.ac.uk

1. Introduction

Stratified binary fluids, in which the stratifying agents diffuse at different rates, exhibit a range of intriguing phenomena not observed in single-component stratified fluids. One such phenomenon is double diffusive instabilities, which arise when one of the stratifying agents acts to destabilise the stratification. In the context of ocean mixing, this leads to the emergence of new types of flows, including diffusive convection, salt fingering, thermohaline intrusions, and thermohaline staircases (Turner 1985; Schmitt 1994; Radko 2013). Theoretical analysis categorises the different dynamically relevant types of stratification based on the density ratio (R_ρ) of the fluid: (1) the doubly stable case ($R_\rho < 0$), when both temperature and salinity are stabilising; (2) salt fingers favourable ($R_\rho > 1$), with salinity as the destabilising agent; (3) diffusive convection favourable ($0 < R_\rho < 1$), where temperature acts as the destabiliser. Understanding how double diffusion modulates mixing in the doubly stable case, and characterising double diffusive instabilities in both laminar and turbulent regimes, are key questions of interest.

To date, linear stability analysis has been the primary approach for identifying the physical parameters that control the development of double diffusive instabilities from a quiescent state. Such an analysis reveals, for instance, that molecular viscosity and diffusion can severely restrict the range of density ratios at which these instabilities can occur, namely,

$$\left.\begin{aligned} \frac{Pr + \tau}{Pr + 1} < R_\rho < 1, \quad &\text{for diffusive convection,} \\ 1 < R_\rho < \frac{1}{\tau}, \quad &\text{for salt fingers} \end{aligned}\right\} \tag{1.1}$$

(e.g. Radko 2013), where $\tau = \kappa_S/\kappa_T$ represents the ratio of salt diffusivity to temperature diffusivity, $Pr = \nu/\kappa_T$ denotes the Prandtl number, and ν is the molecular viscosity. For example, using the typical oceanic values $\tau \approx 0.01$ and $Pr \approx 7$ yields a reduced density ratio range $1 < R_\rho < 100$ for salt fingers and $0.8076 < R_\rho < 1$ for diffusive convection.

While linear stability theory provides valuable insights into double diffusive instabilities, other approaches are necessary to understand their energy source and whether they can persist in a turbulent environment like the oceans. In this regard, the theory of available potential energy (APE) introduced by Lorenz (1955) and applied to the study of turbulent stratified fluids by Winters *et al.* (1995) appears to be one of the most promising avenues of research. According to APE theory, the energetics of a stratified fluid can be separated into two main budgets, one for the total mechanical energy (APE + kinetic energy), and one for the background potential energy (BPE). Physically, the BPE represents the potential energy associated with the state of minimum potential energy achievable from the actual state through an adiabatic (and isohaline in seawater) mass rearrangement. The generic form of these budget equations is

$$\frac{\mathrm{d}(APE + KE)}{\mathrm{d}t} = \text{Forcing} - \int_V \rho(\varepsilon_k + \varepsilon_p)\,\mathrm{d}V, \tag{1.2}$$

$$\frac{\mathrm{d}(BPE)}{\mathrm{d}t} = \int_V \rho(\varepsilon_k + \varepsilon_p)\,\mathrm{d}V, \tag{1.3}$$

where ε_k is the positive definite viscous dissipation rate, and ε_p is the diffusive rate of APE dissipation. As is well known, viscous dissipation is an irreversible process that converts kinetic energy into 'heat', which in APE theory is played by the BPE.

In a single-component compressible fluid, Tailleux (2013c) (for which an improved version is presented in Appendix A) established that the APE dissipation rate ε_p may be broken down into three components:

$$\varepsilon_p = \varepsilon_{p,lam} + \varepsilon_{p,tur} + \varepsilon_{p,eos}, \tag{1.4}$$

with $\varepsilon_{p,lam}$ and $\varepsilon_{p,tur}$ representing the laminar and turbulent parts of ε_p, respectively, and $\varepsilon_{p,eos}$ representing the part of ε_p arising from the nonlinearities of the equation of state. To date, the APE dissipation rate has been primarily defined and studied for single-component Boussinesq fluids with a linear equation of state, for which only $\varepsilon_{p,lam}$ and $\varepsilon_{p,tur}$ subsist. As clarified further in the text, $\varepsilon_{p,lam}$ and $\varepsilon_{p,tur}$ relate to the terms $-\Phi_i$ and Φ_d in the Winters *et al.* (1995) framework. Physically, $\varepsilon_{p,tur}$ is the fundamental quantity for defining the turbulent diapycnal diffusivity

$$K_\rho = \frac{\varepsilon_{p,tur}}{N^2} = \frac{\Gamma \varepsilon_k}{N^2} \tag{1.5}$$

(e.g. Osborn & Cox 1972; Osborn 1980; Lindborg & Brethouwer 2008), often quantified by the dissipation ratio $\Gamma = \varepsilon_{p,tur}/\varepsilon_k$, a common measure of mixing efficiency often assumed to be close to 0.2. Typically, $\varepsilon_{p,tur}$ is estimated from microstructure measurements using formulas such as the widely used expression

$$\varepsilon_{p,tur} = \frac{g\alpha\kappa_T |\nabla\theta'|^2}{d\bar{\theta}/dz} \tag{1.6}$$

(Oakey 1982; Gargett & Holloway 1984), where α is the thermal expansion coefficient, g is the acceleration due to gravity, and $\bar{\theta}$ represents the mean potential temperature profile. Consequently, ε_p, like the viscous dissipation rate, is generally considered a positive quantity that also converts mechanical energy into BPE or 'heat'. (For further views on this, see Tailleux 2009.)

In a double diffusive fluid, double diffusive instabilities can develop even in the absence of mechanical forcing. Within the framework of APE theory, this is possible only if BPE, like APE, can become a source of kinetic energy, and hence if we accept the notion that ε_p can become negative. In this paper, we derive a theoretical expression for the ratio $\varepsilon_p^{dd}/\varepsilon_p^{std}$ of the APE dissipation rates for a double diffusive case over that for a simple fluid. We demonstrate that ε_p^{dd} is negative under the conditions corresponding to the linear stability analysis for both salt finger and diffusive convection regimes. Recently, Middleton & Taylor (2020) proposed to characterise double diffusive instabilities in terms of the sign of the diapycnal component of the molecular buoyancy flux, which is related to the turbulent part $\varepsilon_{p,tur}$ of ε_p (equivalently, Φ_d), rather than ε_p. However, their criterion succeeds in predicting only the occurrence of the diffusive convection instability, but not the salt finger instability, which suggests that it is the sign of the net APE dissipation rate $\varepsilon_{p,lam} + \varepsilon_{p,tur}$ (equivalently $\Phi_d - \Phi_i$) that represents the most fundamental criterion for double diffusive instabilities. Additionally, we show that considering the APE dissipation rate sheds light on other aspects of double diffusive instabilities in turbulent regimes. For example, it correctly predicts the regime associated with diffusive interleaving in the presence of density-compensated isopycnal gradients of temperature and salinity, which are thought to be essential for the formation of thermohaline staircases (Merryfield 2000).

The remainder of this paper is organised as follows. Section 2 establishes the general form of the local APE dissipation rate for a double diffusive fluid. Section 3 outlines the conditions necessary for negative APE dissipation. Section 4 provides a comparison

of the local and APE frameworks. Section 5 discusses the practical issues related to the computation of the reference temperature and salinity profiles used in our definition of spiciness. Finally, § 6 discusses the results and provides future perspectives.

2. Local APE theory and APE dissipation rate

Similarly to Middleton & Taylor (2020), we study the energetics of double diffusive instabilities for a Boussinesq fluid with a linear equation of state for density governed by the system of equations

$$\frac{D\boldsymbol{v}}{Dt} + \frac{1}{\rho_\star}\boldsymbol{\nabla}\delta p = -\frac{g\,\delta\rho}{\rho_\star}\,\boldsymbol{k} + \nu\,\nabla^2\boldsymbol{v}, \tag{2.1}$$

$$\boldsymbol{\nabla}\cdot\boldsymbol{v} = 0, \tag{2.2}$$

$$\frac{D\theta}{Dt} = -\boldsymbol{\nabla}\cdot\boldsymbol{J}_\theta, \quad \frac{DS}{Dt} = -\boldsymbol{\nabla}\cdot\boldsymbol{J}_S, \tag{2.3a,b}$$

$$\rho = \rho_\star[1 - \alpha(\theta - \theta_\star) + \beta(S - S_\star)], \tag{2.4}$$

where ρ is density, p is pressure, g is gravitational acceleration, S is salinity, θ is potential temperature, α is the thermal expansion coefficient, and β is the haline contraction coefficient, while $\rho_\star$, $S_\star$ and $\theta_\star$ are constant reference values for ρ, S and θ, respectively. Moreover, $\delta p = p - p_0(z)$ and $\delta\rho = \rho - \rho_0(z)$ denote the pressure and density anomalies defined relative to the reference pressure and density profiles $p_0(z)$ and $\rho_0(z) = -g^{-1}\,dp_0/dz$ characterising the Lorenz reference state. Simple diffusive laws are assumed for θ and S so that $\boldsymbol{J}_\theta = -\kappa_T\,\boldsymbol{\nabla}\theta$ and $\boldsymbol{J}_s = -\kappa_S\,\boldsymbol{\nabla}S$, where κ_T and κ_S are the molecular diffusivities of heat and salt, respectively. The tracer equations (2.3a,b) may be combined with the equation of state (2.4) to form an equation for the Boussinesq buoyancy $b_{bou} = g[\alpha(\theta - \theta_\star) - \beta(S - S_\star)]$,

$$\frac{Db_{bou}}{Dt} = -\boldsymbol{\nabla}\cdot\boldsymbol{J}_b, \tag{2.5}$$

where

$$\boldsymbol{J}_b = g(\alpha\boldsymbol{J}_\theta - \beta\boldsymbol{J}_s) = -g(\kappa_T\alpha\,\boldsymbol{\nabla}\theta - \kappa_S\beta\,\boldsymbol{\nabla}S) \tag{2.6}$$

is the molecular buoyancy flux. Note that in the absence of salinity, $\boldsymbol{J}_b = -\kappa_T\,\boldsymbol{\nabla}b_{bou}$ is down the gradient of the Boussinesq buoyancy, but not in the general case.

To understand how double diffusion affects APE dissipation, the local APE framework (Tailleux 2013b, 2018; Novak & Tailleux 2018) is preferred over the Winters *et al.* (1995) global APE framework used by Middleton & Taylor (2020). Indeed, we show in § 4 that defining the total APE of a fluid as the volume integral of a non-sign definite integrand makes the global APE framework problematic in several respects; for instance, it introduces unphysical terms in its budget. To avoid such difficulties, it is therefore preferable to define the APE of a fluid as the volume integral of the APE density:

$$e_a = e_a(S, \theta, z) = -\int_{z_r}^z b(S, \theta, z')\,dz', \tag{2.7}$$

where $b(S, \theta, z) = -(g/\rho_\star)(\rho(S, \theta) - \rho_0(z))$ is the buoyancy defined relative to Lorenz reference density profile $\rho_0(z)$ characterising the notional reference state of minimum potential energy obtainable from the actual state by means of an adiabatic and volume-conserving rearrangement. This definition of buoyancy has the

994 A5-4

advantage of vanishing in the Lorenz reference state of rest, which is not the case for the Boussinesq buoyancy. To avoid any ambiguity, note that $b(S, \theta, z') = -(g/\rho_\star)(\rho(S(x, y, z, t), \theta(x, y, z, t)) - \rho_0(z'))$ in the integral (2.7), that is, the values of S and θ are held constant as the parcel moves from z_r to z, consistent with the idea of an adiabatic and isohaline rearrangement. As is well known, the APE density (2.7) represents the work against buoyancy forces that a parcel needs to counteract to move from its reference position at z_r, solution of the level of neutral buoyancy equation

$$b(S, \theta, z_r) = 0, \tag{2.8}$$

to its actual position at z. For simplicity, we disregard the temporal dependence of $\rho_0(z)$ as the APE dissipation depends only on spatial gradients of the tracer fields. Inversion of (2.8) establishes that $z_r = z_r(\rho) = z_r(S, \theta)$ is a material function of S and θ, and is therefore a Lagrangian invariant following fluid parcels in the absence of diffusive sources/sinks of heat and salt.

A local budget equation for the APE density is easily obtained by taking the material derivative of (2.7), which yields

$$\frac{\mathrm{D}e_a}{\mathrm{D}t} = -b \frac{\mathrm{D}z}{\mathrm{D}t} + \underbrace{b(S, \theta, z_r)}_{=0} \frac{\mathrm{D}z_r}{\mathrm{D}t} - (z - z_r) \frac{\mathrm{D}b_{ou}}{\mathrm{D}t}$$
$$= -bw - \nabla \cdot \boldsymbol{J}_a - \varepsilon_p, \tag{2.9}$$

where $\boldsymbol{J}_a$ is the diffusive flux of APE, and ε_p is the APE dissipation rate, given by

$$\boldsymbol{J}_a = -(z - z_r)\boldsymbol{J}_b, \tag{2.10}$$

$$\varepsilon_p = \boldsymbol{J}_b \cdot \nabla(z - z_r) = \underbrace{-g \left(\alpha \kappa_T \frac{\partial \theta}{\partial z} - \beta \kappa_S \frac{\partial S}{\partial z} \right)}_{\varepsilon_{p,lam}} + \underbrace{g(\alpha \kappa_T \nabla \theta - \beta \kappa_S \nabla S) \cdot \nabla z_r}_{\varepsilon_{p,tur}}. \tag{2.11}$$

As indicated by (2.11), ε_p can be decomposed into 'laminar' and 'turbulent' components $\varepsilon_{p,lam}$ and $\varepsilon_{p,tur}$ whose volume integrals

$$\int_V \varepsilon_{p,lam} \, \rho_\star \, \mathrm{d}V = -\Phi_i, \quad \int_V \varepsilon_{p,tur} \, \rho_\star \, \mathrm{d}V = \Phi_d, \tag{2.12a,b}$$

can easily be shown to correspond to Winters *et al.* (1995) global energy conversions $-\Phi_i$ and Φ_d in the global APE framework. So far, the turbulent mixing community has generally assumed Φ_i and Φ_d to represent distinct types of energy conversions, the former with internal energy and the latter with the BPE; however, the analysis of such conversions in real fluids detailed in § 4 clearly indicates that there is no physical basis for such a view, and that in reality, ε_p, Φ_d and Φ_i all represent conversions with internal energy as established previously by Tailleux (2009). In other words, $\varepsilon_{p,lam}$ and $\varepsilon_{p,tur}$ are parts of the same energy conversion.

Physically, (2.9) states that locally, e_a can be modified through: (1) conversion with kinetic energy via the buoyancy flux bw; (2) diffusive and advective transport via $\boldsymbol{J}_a$ and advection; (3) 'dissipation' that can be occasionally be negative associated with ε_p. Physically, our local APE budget (2.9) is simpler in form than that derived previously by Scotti & White (2014) for a single-component fluid, although the two can be verified to be

mathematically equivalent. One of the reasons is due to Scotti & White (2014) imposing the diffusive flux of APE density to be down-gradient, namely,

$$J_{a,sw14} = -\kappa \, \nabla e_a = \kappa(z - z_r) \, \nabla b_{bou} + \kappa b \boldsymbol{k} = J_a + \kappa b \boldsymbol{k}, \qquad (2.13)$$

which differs from J_a by the last term, with the consequence of adding extra terms in their budget equation. Physically, however, the Scotti & White (2014) approach is inconsistent with the analysis of the APE budget for a real fluid (cf. §4 or Tailleux 2009), which reveals that decomposing the potential energy into available and unavailable components leads to a decomposition of the heat flux $J_q = \Upsilon J_q + (1 - \Upsilon)J_q$, with $\Upsilon = (T - T_r)/T$ a Carnot-like thermodynamic efficiency. Physically, it is the first term $J_a = \Upsilon J_q$ that represents the diffusive flux of APE density, whose Boussinesq approximation is given by (2.10). Another reason to be sceptical of (2.13) is that it cannot generalise to the case of a double diffusive two-component fluid, whereas our definition of J_a in (2.10) obviously does.

In their study, Middleton & Taylor (2020) derived an expression for the diapycnal component of the buoyancy flux $J_b \cdot \hat{\boldsymbol{n}}$ in terms of three parameters – the diffusivity ratio $\tau = \kappa_S/\kappa_T$, the gradient ratio $G_\rho = (\alpha \, |\nabla\theta|)/(\beta \, |\nabla S|)$ and the angle γ between $\nabla\theta$ and ∇S – such that $\cos\gamma = \nabla\theta \cdot \nabla S/(|\nabla\theta| \, |\nabla S|)$, with $\hat{\boldsymbol{n}} = \nabla z_r/|\nabla z_r|$, that could similarly be used to study ε_p. However, while G_ρ and $\cos\gamma$ might be diagnosed easily from the output of a numerical experiment, they cannot be inferred easily from measured oceanic properties, nor are they naturally related to physical parameters known to play a key role in double diffusive instabilities such as density-compensated thermohaline variations. For instance, in the canonical salt finger experiment that they discuss in their §2.5, Middleton & Taylor (2020) find that Φ_d becomes negative as the result of the gradients of salinity growing more rapidly than temperature gradients, resulting in the creation of density-compensated thermohaline variations (usually referred to as 'spice'). In oceanography, the term spice was perhaps first used by Munk (1981) to quantify the range of possible θ/S behaviour of a seawater sample of given density, from warm and salty (spicy) to fresh and cold (minty). Physically, the creation of spice is of considerable dynamical importance because lateral stirring along isopycnal surfaces is not opposed by restoring buoyancy forces and is therefore considerably more efficient at dissipating tracer variances than vertical stirring. To study the role of spice, Middleton *et al.* (2021) subsequently reformulated the Middleton & Taylor (2020) expression in terms of the 'spiciness' variable $\tau = \alpha\theta + \beta S$ (denoted S_p in their paper), which may be regarded as the 'linear' version of the spiciness variables developed by Jacket & McDougall (1985) or Flament (2002), for instance. Recently, the theory of spiciness was revisited by Tailleux (2021), who argued that to be physically meaningful, spice should ideally vanish in a spiceless ocean, and hence that spice variables should be defined as isopycnal anomalies, which is not the case for τ or any related variable. In the present case, this can be implemented in practice by decomposing θ and S as

$$\theta = \theta_0(z_r) + \theta_\xi, \quad S = S_0(z_r) + S_\xi, \qquad (2.14a,b)$$

with the reference profiles $\theta_0(z_r)$ and $S_0(z_r)$ defining the temperature and salinity profiles of the assumed spiceless state, and $z_r = z_r(\rho)$ denoting the reference position of a fluid parcel as before. Provided that $\theta_0(z_r)$ and $S_0(z_r)$ are defined as the (thickness-weighted) isopycnal mean temperature and salinity, respectively, they contain all necessary information to compute density $\rho = \rho_\star[1 - \alpha(\theta_0(z_r) - \theta_\star) - \beta(S_0(z_r) - S_\star)]$, thus allowing them to be interpreted as the 'active' components of θ and S. In contrast, the isopycnal anomalies θ_ξ and S_ξ are density-compensated $\alpha\theta_\xi = \beta S_\xi$ and

therefore do not contribute to density, thus allowing them to be interpreted as the passive components of θ and S. Physically, defining the spiceless stratification $\theta_0(z)$, $S_0(z)$ in terms of thickness-weighted isopycnal means can be justified on the grounds that such a state is the one that would be expected to result in the idealised limit of infinitely fast (resp. infinitely slow) isopycnal (resp. diapycnal) mixing. How to estimate these in practice are discussed in § 5. As shown below, defining spiciness in this way allows us to define a new set of parameters with which to study the behaviour of ε_p, different from those used by Middleton & Taylor (2020) and Middleton *et al.* (2021), and which we consider to be more physical and more relevant.

In the following, S_ξ and θ_ξ are combined into the variable $\xi = \rho_\star(\alpha\theta_\xi + \beta S_\xi)$ to define a single measure of spiciness having the dimension of density, as is commonly done in the literature. After some manipulation, it is possible to rewrite ε_p in the form

$$
\begin{aligned}
\varepsilon_p &= g\left(\kappa_T\alpha\,\frac{\partial\tilde{\theta}}{\partial z_r} - \kappa_S\beta\,\frac{\partial\tilde{S}}{\partial z_r}\right)\left(|\nabla z_r|^2 - \frac{\partial z_r}{\partial z}\right) + \frac{g(\kappa_T - \kappa_S)}{2\rho_\star}\left(\nabla z_r \cdot \nabla\xi - \frac{\partial\xi}{\partial z}\right) \\
&= K_{v0}\,N_0^2(z_r)\left[\frac{R_\rho - \tau}{R_\rho - 1} + (1 - \tau)\mathcal{M}\right]\left(1 - \frac{\partial_z z_r}{|\nabla z_r|^2}\right),
\end{aligned}
\tag{2.15}
$$

which depends on the parameters

$$
\left.\begin{aligned}
\mathcal{M} &\equiv \frac{g}{2\rho_\star N_0^2(z_r)}\frac{\nabla z_r \cdot \nabla\xi - \partial_z\xi}{|\nabla z_r|^2 - \partial_z z_r} \quad \text{(spiciness parameter)}, \\
K_{v0} &\equiv \kappa_T\,|\nabla z_r|^2 = \kappa_T\left(\frac{\mathrm{d}\rho_0}{\mathrm{d}z_r}(z_r)\right)^{-2}|\nabla\rho|^2 \quad \text{(effective diffusivity)}, \\
N_0^2(z_r) &\equiv -\frac{g}{\rho_\star}\frac{\mathrm{d}\rho_0}{\mathrm{d}z_r}(z_r) = g\left(\alpha\,\frac{\mathrm{d}\theta_0}{\mathrm{d}z_r} - \beta\,\frac{\mathrm{d}S_0}{\mathrm{d}z_r}\right)(z_r) \quad \text{(reference } N^2), \\
R_\rho &\equiv R_\rho(z_r) = \alpha\,\frac{\mathrm{d}\theta_0}{\mathrm{d}z_r}\bigg/ \beta\,\frac{\mathrm{d}S_0}{\mathrm{d}z_r} \quad \text{(density ratio)}, \\
\tau &\equiv \frac{\kappa_S}{\kappa_T} \quad \text{(diffusivity ratio)}.
\end{aligned}\right\}
\tag{2.16}
$$

Our choices of control parameters R_ρ and $\mathcal{M}$ differ from those of Middleton & Taylor (2020), who chose the gradient ratio $G_\rho \equiv \alpha\,|\nabla\theta|/(\beta\,|\nabla S|)$, and the angle ϕ between $\nabla\theta$ and ∇S defined by $\cos\phi = -\nabla\theta \cdot \nabla S/(|\nabla\theta|\,|\nabla S|)$ instead. The present approach is preferred here because R_ρ is a more commonly encountered parameter in the literature, while $\mathcal{M}$ is more naturally connected to the physics of lateral thermohaline intrusions, whose importance is well established. Moreover, K_{v0} represents a local version of the effective diffusivity concept originally proposed by Winters & D'Asaro (1996) and Nakamura (1996) based on thermal diffusion, and is positive definite by construction. Also, K_{v0} is different from the effective diffusivity K_{eff} introduced in § 5 and governing the evolution of the sorted density profile. Note that although the term within square brackets in (2.15) goes to infinity as $R_\rho \to 1$, ε_p remains finite as $N_0^2 \to 0$.

3. APE dissipation and diffusive stability

3.1. *Single-component fluid*

Even in the case of a single component fluid, the APE dissipation rate ε_p, unlike the viscous dissipation rate ε_k, is not guaranteed to be always positive definite. To see this, let us first

set $\mathcal{M} = 0$ and $\tau = 1$ in (2.15), leading to

$$\varepsilon_p \equiv \varepsilon_p^{std} = K_{v0} N_0^2(z_r) \left(1 - \frac{\partial_z z_r}{|\nabla z_r|^2}\right) = \kappa_T \left(|\nabla z_r|^2 - \frac{\partial z_r}{\partial z}\right) N_0^2. \tag{3.1}$$

Since κ_T and N_0^2 are positive by construction, (3.1) shows that ε_p can occasionally be negative provided that

$$\mathcal{F} \equiv |\nabla z_r|^2 - \frac{\partial z_r}{\partial z} = |\nabla_h z_r|^2 + \left(\frac{\partial z_r}{\partial z}\right)^2 - \frac{\partial z_r}{\partial z} < 0, \tag{3.2}$$

where ∇_h denotes the horizontal gradient. If viewed as a quadratic polynomial in $\partial z_r/\partial z$ with discriminant $\Delta = 1 - 4|\nabla_h z_r|^2$, $\mathcal{F}$ can be negative only when possessing two real roots, that is, when $\Delta < 0$, or equivalently when

$$|\nabla_h z_r| < \tfrac{1}{2}. \tag{3.3}$$

In that case, $\mathcal{F}$ and ε_p^{std} achieve their maximally negative values for $\partial z_r/\partial z = 1/2$:

$$|\varepsilon_{p,min}^{std}| = \kappa_T N_0^2 \left(\frac{1}{4} - |\nabla_h z_r|^2\right) < \frac{\kappa_T N_0^2}{4}, \tag{3.4}$$

which is bounded above by $\kappa_T N_0^2/4$, which is laminar-like in character and therefore extremely small relative to turbulent values. Nevertheless, because a negative ε_p can potentially cause perturbations to grow with time, it seems of interest to estimate the time scale $\tau_{APE} = e_a/\varepsilon_p$ of the latter. Using the fact that the APE density scales as $N_0^2 \zeta^2/2$, where $\zeta = z - z_r$, this leads to

$$\tau_{APE} = \frac{N_0^2 \zeta^2/2}{\kappa_T N_0^2/4} = \frac{2\zeta^2}{\kappa_T}. \tag{3.5}$$

For illustration, using $\kappa_T = 10^{-7} m^2 s^{-1}$ and $\zeta = 1$ cm $= 10^{-2}$ m yields $\tau = 2 \times 10^{-4}/10^{-7} = 2 \times 10^3$ s ≈ 33 min, which is fast enough to be observable in the laboratory, at least in principle. Inequality (3.3), which may be rewritten as $|\nabla_h \zeta| < 1/2$, also requires that the horizontal scales of the unstable perturbations be larger than their vertical scales. *A priori*, perturbations with the required characteristics cannot occur spontaneously; they can only be created with some form of appropriate gentle mechanical stirring. These considerations appear to be consistent with the experimental observation of steps formation by Park, Whitehead & Gnanadesikan (1994), who observed the smaller steps to grow faster than larger ones. However, these conclusions are only speculative at this stage, and more research is needed to confirm or disprove the relevance of our results.

3.2. *Binary fluid: spiceless case $\mathcal{M} = 0$*

In the double diffusive case but in the absence of spiciness, (2.15) reduces to

$$\varepsilon_p = K_{v0} N_0^2(z_r) \left(\frac{R_\rho - \tau}{R_\rho - 1}\right) \left(1 - \frac{\partial_z z_r}{|\nabla z_r|^2}\right). \tag{3.6}$$

This time, the sign of ε_p is controlled by the sign of the product of $\mathcal{F}$ times the density-ratio-dependent term within parentheses. Restricting attention to the case $\tau < 1$ pertaining to thermosolutal convection, table 1 shows that ε_p can be negative for all possible values of R_ρ, although not necessarily in both laminar and turbulent regimes.

994 A5-8

	$-\infty < R_\rho \le \tau$	$\tau < R_\rho \le 1$	$1 < R_\rho < +\infty$
$\mathcal{F} > 0$	$\varepsilon_p > 0$	$\boxed{\varepsilon_p < 0}$	$\varepsilon_p > 0$
$\mathcal{F} < 0$	$\boxed{\varepsilon_p < 0}$	$\varepsilon_p > 0$	$\boxed{\varepsilon_p < 0}$

Table 1. Sign of the APE dissipation rate ε_p as a function of the density ratio R_ρ and the sign of $\mathcal{F} = |\nabla z_r|^2 - \partial_z z_r$ in the absence of spiciness ($\mathcal{M} = 0$), in the case where $\tau = \kappa_S/\kappa_T < 1$, characteristic of salty water.

3.2.1. *Diffusive convection: $\tau < R_\rho < 1$*

In this case, ε_p is negative only if $\mathcal{F} > 0$, in which case the diapycnal component of the buoyancy flux is associated with up-gradient diffusion, as shown by Middleton & Taylor (2020). Note that the range $\tau < R_\rho < 1$ is associated with diffusive convection instability only for inviscid flows. In reality, stability analysis shows that viscosity restricts the range of instabilities to the range

$$\frac{Pr + \tau}{Pr + 1} < R_\rho < 1, \tag{3.7}$$

where $Pr = \nu/\kappa_T$ is the Prandtl number, with ν the molecular viscosity. For typical oceanic conditions, $\tau \approx 0.01$ and $Pr \approx 7$ yield the restricted range $0.876 < R_\rho < 1$.

3.2.2. *'Salt' finger instability: $R_\rho > 1$*

According to classical linear stability analysis, salt fingers are expected to develop in the density ratio range

$$1 < R_\rho < \frac{1}{\tau} \tag{3.8}$$

(e.g. Radko 2013). In that range, table 1 shows that APE dissipation can be negative only if $\mathcal{F} < 0$, and hence that the appropriate expression for ε_p is

$$\varepsilon_p = \kappa_T N_0^2 \left(|\nabla z_r|^2 - \frac{\partial z_r}{\partial z} \right) \left(\frac{R_\rho - \tau}{R_\rho - 1} \right) = \kappa_T N_0^2 \mathcal{F} \left(\frac{R_\rho - \tau}{R_\rho - 1} \right). \tag{3.9}$$

Equation (3.9) shows that negative APE dissipation in that case corresponds to the amplification of the negative APE dissipation in a single-component fluid multiplied by the amplification factor

$$\frac{R_\rho - \tau}{R_\rho - 1} > 1. \tag{3.10}$$

Since the time scale for the growth of an unstable perturbation is *a priori* inversely proportional to ε_p, the expectation is that the same perturbations in single-component and double diffusive fluids will grow faster in the former than in the latter. A key difference, however, is that in a double diffusive fluid, diffusive instabilities are able to grow spontaneously, whereas in a single-component fluid, their development requires initiation by means of external mechanical stirring of the right kind.

Since $\mathcal{F}$ can be negative only for laminar-like conditions, it follows that density-compensated thermohaline variations need to develop in order for the APE dissipation to remain negative as the flow transitions to turbulence. In fact, this is precisely what appears to happen in the salt finger experiments of Middleton & Taylor (2020), and as was previously advocated by Merryfield (2000).
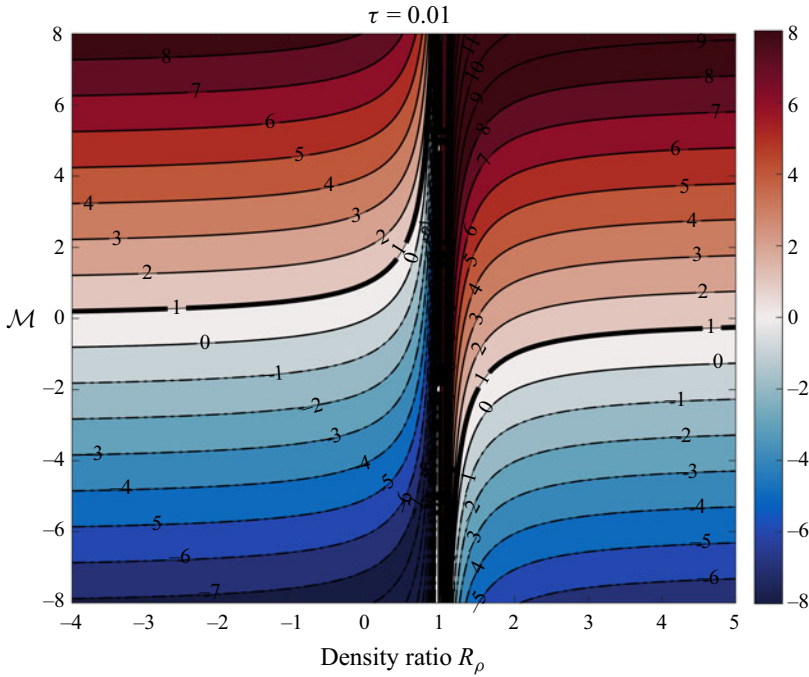
Figure 1. The amplification factor $f(R_\rho, \tau, \mathcal{M})$ (given by (3.12)) as a function of the spiciness parameter $\mathcal{M}$ and density ratio R_ρ, for $\tau = 0.01$.

3.3. *Impact of spiciness anomalies on diffusive instability*

Density-compensated thermohaline variations are a characteristic feature of ocean stratification in all regions and at all depths (Tailleux 2021). Consequently, spiciness and a non-zero value of $\mathcal{M}$ are likely to be representative of the general situation. Assuming a prevalence of the turbulent regime, we can make the approximation $1 - \partial_z z_r / |\nabla z_r|^2 \approx 1$, and the relevant expression for ε_p in the oceanic case appears as

$$\varepsilon_p = K_{v0} N_0^2(z_r) \left[\frac{R_\rho - \tau}{R_\rho - 1} + (1 - \tau)\mathcal{M} \right]. \tag{3.11}$$

This equation shows that the single-component APE dissipation rate $\varepsilon_p^{std} = K_{v0} N_0^2(z_r)$ is modulated by the factor

$$f(R_\rho, \tau, \mathcal{M}) = \frac{R_\rho - \tau}{R_\rho - 1} + (1 - \tau)\mathcal{M}. \tag{3.12}$$

To illustrate the effect of this factor, figure 1 depicts $f(R_\rho, \tau, \mathcal{M})$ as a function of R_ρ and $\mathcal{M}$ for $\tau = \kappa_S / \kappa_T = 0.01$. A notable difference from the spiceless case is the possibility of negative ε_p for a wide range of stratification, depending on the values of $\mathcal{M}$:

$$\mathcal{M} < -\frac{1}{1 - \tau} \frac{R_\rho - \tau}{R_\rho - 1}. \tag{3.13}$$

This opens up various scenarios, ranging from complete inhibition of mixing ($f = 0$) to enhanced mixing ($f > 1$) and even diffusive instability ($f < 0$). In the turbulent case, the

994 A5-10

value of $\mathcal{M}$ approximates to

$$\mathcal{M} = \frac{g}{2\rho_\star N_0^2} \frac{\nabla z_r \cdot \nabla \xi - \partial_z \xi}{|\nabla z_r|^2 - \partial_z z_r} \approx \frac{g}{2\rho_\star N_0^2} \frac{\nabla z_r \cdot \nabla \xi}{|\nabla z_r|^2} = -\frac{\nabla \xi \cdot \nabla \rho}{2 |\nabla \rho|^2}. \tag{3.14}$$

This reveals that for $\mathcal{M}$ to be negative, spiciness and density gradients need to be positively correlated. Exploring the possibility and magnitude of such a situation requires a way to express in terms of the large-scale gradients of ξ and ρ, which warrants further investigation beyond the scope of this study.

4. Comparison of APE frameworks

Holliday & McIntyre ([1981](#)) and Andrews ([1981](#)) had already shown the possibility of building Lorenz APE theory from a local principle by the time Winters *et al.* ([1995](#)) published their study in the mid-1990s. However, the local APE framework was still rudimentary and had not been used for any concrete applications. This situation has improved in recent years, as the local APE framework has progressed rapidly and achieved full maturity, being able to deal with a fully compressible multi-component stratified fluid (Tailleux [2018](#)). In this paper, we prefer the local APE framework to the global APE framework, because it is objectively simpler and more physical, even though this is not necessarily widely appreciated or acknowledged. This section intends to reveal some of the unphysical aspects of the global APE framework that are commonly overlooked, but easily uncovered by the local APE framework.

We start by reviewing the main results of the global APE framework as first derived by Winters *et al.* ([1995](#)). Following Lorenz ([1955](#)), Winters *et al.* ([1995](#)) (indicated by 'W95' in equations) defined the APE of a fluid as the potential energy of the actual state minus that of the adiabatically sorted state, i.e.

$$E_a^{W95} = \int_V \rho g z \, \mathrm{d}V - \int_V \rho_r g z_r \, \mathrm{d}V = \int_V \rho g (z - z_r) \, \mathrm{d}V, \tag{4.1}$$

and showed that its budget equation may be written in the form

$$\frac{\mathrm{d}E_a^{W95}}{\mathrm{d}t} = \underbrace{\int_V \rho g w \, \mathrm{d}V}_{\Phi_z} \underbrace{- g \oint_S (z\rho - \psi) \boldsymbol{v} \cdot \boldsymbol{n} \, \mathrm{d}S}_{S_{APE}^{adv}} + \underbrace{\kappa g \oint_S (z - z_r) \nabla \rho \cdot \boldsymbol{n} \, \mathrm{d}S}_{S_{APE}^{diff}} - (\Phi_d - \Phi_i),$$

$$\tag{4.2}$$

where Φ_z represents the conversion with kinetic energy, S_{APE}^{adv} represents the advective flux of APE through the boundaries enclosing the control volume considered, and S_{APE}^{diff} represents the diffusive flux of APE through the same boundaries. The expressions for Φ_i and Φ_d are given by

$$\Phi_i = -\kappa g A (\bar{\rho}_{top} - \bar{\rho}_{bottom}), \tag{4.3}$$

$$\Phi_d = -\kappa g \int_V \frac{\partial z_r}{\partial \rho} |\nabla \rho|^2 \, \mathrm{d}V, \tag{4.4}$$

where A is the cross-section of the domain, assumed constant and the overbar denotes surface average, while the quantity ψ entering the expression for the advective flux is

given by

$$\psi = \int^{\rho} z_r(\tilde{\rho}, t) \, d\tilde{\rho}. \tag{4.5}$$

For the most part, the derivation of (4.2) is straightforward, except for the demonstration that

$$\int_V \rho g \frac{\partial z_r}{\partial t} \, dV = 0, \tag{4.6}$$

which is non-trivial. A key point of this paper concerns the proper interpretation of Φ_i and Φ_d. Following Winters *et al.* (1995), Φ_d and Φ_i have been interpreted as representing two different types of energy conversion, with BPE and internal energy, respectively. The need for revisiting this interpretation is explained below.

The global APE budget (4.2) is now compared with that obtained by integrating the local APE density budget. This is done here for a two-constituent fluid with time-dependent Lorenz reference state, for which the local APE density takes the form

$$e_a = e_a(S, \theta, z, t) = -\int_{z_r}^{z} b(S, \theta, \tilde{z}, t) \, d\tilde{z} = \frac{g}{\rho_\star} \int_{z_r}^{z} [\rho(S, \theta) - \rho_0(\tilde{z}, t)] \, d\tilde{z}, \tag{4.7}$$

with the global APE being now defined as the volume-integrated APE density

$$E_a = \int_V \rho_\star e_a \, dV. \tag{4.8}$$

Because (4.7) implies that $\rho_\star e_a = \rho g(z - z_r) + p_0(z, t) - p_0(z_r, t)$, it follows that E_a can be regarded as the sum of E_a^{W95} plus an additional term related to the work against the vertical reference pressure gradient, as

$$E_a = E_a^{W95} + \int_V [p_0(z, t) - p_0(z_r, t)] \, dV. \tag{4.9}$$

Equation (4.9) shows that the validity of the global definition of APE E_a^{W95} requires the second term pertaining to the work against the reference pressure gradient to vanish. This is easily verified to be the case if the volume V enclosing the fluid remains unchanged under the isochoric adiabatic rearrangement $x \to x_r$ used to construct $\rho_0(z, t)$, as then $\partial(x, y, z)/\partial(x_r, y_r, z_r) = 1$, $dV = dV_r$, $V_r = V$, thus ensuring that

$$\int_V p_0(z_r, t) \, dV = \int_{V_r} p_0(z_r) \, dV_r = \int_V p_0(z, t) \, dV, \tag{4.10}$$

as requested (e.g. Roullet & Klein 2009; Tailleux 2013*b*; Winters & Barkan 2013). However, the identity $E_a = E_a^{W95}$ is not expected to necessarily hold for more complex situations, although the precise circumstances under which this happens have yet to be fully understood and discussed in the literature.

994 A5-12

For a time-dependent Lorenz reference state, the local APE budget (2.9) derived earlier is easily shown to be given by

$$\frac{\partial e_a}{\partial t} = -\boldsymbol{v} \cdot \nabla e_a - bw - \nabla \cdot \boldsymbol{J}_a - \varepsilon_p - \frac{g}{\rho_\star} \int_{z_r}^{z} \frac{\partial \rho_0}{\partial t}(\tilde{z}, t) \, \mathrm{d}\tilde{z}, \qquad (4.11)$$

where $\boldsymbol{J}_a = -(z - z_r)\boldsymbol{J}_b$ and $\varepsilon_p = \boldsymbol{J}_b \cdot \nabla(z - z_r)$ as before. By taking the volume integral of (4.11), the evolution equation for the global APE E_a is easily shown to be

$$\frac{\mathrm{d}E_a}{\mathrm{d}t} = \underbrace{- \int_V \rho_\star bw \, \mathrm{d}V}_{-\Phi_z} \underbrace{- \oint_S \rho_\star e_a \boldsymbol{v} \cdot \boldsymbol{n} \, \mathrm{d}S}_{S_{APE}^{adv}} + \underbrace{\oint_S \rho_\star(z - z_r)\boldsymbol{J}_b \cdot \boldsymbol{n} \, \mathrm{d}S}_{S_{APE}^{diff}} \underbrace{- \int_V \rho_\star \varepsilon_p \, \mathrm{d}V}_{\Phi_d - \Phi_i}, \quad (4.12)$$

where the expressions for Φ_i and Φ_d are

$$\Phi_i = -\int_V \rho_\star \varepsilon_{p,lam} \, \mathrm{d}V = \rho_\star g A [\alpha \kappa_T \, \Delta\theta - \beta \kappa_S \, \Delta S]$$

$$= \rho_\star g A \kappa_T \beta \, \Delta S \left[\frac{\alpha \, \Delta\theta}{\beta \, \Delta S} - \frac{\kappa_S}{\kappa_T} \right] = \rho_\star g A \kappa_T \beta \, \Delta S \, (R_\rho^{bulk} - \tau), \qquad (4.13)$$

$$\Phi_d = \int_V \rho_\star \varepsilon_{p,tur} \, \mathrm{d}V = \int_V g(\alpha \kappa_T \nabla\theta - \beta \kappa_S \nabla S) \cdot \nabla z_r \rho_\star \, \mathrm{d}V, \qquad (4.14)$$

where $\Delta\theta = \bar{\theta}_{top} - \bar{\theta}_{bottom}$, and $\Delta S = \bar{S}_{top} - \bar{S}_{bottom}$, while R_ρ^{bulk} is a bulk density ratio. For canonical salt finger instability, $1 < R_\rho^{bulk} < 1/\tau$, in which case (4.13) predicts $\Phi_i > 0$ as required to act as a source of APE. Note that the last term in (4.11) may be written in the form $F(z, t) - F(z_r, t)$, with F such that $(\partial F/\partial z)(z, t) = (\partial\rho_0/\partial t)(z, t)$. As a result, its volume integral must vanish whenever the second term of (4.9) also vanishes for the same reason, as explained above.

Winters & Barkan (2013) assumed that when E_a equals E_a^{W95}, their respective budgets should be equivalent and described by the equations for an open domain derived previously by Winters *et al.* (1995). We disagree with this view, however, because the budgets of E_a and E_a^{W95} are closely related to the local budgets of their integrands, but these are not equivalent, as established above. Physically, the non-equivalence of the budgets for E_a and E_a^{W95} is due to the integrand of E_a^{W95} lacking a vital piece of information about the energetics of stratified fluids compared to E_a, namely the fact that when a fluid parcel moves from its reference position z_r to its actual position, it has to work not only against gravity, but also against the vertical reference pressure gradient. More generally, Tailleux (2013b) shows that work against buoyancy forces actually involves three terms: (1) one associated with changes in gravitational potential energy; (2) one associated with changes in internal energy; and (3) one associated with work against the reference pressure gradient. Even if the latter term integrates to zero over the whole fluid domain, it is still necessary to retain it for correctly predicting the advective flux of APE through the boundaries of open domains. Two key differences characterise the non-equivalence of the budgets for E_a and E_a^{W95}: (1) in Winters *et al.* (1995), the reversible conversion with kinetic energy Φ_z is associated with the density flux $\rho g w$, but with the buoyancy flux $-bw$ in the local APE framework; (2) in Winters *et al.* (1995), the advective flux of APE through boundaries S_{APE}^{adv} is associated with the flux of the bizarre quantity $g(z\rho - \psi)$ but with the flux of APE density e_a in the local framework, which makes much more sense. The unphysical character of the advective flux of APE in Winters *et al.* (1995) highlights

the well-known problematic character of the global APE framework for regional studies, which was one of the key motivations for the quest of a local available energy framework (Tailleux 2013*a*).

The unphysical character of the advective flux of APE in the Winters *et al.* (1995) global APE budget for open domains is problematic and can be corrected only by using the local APE framework. The issue is vital, because recent studies of local energetics have demonstrated the key importance of the lateral advection of APE density for understanding important physical processes, such as storm track dynamics (Novak & Tailleux 2018; Federer *et al.* 2024), tropical cyclone intensification (Harris *et al.* 2022), or the energetics of the thermohaline circulation (Gregory & Tailleux 2011; Sijp *et al.* 2012). Importantly, it is worth noting that E_a and E_a^{W95} are also non-equivalent numerically. Indeed, in the global framework, the APE of the fluid is estimated as a small residual between two very large terms, which is ill-conditioned and prone to large numerical errors. In the local framework, however, the APE of a fluid is estimated as the sum of small non-negative numbers, which is numerically robust and very accurate. As a result, the field would benefit greatly from switching to the local APE framework, as this is the better and more physical-based practice.

5. Reference state variables

The practical implementation of the above theoretical results requires the computation of the Lorenz reference density profile $\rho_0(z, t)$, as well as of the associated isopycnal temperature and salinity averages necessary for defining spiciness. Here, we provide the mathematical foundations for defining the latter in terms of probability density functions (p.d.f.s), which allows one to compute these in practice without the use of sorting. These results pertain to arbitrary domains and unify the p.d.f. approaches of Tseng & Ferziger (2001) and Saenz *et al.* (2015), while also making use of previous ideas by Hochet & Tailleux (2019) and Hochet *et al.* (2021).

5.1. *Reference position $z_r(\rho, t)$ and density profile $\rho_0(z, t)$*

The classical interpretation of the Lorenz reference state as an isochoric rearrangement of fluid parcels is easy to understand and implement algorithmically. However, it does not easily lend itself to mathematical analysis. Here, we derive expressions for defining and constructing the Lorenz reference density profile $\rho_0(z, t)$ and its inverse function $z_r(\rho, t)$ (such that $\rho_0(z_r(\rho, t), t) = \rho$ at all times) in a way that greatly facilitates the study of its mathematical properties. Our approach is not entirely new, as elements of it can be found in Winters *et al.* (1995) or in the p.d.f. approaches of Tseng & Ferziger (2001), Saenz *et al.* (2015), Hochet & Tailleux (2019), Hochet *et al.* (2021) and Penney *et al.* (2020), but is useful to introduce the approaches discussed next. To that end, let us assume that the domain containing the fluid is bounded at the top by a rigid lid at $z = 0$, and at its bottom by a spatially varying bottom topography at $z = -H(x, y)$. We denote by $\hat{V}(z)$ the volume between an arbitrary depth z and the surface, and by $A(z) = -\mathrm{d}V(z)/\mathrm{d}z$ the cross-sectional area of the domain at depth z.

Physically, the reference depth $z = z_r(\rho, t)$ of a fluid parcel must be such that the volume of water between $z_r(\rho, t)$ and the surface, namely,

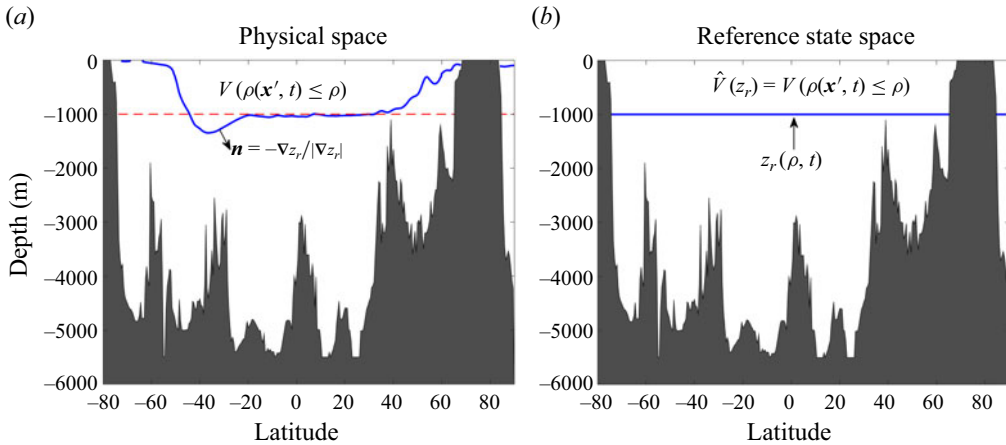$$\hat{V}(z_r) = \int_{z_r}^{0} A(z) \, \mathrm{d}z, \tag{5.1}$$

Figure 2. Mapping of (*a*) physical space onto (*b*) reference state space.

must equal the volume of water in physical space made up of all the fluid parcels of densities less than ρ, namely,

$$V(\rho, t) = \int_{V(\rho(\boldsymbol{x}', t) \le \rho)} \mathrm{d}V' = \int_V H(\rho - \rho(\boldsymbol{x}', t)) \, \mathrm{d}V', \tag{5.2}$$

where H denotes the Heaviside step function defined as in Winters *et al.* (1995):

$$H(\rho(\boldsymbol{x}, t) - \rho(\boldsymbol{x}_0, t)) = \begin{cases} 0, & \rho(\boldsymbol{x}, t) < \rho(\boldsymbol{x}_0, t), \\ \dfrac{1}{2}, & \rho(\boldsymbol{x}, t) = \rho(\boldsymbol{x}_0, t), \\ 1, & \rho(\boldsymbol{x}, t) > \rho(\boldsymbol{x}_0, t). \end{cases} \tag{5.3}$$

Equating the two volumes $\hat{V}(z_r) = V(\rho, t)$ yields an implicit equation for $z_r(\rho, t)$, which may be inverted to yield

$$z_r(\rho, t) = \hat{V}^{-1}[V(\rho, t)], \tag{5.4}$$

which in the case where $A(z) = \text{constant}$, reduces to

$$z_r(\rho, t) = -\frac{V(\rho, t)}{A}, \tag{5.5}$$

as derived previously by Winters *et al.* (1995). As an illustration, figure 2(*a*) shows the constant Lorenz reference depth $z_r(\rho, t) = -1000$ m for a realistic ocean, with figure 2(*b*) showing the corresponding position in reference state space. Construction of the Lorenz reference state for a realistic ocean has been discussed in Saenz *et al.* (2015) and Tailleux (2016, 2021). Once $z_r(\rho, t)$ has been determined for a number of target densities, the reference density profile $\rho_0(z, t)$ may be obtained simply by solving the implicit relation $\rho_0(z_r(\rho, t), t) = \rho$ at a number of discrete reference depths, from which an interpolator may be constructed to infer the reference density for other values. Alternatively, using the fact that $\rho = \rho_0(z_r, t)$ by definition, one may also solve the equation $\hat{V}(z_r) = V(\rho, t)$ in the form $\hat{V}(z_r) = V(\rho_0(z_r, t), t)$ to infer $\rho_0(z_r, t)$ at a number of predetermined target reference depths.
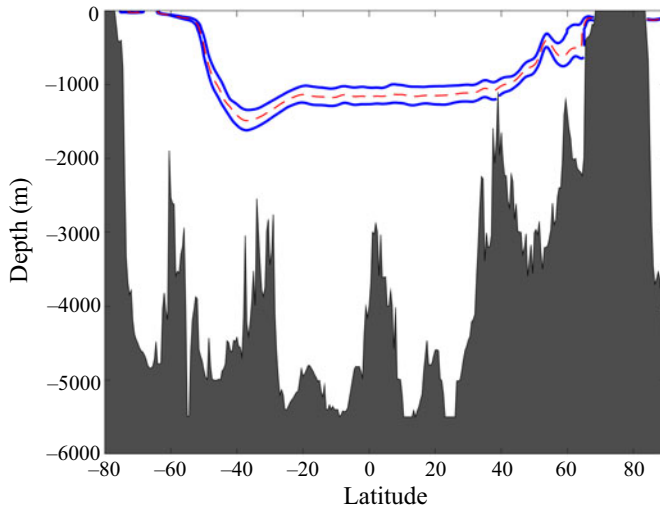
Figure 3. Schematics of an isopycnal surface (dashed red line) bounded by two nearby isopycnal surfaces (thick blue lines) defining the volume of fluid used to define the thickness-weighted average value of any arbitrary tracer for the isopycnal surface considered.

Note that by differentiating the equality $\hat{V}(z_r) = V(\rho, t) = V(\rho_0(z_r, t), t)$ with respect to z_r, the following useful expression for $\partial \rho_0 / \partial z_r$ may be obtained:

$$\frac{\partial \rho_0}{\partial z}(z_r, t) = -\left(\frac{\partial V}{\partial \rho}\right)^{-1} A(z_r). \tag{5.6}$$

The above expressions establish that the practical computation of $z_r(\rho, t)$ and $\rho_0(z, t)$ can be achieved without resorting to a sorting algorithm. Indeed, all one has to do is to define a number of target densities ρ_i, $i = 1, \ldots, N$, at which to compute the corresponding volumes $V(\rho_i, t)$. Having defined the inverse function $\hat{V}^{-1}(z)$, one may simply compute the corresponding target reference depths $z_r(\rho_i, t)$ at which $\rho_0(z, t)$ may be evaluated. Once all quantities have been determined at the discrete target depths, interpolators can be constructed to find values at other points.

5.2. *Construction of isopycnal mean profiles*

Physically, the isopycnally averaged temperature and salinity profiles $\theta_0(z, t)$ and $S_0(z, t)$ needed to isolate the passive spice component need to be obtained as thickness-weighted isopycnal averages, as illustrated in figure 3, as this is the only physically meaningful way to define an isopycnal average (Young 2012). Mathematically, $\theta_0(z, t)$ can be defined as a solution of the equation

$$\int_{z_r(\rho, t)}^{0} A(z) \, \theta_0(z, t) \, \mathrm{d}z = \underbrace{\int_{V(\rho, t)} \theta(\boldsymbol{x}, t) \, \mathrm{d}V}_{H(\rho, t)}. \tag{5.7}$$

Physically, the right-hand side of (5.7) represents the volume integrated temperature over the body of water made up of fluid parcels with densities less than ρ, which we denote $H(\rho, t)$. The right-hand side is the volume integral of the thickness-weighted isopycnally
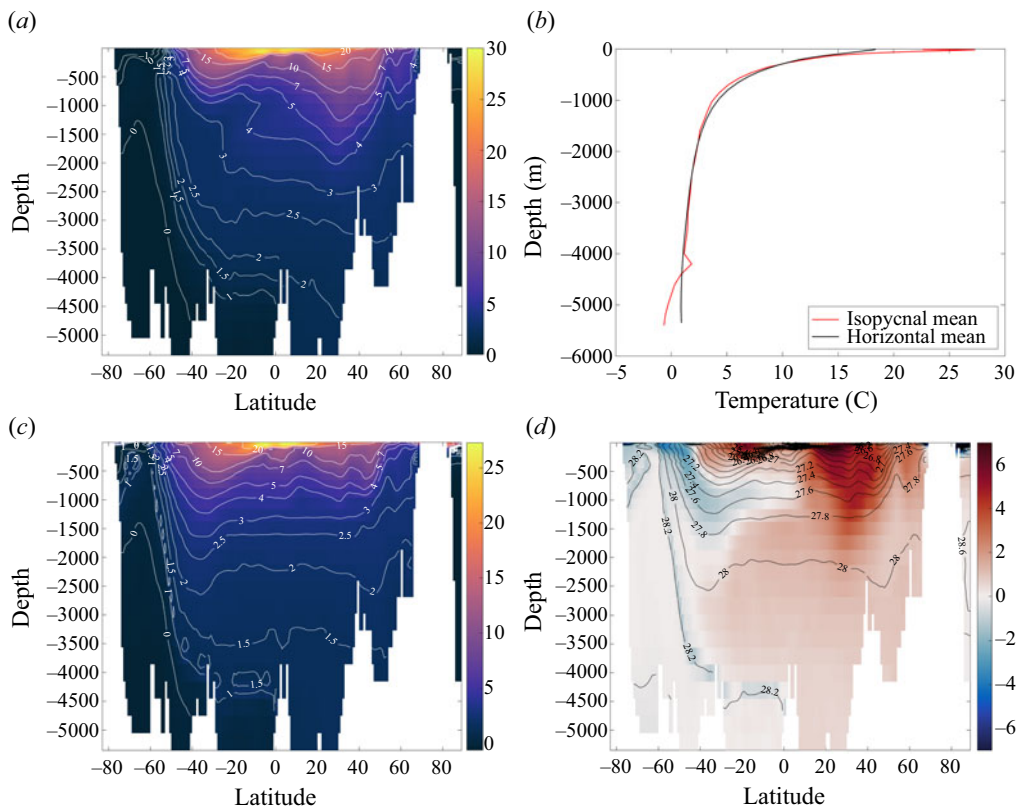
994 A5-16

Figure 4. (*a*) Latitude/depth section of Conservative Temperature along $30°$W in the Atlantic Ocean. (*b*) Global isopycnal mean (red line) profile used for the spice/heave decomposition of Conservative Temperature shown in (*c*) and (*d*). For comparison, a horizontal mean profile (black line) is also depicted. (*c*) Active heave component of Conservative Temperature along the same $30°$W section as in (*a*). (*d*) Passive spice component of Conservative Temperature along the same $30°$W section as in (*a*). Black contour lines represent constant Lorenz reference density surfaces labelled in terms of thermodynamic neutral density γ^T.

averaged temperature between $z_r(\rho, t)$ and the surface. An explicit expression for $\theta_0(z, t)$ at $z = z_r$ may be obtained by differentiating (5.7) with respect to z_r, which leads to

$$-A(z_r)\,\theta_0(z_r, t) = \frac{\partial H}{\partial \rho}\,\frac{\partial \rho_0}{\partial z_r}(z_r, t) \quad \Rightarrow \quad \theta_0(z_r, t) = \left(\frac{\partial V}{\partial \rho}\right)^{-1}\frac{\partial H}{\partial \rho}, \qquad (5.8)$$

which made use of (5.6).

Equation (5.8) is easily implemented in practice, at least for the type of data considered here. As an illustration, figure 4 shows a decomposition of temperature into active and passive components. Specifically, figure 4(*a*) shows a meridional section of Conservative Temperature along $30°$W in the Atlantic Ocean. Figure 4(*b*) shows the thickness-weighted isopycnal temperature average profile compared with the more standard horizontal mean profile. The thickness-weighted averaged temperature is illustrated in figure 4(*c*), while figure 4(*d*) illustrates the passive spice component, obtained as $\theta_{spice} = \theta - \theta_0(z_r, t)$. Our approach to defining spice, based on Tailleux (2021), is strongly supported by the fact that the patterns in figure 4(*d*) all appear to coincide with known ocean water masses in the Atlantic Ocean, namely Antarctic Bottom Water, North Atlantic Deep Water, Antarctic Intermediate Water and Mediterranean Intermediate Water.

5.3. *Evolution of the reference density profile*

By definition, $\rho_0(z, t)$ must coincide with its thickness-weighted isopycnal average and therefore satisfy (5.7), namely,

$$\int_{z_r}^0 A(z)\, \rho_0(z, t)\, \mathrm{d}z = \underbrace{\int_{V(\rho, t)} \rho(\boldsymbol{x}', t)\, \mathrm{d}V'}_{R(\rho, t)}. \tag{5.9}$$

We now show that this equation can be manipulated to yield an exact expression for the evolution equation satisfied by $\rho_0(z, t)$ as well as for the effective diffusivity first introduced by Winters & D'Asaro (1996) and Nakamura (1996). For simplicity, we focus only on the case where the buoyancy flux vanishes through all solid boundaries as well as the surface, but the derivation can easily be extended to account for a non-zero surface buoyancy flux.

To obtain the sought-for evolution equation for ρ_0, let us take the time derivative of (5.9), which yields

$$\int_{z_r}^0 A(z)\, \frac{\partial \rho_0}{\partial t}(z, t)\, \mathrm{d}z = \int_{V(\rho, t)} \frac{\partial \rho}{\partial t}(\boldsymbol{x}', t)\, \mathrm{d}V'. \tag{5.10}$$

Now, using the fact that the evolution equation for density may be written as

$$\frac{\partial \rho}{\partial t} = -\boldsymbol{v} \cdot \boldsymbol{\nabla}\rho + \boldsymbol{\nabla} \cdot (\rho_\star \boldsymbol{J}_b/g), \tag{5.11}$$

(5.10) may be rewritten as

$$\begin{aligned}
\int_{z_r}^0 A(z)\, \frac{\partial \rho_0}{\partial t}(z, t)\, \mathrm{d}z &= \int_{V(\rho, t)} [\boldsymbol{\nabla} \cdot (\rho_\star \boldsymbol{J}_b/g) - \boldsymbol{v} \cdot \boldsymbol{\nabla}\rho]\, \mathrm{d}V' \\
&= \oint_{S_\rho} (\rho_\star/g) \boldsymbol{J}_b \cdot \boldsymbol{n}\, \mathrm{d}S = \oint_{S_\rho} (\rho_\star/g) \boldsymbol{J}_b \cdot \frac{\boldsymbol{\nabla}\rho}{|\boldsymbol{\nabla}\rho|}\, \mathrm{d}S \\
&= \frac{\partial \rho_0}{\partial z}(z_r, t) \oint_{S_\rho} (\rho_\star/g) \frac{\boldsymbol{J}_b \cdot \boldsymbol{\nabla}z_r}{|\boldsymbol{\nabla}\rho|}\, \mathrm{d}S \\
&= -\frac{\partial \rho_0}{\partial z}(z_r, t) \oint_{S_\rho} (\rho_\star/g) \frac{\varepsilon_{p,tur}}{|\boldsymbol{\nabla}\rho|}\, \mathrm{d}S.
\end{aligned} \tag{5.12}$$

In the second line, S_ρ refers to the isopycnal surface $\rho = \text{const.}$, meaning that the outward normal unit vector is $\boldsymbol{n} = \boldsymbol{\nabla}\rho/|\boldsymbol{\nabla}\rho| = -\boldsymbol{\nabla}z_r/|\boldsymbol{\nabla}z_r|$, as illustrated schematically in figure 2(*a*). To go from the second line to the third line, we used the fact that as $\rho = \rho_0(z_r, t)$, we have $\boldsymbol{\nabla}\rho = (\partial \rho_0/\partial z_r)(z_r, t) \boldsymbol{\nabla}z_r$. To go from the third line to the fourth line, we used the fact that by definition, $\boldsymbol{J}_b \cdot \boldsymbol{\nabla}z_r = -\varepsilon_{p,tur}$.

By defining the reference squared buoyancy profile as

$$N_0^2(z_r, t) = -\frac{g}{\rho_\star} \frac{\partial \rho_0}{\partial z}(z_r, t), \tag{5.13}$$

and the locally defined effective diffusivity as

$$\kappa_{eff} = \frac{\varepsilon_{p,turb}}{N_0^2(z_r)}, \tag{5.14}$$

(5.12) may be rewritten as

$$\int_{z_r}^{0} A(z) \frac{\partial \rho_0}{\partial t}(z, t)\, \mathrm{d}z = -A(z_r)\, K_{eff}\, \frac{\partial \rho_0}{\partial z_r}(z_r, t) = \frac{\partial R}{\partial t}, \tag{5.15}$$

where the globally defined effective diffusivity is related to the locally defined one by

$$K_{eff} = \frac{1}{A(z_r)} \oint_{S_\rho} \frac{\kappa_{eff}}{|\boldsymbol{\nabla} z_r|}\, \mathrm{d}S, \tag{5.16}$$

as established previously by Hochet *et al.* (2019). The final step of the derivation consists in differentiating (5.15) with respect to z_r, which yields

$$\frac{\partial \rho_0}{\partial t}(z_r, t) = \frac{1}{A(z_r)} \frac{\partial}{\partial z_r} \left[A(z_r)\, K_{eff}\, \frac{\partial \rho_0}{\partial z_r} \right](z_r, t). \tag{5.17}$$

Note that in practice, K_{eff} can also be determined by inverting the diffusion equation (5.17) from the knowledge of the temporal variations of the sorted density profile $\rho_0(z, t)$. Alternatively, one may combine (5.6) and (5.15) to verify that the globally defined effective diffusivity may be written as

$$K_{eff}(z_r, t) = \frac{1}{A^2(z_r)} \frac{\partial V}{\partial \rho}(\rho, t)\, \frac{\partial R}{\partial t}(\rho, t), \tag{5.18}$$

which in principle allows one to calculate K_{eff} using simple binning techniques. The above results establish that the effective diffusivity depends only on the turbulent part $\varepsilon_{p,tur}$ of ε_p, consistent with established results.

6. Summary, discussion and conclusions

Double diffusive instabilities are mysterious as they can seemingly develop from a state with zero initial mechanical energy, without the need for any external mechanical source(s) of energy (except for that of the initial perturbation triggering the instability). In the context of APE theory, such instabilities are possible only if the mechanical energy KE + APE can grow at the expense of BPE. Since viscous dissipation ε_k is always positive, this can happen only when the APE dissipation rate ε_p is negative. Therefore, it is more logical and physically plausible to consider the sign of the APE dissipation rate $\varepsilon_p = \boldsymbol{J}_b \cdot \boldsymbol{\nabla}(z - z_r)$, rather than the sign of the diapycnal component of the buoyancy flux proposed by Middleton & Taylor (2020), as the most fundamental criterion for characterising double diffusive instabilities. In the context of the Winters *et al.* (1995) framework, this is equivalent to predicting the development of double diffusive instabilities from the sign of $\Phi_d - \Phi_i$ rather than that of Φ_d only. This is justified by the fact that Φ_d is negative only for diffusive convection instability, not for salt finger instability, while $\Phi_d - \Phi_i$ is negative for both.

While our current findings unequivocally acknowledge the existence and necessity of negative APE dissipation in single- and two-component stratified fluids, the precise underlying mechanisms remain elusive. The controversy revolves primarily about the nature of BPE and of the energy conversion type between APE and BPE. The primary challenge stems from the fact that in fully compressible fluids, APE dissipation appears naturally as a conversion between APE and the internal energy component of BPE, akin to viscous dissipation (Tailleux 2009, 2013c). However, this interpretation is not the most intuitive within standard Boussinesq models, as the BPE lacks an explicit link to internal

energy, thereby complicating our understanding of its energetics. Nevertheless, the crucial role of internal energy in understanding the energetics of turbulent stratified fluids is underscored by the term $gz\,\mathrm{D}\rho/\mathrm{D}t$, which approximates the compressible work $p\,\mathrm{D}\upsilon/\mathrm{D}t$ in Boussinesq models (Young 2010; Tailleux 2012), implying a conversion with internal energy. While some may find considering thermodynamics central to the understanding of the energetics of turbulent stratified fluids unpalatable, it is noteworthy that in the current state of knowledge, there does not appear to be any other way to explain the occurrence of negative APE dissipation in single-component fluids or doubly stable two-component fluids. Indeed, existing explanations for the release of BPE into APE by Smyth, Nash & Moum (2005) and Ma & Peltier (2021) have linked it to the 'APE' associated with a destabilising stratifying tracer. However, this approach is difficult to justify theoretically, because in reality, heat and salt can be separated only diabatically (due to intermolecular forces). Another difficulty is that the approach breaks down in the limit $\kappa_S = \kappa_T$, which reduces to the case of a simple fluid, or for doubly stable stratified fluids. On the other hand, the thermodynamic theory of exergy offers a potentially more promising avenue. Exergy theory posits that because internal energy is convex with respect to its canonical variables (specific entropy, salinity and specific volume for two-component seawater), a portion of it – referred to as exergy, proportional to the deviation from thermodynamic equilibrium – is available for conversions into work (Bannon 2013; Tailleux 2013*a*; Bannon & Najjar 2014) under diabatic transformation, regardless of whether the stratifying tracers are stabilising or destabilising. For single-component fluids, exergy is directly proportional to temperature variance at leading order, providing a means to identify and quantify the portion of BPE that can be converted into APE in simple and double diffusive flows. However, this approach necessitates accepting APE dissipation as a conversion between APE and the internal energy component of BPE, thereby challenging the prevailing assumption. Pursuing this line of enquiry is beyond the scope of this paper and is deferred to future studies. We anticipate that this will require the development of a new class of Boussinesq models with explicit and thermodynamically consistent internal energy. The development of such models, building upon the recent findings of Tailleux & Dubos (2024), is currently underway and will be presented in a subsequent study.

In this study, the local APE framework was preferred over the Winters *et al.* (1995) global APE framework owing to its greater simplicity, physical justification, and flexibility. Despite its popularity, the global APE framework has several unphysical aspects that can sometimes lead to erroneous conclusions and/or interpretations. Examples include its unsuitability for regional studies in open domains due to the unphysical character of its advective fluxes, and the inability to meaningfully decompose the sign-indefinite integrand $\rho g(z - z_r)$ into mean and eddy components. These issues, which tend to be overlooked in the turbulent mixing community, can be corrected only by using the local APE framework. While the non-intuitive character of the early formulations of the local APE theory by Andrews (1981), Holliday & McIntyre (1981) and Shepherd (1993) made it somewhat difficult to use or apply, the most recent formulations have greatly clarified it, and several applications have since demonstrated its usefulness and ease of use, e.g. Roullet & Klein (2009), Zemskova, White & Scotti (2015), Novak & Tailleux (2018) and Harris & Tailleux (2018). Given that the local APE framework has made the global APE framework largely obsolete, there is no longer any real justification for continuing to use the global framework, especially as it is obviously unsuitable for the study of individual turbulent mixing events in large domains such as that pertaining to the oceans. We hope that the simplicity of our derivations, which we believe to be more intuitive and transparent than those of Scotti & White (2014), will convince some readers to switch frameworks.

For a single-component fluid, the Boussinesq APE dissipation rate $\varepsilon_p = \boldsymbol{J}_b \cdot \boldsymbol{\nabla}(z - z_r)$ can be shown to approximate the quantity

$$\rho \varepsilon_p = \kappa c_p \, \boldsymbol{\nabla} T \cdot \boldsymbol{\nabla} \left(\frac{T - T_r}{T} \right) \tag{6.1}$$

(Tailleux 2009, 2013c). Equation (6.1) features a Carnot-like thermodynamic efficiency $(T - T_r)/T$, underscoring the inherent thermodynamic character of turbulent stratified mixing and APE dissipation, which is hidden by the Boussinesq approximation. Tailleux (2013c) demonstrated that (6.1) could be broken down into three parts (see Appendix A for a simpler and clearer derivation of the result): a laminar component $\varepsilon_{p,lam}$ related to the term $-\Phi_i$ in Winters *et al.* (1995); a turbulent term $\varepsilon_{p,turb}$ related to the term Φ_d in Winters *et al.* (1995); and a term $\varepsilon_{p,eos}$ that arises from the nonlinearities of the equation of state and has no equivalent in the standard Boussinesq model and Winters *et al.* (1995). As far as we understand the issue, these components are indissociable parts of ε_p and therefore necessarily of the same type, thus implying that Φ_d and Φ_i should both be interpreted as conversions between APE and internal energy, in contrast to the prevailing interpretation. In the literature, Φ_i and $\varepsilon_{p,lam}$ are often neglected based on the assumption that they are negligible in fully turbulent flows. However, this may not necessarily be the case, as Φ_i appears to be a significant term in many published studies. For example, Φ_i is comparable to Φ_d in the simulations discussed by Winters *et al.* (1995). The present theory predicts that Φ_i should also be dominant in the early stages of the canonical salt finger instability experiment discussed by Middleton & Taylor (2020); it is therefore unfortunate that the authors did not diagnose Φ_i in their experiments, as this would undoubtedly have been useful to demonstrate its fundamental importance. We hope to illustrate this behaviour in the future by using an energy complete Boussinesq approximation currently under development, based on the findings of Tailleux & Dubos (2024).

Like Middleton & Taylor (2020), we find that three physical parameters are sufficient to fully describe the behaviour of Φ_d or ε_p, but apart from the diffusivity ratio κ_S/κ_T, how to chose the two remaining parameters is not unique. The parameters chosen by Middleton & Taylor (2020) are strongly affected by turbulence and the microstructures of θ and S, so diagnosing them from observational data or environmental parameters, or even linking them to parameters used in previous studies (such as the density ratio), is not straightforward. To overcome some of these difficulties, Middleton *et al.* (2021) subsequently reformulated the problem in terms of density/spiciness variable, using a linear spiciness variable to achieve a more physical and useful characterisation of Φ_d. However, this spiciness variable is arguably suboptimal for measuring spiciness because it is not properly calibrated to vanish in a spiceless fluid. For this reason, Tailleux (2021) recommended that spiciness be defined as an isopycnal anomaly instead. On this basis, we proposed a new set of physical parameters that we used to characterise ε_p: a density ratio and a new spiciness parameter more in line with the parameters used in previous studies. However, estimating these parameters in numerical simulations or in the field requires the computation of thickness-weighted isopycnal averages, with which most non-oceanographers are likely to be unfamiliar. For this reason, and to facilitate the adoption of our ideas, we provided explicit mathematical expressions and practical examples to show how to implement this form of averaging in practice, valid even for a fully compressible and nonlinear ocean, and illustrated it on a realistic oceanographic example with real data.

While the presence of double diffusive instabilities is determined by the sign of the full APE dissipation rate $\varepsilon_p = \varepsilon_{p,lam} + \varepsilon_{p,tur}$, the effective diapycnal diffusivity controlling

the evolution of the sorted density profile is determined only by the turbulent part $\varepsilon_{p,tur}$. Consequently, the conventional expressions for turbulent diapycnal diffusivity and dissipation ratio should be formulated as $K_v = \varepsilon_{p,turb}/N^2$ and $\Gamma = \varepsilon_p/\varepsilon_k = \varepsilon_{p,turb}/\varepsilon_k$ rather than in terms of the full ε_p. This is consistent with common practice, but it may not necessarily be obvious if one considers that $\varepsilon_{p,lam}$ (equivalently Φ_i) is part of the definition of the net APE dissipation rate, which departs from the prevailing view that only $\varepsilon_{p,turb}$ or Φ_d contributes to it. As a consequence, not all double diffusive instabilities can be expected to exhibit negative effective diffusivities, which, for instance, are not present in the initial stages of the canonical salt finger instability discussed by Middleton & Taylor (2020). Our theory shows that double diffusion can enhance, reduce or even change the sign of K_v, whose consequences for the study of mixing in the oceans need to be investigated further. The implications for the theoretical understanding of thermohaline staircases is unclear, as recent theories tend to emphasise the spatial variations (linked to the functional dependence on some turbulent parameters such as the buoyancy Reynolds number) of an otherwise positive effective diffusivity as the main cause; e.g. see Radko (2005), Taylor & Zhou (2017) or Ma & Peltier (2022*a*,*b*) and references therein for some recent discussion. Further research is clearly needed to clarify the issue, but is beyond the scope of this paper.

Declaration of interests. The author reports no conflict of interest.

Data availability statement. Matlab software to reproduce the results and figures of this paper are openly available in the Zenodo repository at https://doi.org/10.5281/zenodo.12544932, cited as Tailleux (2024).

Author ORCIDs.
R. Tailleux https://orcid.org/0000-0001-8998-9107.

Appendix A. APE dissipation in compressible and Boussinesq fluids

In contrast to the concept of viscous dissipation, whose origin can be traced back to the development of the Navier–Stokes equations in the early 19th century, the concept of APE dissipation is comparatively much more recent, as it appears to have been developed in a somewhat *ad hoc* way from a re-scaling of the dissipation of temperature variance by Gargett & Holloway (1984) and Oakey (1982). In particular, it was not obtained as part of a local APE budget, even though the possibility of constructing Lorenz (1955) APE theory from a local principle had been established by Andrews (1981) and Holliday & McIntyre (1981). Local balance equations for a local APE density were perhaps first derived by Molemaker & McWilliams (2010). As far as we are aware, a theoretical formulation of APE dissipation for a fully compressible fluid clarifying how it relates to its Boussinesq counterpart was developed only much more recently by Tailleux (2009, 2013*c*). Here, we revisit our earlier derivation to provide a somewhat simpler treatment, in which dependence of our expressions in terms of entropy is replaced by a dependence on potential temperature. Concretely, the task here is to understand how to relate the expression for the

APE dissipation rate for a fully compressible fluid, i.e.

$$\varepsilon_p = \kappa_T c_p \, \boldsymbol{\nabla} T \cdot \boldsymbol{\nabla} \Upsilon = -\kappa_T c_p \, \boldsymbol{\nabla} T \cdot \boldsymbol{\nabla} \left(\frac{T_r}{T} \right), \tag{A1}$$

to the terms Φ_i and Φ_d that appear in the Winters *et al.* (1995) global APE framework, where $T_r = T(\eta, p_r)$ and $T = T(\eta, p)$ are the *in situ* temperatures at the reference pressure p_r and actual pressure p, respectively, with $\Upsilon = (T - T_r)/T$ the thermodynamic efficiency.

As in Tailleux (2013*a*), we take as our starting point the following thermodynamic relation for the total differential of specific entropy η regarded as a function of either (T, p) or θ, namely,

$$d\eta = \frac{c_p}{T} \, dT - \frac{\alpha}{\rho} \, dp = \frac{c_{p\theta}}{\theta} \, d\theta, \tag{A2}$$

in which η is the specific entropy, α is the thermal expansion coefficient, c_p is the specific heat capacity at constant pressure, $\theta = T(\eta, p_a)$ is the potential temperature (i.e. the temperature referenced to the mean surface atmospheric pressure p_a), and $c_{p\theta} = c_p(\theta, p_a)$. The thermodynamic relation (A2) can be rearranged to yield an expression for the total differential of the logarithm of the temperature $\ln T$, namely,

$$\frac{dT}{T} = \frac{\alpha}{\rho c_p} \, dp + \frac{c_{p\theta}}{c_p} \frac{d\theta}{\theta}, \tag{A3}$$

thus allowing one to regard $T = T(\theta, p)$ as a function of potential temperature and pressure. Because dT/T is an exact differential, (A3) implies the following Maxwell relationship (i.e. the equality of the cross-derivatives):

$$\frac{\partial}{\partial \theta} \left(\frac{\alpha}{\rho c_p} \right) = \frac{c_{p\theta}}{\theta} \frac{\partial}{\partial p} \left(\frac{1}{c_p} \right), \tag{A4}$$

which will prove useful in the following. For the cases of interest, the reference temperatures entering calculations are 'potential temperatures' of some kind referenced to an arbitrary pressure p_r. It follows that if we denote $T_r = T(\theta, p_r)$, then we may write

$$\ln \frac{T_r}{T} = \int_p^{p_r} \frac{\alpha}{\rho c_p} (\theta, \tilde{p}) \, d\tilde{p}. \tag{A5}$$

Taking the gradient, we then obtain

$$\frac{T}{T_r} \boldsymbol{\nabla} \frac{T_r}{T} = -\frac{\alpha}{\rho c_p} (\theta, p) \, \boldsymbol{\nabla} p + \left[\int_p^{p_r} \frac{\partial}{\partial \theta} \left(\frac{\alpha}{\rho c_p} \right) (\theta, \tilde{p}) \, d\tilde{p} + \frac{\alpha}{\rho c_p} (\theta, p_r) \frac{dp_r}{d\theta} \right] \boldsymbol{\nabla} \theta. \tag{A6}$$

Using the above Maxwell relationship (A4) gives

$$\int_p^{p_r} \frac{\partial}{\partial \theta} \left(\frac{\alpha}{\rho c_p} \right) (\theta, \tilde{p}) \, d\tilde{p} = \frac{c_{p\theta}}{\theta} \int_p^{p_r} \frac{\partial}{\partial p} \left(\frac{1}{c_p} \right) (\theta, \tilde{p}) \, d\tilde{p} = \frac{c_{p\theta}}{\theta} \left(\frac{1}{c_{pr}} - \frac{1}{c_p} \right), \tag{A7}$$

so that (A6) may be rewritten as

$$\boldsymbol{\nabla} \left(\frac{T_r}{T} \right) = -\frac{\alpha T_r}{\rho c_p T} \boldsymbol{\nabla} p + \frac{\alpha_r T_r}{\rho_R c_{pr} T} \boldsymbol{\nabla} p_r + \frac{c_{p\theta} T_r}{T \theta} \left(\frac{1}{c_{pr}} - \frac{1}{c_p} \right) \boldsymbol{\nabla} \theta, \tag{A8}$$

with $c_{pr} = c_p(\theta, p_r)$, $\alpha_r = \alpha(\theta, p_r)$ and $\rho_r = \rho(\theta, p_r)$. Note here that the notation $Q = Q(\theta, p)$ is an abuse of notations for the function $Q = Q(T(\theta, p), p)$, i.e. it is computed

from the expressions in terms of *in situ* temperature and pressure in which T is replaced by its expression as a function of potential temperature and pressure.

The above results may be used to express the non-viscous non-conservation term ε_p as the sum

$$\varepsilon_p = \varepsilon_{p,lam} + \varepsilon_{p,tur} + \varepsilon_{p,eos}. \tag{A9}$$

Moreover, if one uses the fact that $p_r = p_0(z_r(\theta))$ so that $\nabla p_r = -\rho_r g \, (\mathrm{d}z_r/\mathrm{d}\theta) \nabla\theta$, then one may write

$$\varepsilon_{p,lam} = \frac{\alpha T_r}{\rho T} \kappa_T \nabla T \cdot \nabla p, \tag{A10}$$

$$\varepsilon_{p,tur} = \frac{c_p T_r \alpha_r g}{c_{pr} T} \frac{\mathrm{d}z_r}{\mathrm{d}\theta} \kappa_T \nabla T \cdot \nabla\theta, \tag{A11}$$

$$\varepsilon_{p,eos} = -\frac{c_{p\theta} T_r}{c_{pr} T} \frac{c_p - c_{pr}}{\theta} \kappa_T \nabla T \cdot \nabla\theta. \tag{A12}$$

Importantly, the term $\varepsilon_{p,lam}$ vanishes only at standard thermodynamic equilibrium, characterised by uniform *in situ* temperature T. However, both $\varepsilon_{p,tur}$ and $\varepsilon_{p,eos}$ vanish at standard thermodynamic equilibrium, but also for the 'turbulent' thermodynamic equilibrium characterised by uniform potential temperature θ.

A.1. *Link with the Boussinesq approximation*

The above expressions show that the exact counterparts of the Φ_i term in Winters *et al.* (1995) are the expressions

$$\Phi_i^{exact} = -\int_V \frac{\alpha T_r}{T} \kappa_T \nabla T \cdot \nabla p \, \mathrm{d}V, \tag{A13}$$

$$\Phi_d^{exact} = \int_V \frac{c_p T_r \rho \alpha_r g}{c_{pr} T} \frac{\mathrm{d}z_r}{\mathrm{d}\theta} \kappa_T \nabla T \cdot \nabla\theta \, \mathrm{d}V. \tag{A14}$$

In the Boussinesq approximation, $p \approx -\rho_\star g z$, $T_r \approx T$ and α is treated as a constant. As a result,

$$\Phi_i^{exact} \approx \int_V \rho_\star \alpha g \kappa_T \frac{\partial T}{\partial z} \, \mathrm{d}V = \rho_\star \alpha g \kappa_T (\langle T \rangle_{top} - \langle T \rangle_{bottom}) = \Phi_i. \tag{A15}$$

As is well known, this term integrates to a term that depends on the difference between the surface integrated temperature at the upper boundary minus that at the lower boundary. For a single-component fluid, Φ_i is generally positive for a stable stratification. In particular, Φ_i is independent of the turbulent character of the fluid. It is important that this is not the case of the exact expression. An important open question is whether strong turbulence could potentially make Φ_i^{exact} significantly different from Φ_i.

Regarding the Φ_d term,

$$\Phi_d^{exact} \approx \int_V \rho_\star \alpha g \frac{\mathrm{d}z_r}{\mathrm{d}\theta} \kappa_T |\nabla\theta|^2 \, \mathrm{d}V = -\int_V g \frac{\mathrm{d}z_r}{\mathrm{d}\rho} \kappa_T |\nabla\rho|^2 \, \mathrm{d}V, \tag{A16}$$

assuming $\mathrm{d}\rho = -\rho_\star \alpha \, \mathrm{d}\theta$. The Boussinesq approximation of Φ_d^{exact} is therefore potentially much more accurate than that of Φ_i.

994 A5-24

Note that

$$\varepsilon_{p,eos} \approx -\frac{c_p - c_{pr}}{\theta} \kappa_T \, \nabla T \cdot \nabla \theta \qquad (A17)$$

is related to a nonlinear equation of state. This term vanishes if the pressure variations of c_p can be neglected. As shown by the Maxwell relation (A4), this is the case if

$$\frac{\partial}{\partial \theta} \left(\frac{\alpha}{\rho c_p} \right) = 0. \qquad (A18)$$

Therefore, the condition $\alpha/(\rho c_p) = \text{const.}$ was taken as the appropriate definition of a linear equation of state underlying the construction of the Boussinesq approximation in Tailleux (2013*c*).

REFERENCES

ANDREWS, D.G. 1981 A note on potential energy density in a stratified compressible fluid. *J. Fluid Mech.* **107**, 227–236.

BANNON, P.R. 2013 Available energy of geophysical systems. *J. Atmos. Sci.* **70**, 2650–2654.

BANNON, P.R. & NAJJAR, R.G. 2014 Available energy of the world ocean. *J. Mar. Res.* **72**, 219–242.

FEDERER, M., PAPRITZ, L., SPRENGER, M., GRAMS, C.M. & WENTA, M. 2024 On the local potential energy perspective of baroclinc wave development. *J. Atmos. Sci.* **81**, 871–886.

FLAMENT, P. 2002 A state variable for characterizing water masses and their diffusive stability: spiciness. *Prog. Oceanogr.* **54**, 493–501.

GARGETT, A.E. & HOLLOWAY, G. 1984 Dissipation and diffusion by internal wave breaking. *J. Mar. Res.* **42**, 15–27.

GREGORY, J.M. & TAILLEUX, R. 2011 Kinetic energy analysis of the response of the Atlantic meridional overturning circulation to CO_2-forced climate change. *Clim. Dyn.* **37**, 893–914.

HARRIS, B.L. & TAILLEUX, R. 2018 Assessment of algorithms for computing moist available potential energy. *Q. J. R. Meteorol. Soc.* **144**, 1501–1510.

HARRIS, B.L., TAILLEUX, R., HOLLOWAY, C.E. & VIDALE, P.L. 2022 A moist available potential energy budget for an axisymmetric tropical cyclone. *J. Atmos. Sci.* **79**, 2493–2513.

HOCHET, A. & TAILLEUX, R. 2019 Comments on 'diathermal heat transport in a global ocean model'. *J. Phys. Oceanogr.* **49**, 2189–2193.

HOCHET, A., TAILLEUX, R., FERREIRA, D. & KUHLBRODT, T. 2019 Isoneutral control of effective diapycnal mixing in numerical ocean models with neutral rotated diffusion tensors. *Ocean Sci.* **15**, 21–32.

HOCHET, A., TAILLEUX, R., KUHLBRODT, T. & FERREIRA, D. 2021 Global heat balance and heat uptake in potential temperature coordinates. *Clim. Dyn.* **57**, 2021–2035.

HOLLIDAY, D. & MCINTYRE, M.E. 1981 On potential energy density in an incompressible, stratified fluid. *J. Fluid Mech.* **107**, 221–225.

JACKET, D.R. & MCDOUGALL, T.J. 1985 An oceanographic variable for the characterizion of intrusions and water masses. *Deep Sea Res.* **32**, 1195–2207.

LINDBORG, E. & BRETHOUWER, G. 2008 Vertical dispersion by stratified turbulence. *J. Fluid Mech.* **614**, 303–314.

LORENZ, E.N. 1955 Available potential energy and the maintenance of the general circulation. *Tellus* **7**, 157–167.

MA, Y. & PELTIER, W.R. 2021 Parametrization of irreversible diapycnal diffusivity in salt-fingering turbulence using DNS. *J. Fluid Mech.* **911**, A9.

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

MA, Y. & PELTIER, W.R. 2022*b* Thermohaline–turbulence instability and thermohaline staircase formation in the polar oceans. *Phys. Rev. Fluids* **7**, 083891.

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

MIDDLETON, L., FINE, E.C., MACKINNON, J.A., ALFORD, M.H. & TAYLOR, J.R. 2021 Estimating dissipation rates associated with double diffusion. *Geophys. Res. Lett.* **48**, e2021GL092779.

MIDDLETON, L. & TAYLOR, J.R. 2020 A general criterion for the release of background potential energy through double diffusion. *J. Fluid Mech.* **893**, R3.

MOLEMAKER, M.J. & MCWILLIAMS, J.C. 2010 Local balance and cross-scale flux of available potential energy. *J. Fluid Mech.* **645**, 295–314.

MUNK, W. 1981 Internal waves and small-scale processes. In *Evolution of Physical Oceanography* (ed. C. Wunsch & B.A. Warren), chap. 9, pp. 264–291. MIT Press.

NAKAMURA, N. 1996 Two-dimensional mixing, edge formation, and permeability diagnosed in an area coordinate. *J. Atmos. Sci.* **53**, 1524–1537.

NOVAK, L. & TAILLEUX, R. 2018 On the local view of atmospheric available potential energy. *J. Atmos. Sci.* **75**, 1891–1907.

OAKEY, N.S. 1982 Determination of the rate of dissipation of turbulent energy from simultaneous temperature and velocity shear microstructure measurements. *J. Phys. Oceanogr.* **12**, 256–271.

OSBORN, T.R. 1980 Estimates of the local rate of vertical diffusion from dissipation measurements. *J. Phys. Oceanogr.* **10**, 83–89.

OSBORN, T.R. & COX, C.S. 1972 Oceanic fine structure. *Geophys. Fluid Dyn.* **3**, 321–345.

PARK, Y.-G., WHITEHEAD, J.A. & GNANADESIKAN, A. 1994 Turbulent mixing in stratified fluids: layer formation and energetics. *J. Fluid Mech.* **279**, 279–311.

PENNEY, J., MOREL, Y., HAYNES, P., AUCLAIR, F. & NGUYEN, C. 2020 Diapycnal mixing of passive tracers by Kelvin–Helmholtz instabilities. *J. Fluid Mech.* **900**, A26.

RADKO, T. 2005 What determines the thickness of layers in a thermohaline staircase? *J. Fluid Mech.* **523**, 79–98.

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

ROULLET, G. & KLEIN, P. 2009 Available potential energy diagnosis in a direct numerical simulation of rotating stratified turbulence. *J. Fluid Mech.* **624**, 45–55.

SAENZ, J.A., TAILLEUX, R., BUTLER, E.D., HUGHES, G.O. & OLIVER, K.I.C. 2015 Estimating Lorenz's reference state in an ocean with a nonlinear equation of state for seawater. *J. Phys. Oceanogr.* **45**, 1242–1257.

SCHMITT, R.W. 1994 Double diffusion in oceanography. *Annu. Rev. Fluid Mech.* **26**, 255–285.

SCOTTI, A. & WHITE, B. 2014 Diagnosing mixing in stratified turbulent flows with a locally defined available potential energy. *J. Fluid Mech.* **740**, 114–135.

SHEPHERD, T.G. 1993 A unified theory of available potential energy. *Atmos. Ocean* **31**, 1–26.

SIJP, W.P., GREGORY, J.M., TAILLEUX, R. & SPENCE, P. 2012 The key role of the western boundary in linking the AMOC strength to the north–south pressure gradient. *J. Phys. Oceanogr.* **42**, 628–643.

SMYTH, W.D., NASH, J.D. & MOUM, J.N. 2005 Differential diffusion in breaking Kelvin–Helmholtz billows. *J. Phys. Oceanogr.* **35**, 1004–1022.

TAILLEUX, R. 2009 On the energetics of turbulent stratified mixing, irreversible thermodynamics, Boussinesq models, and the ocean heat engine controversy. *J. Fluid Mech.* **638**, 339–382.

TAILLEUX, R. 2012 Thermodynamics/dynamics coupling in weakly compressible turbulent stratified fluids. *ISRN Thermodyn.* **2012**, 609701.

TAILLEUX, R. 2013*a* Available potential energy and exergy in stratified fluids. *Annu. Rev. Fluid Mech.* **45**, 35–58.

TAILLEUX, R. 2013*b* Available potential energy density for a multicomponent Boussinesq fluid with a nonlinear equation of state. *J. Fluid Mech.* **735**, 499–518.

TAILLEUX, R. 2013*c* Irreversible compressible work and available potential energy dissipation. *Phys. Scr.* **842** (T155), 014033.

TAILLEUX, R. 2016 Generalized patched potential density and thermodynamic neutral density: two new physically based quasi-neutral density variables for ocean water masses analyses and circulation studies. *J. Phys. Oceanogr.* **46**, 3571–3584.

TAILLEUX, R. 2018 Local available energetics of multicomponent compressible stratified fluids. *J. Fluid Mech.* **842**, R1.

TAILLEUX, R. 2021 Spiciness theory revisited, with new views on neutral density, orthogonality and passiveness. *Ocean Sci.* **17**, 203–219.

TAILLEUX, R. 2024 Supporting software for: APE dissipation as the most fundamental criterion for double diffusive instabilities. Zenodo repository.

TAILLEUX, R. & DUBOS, T. 2024 A simple and transparent method for improving the energetics and thermodynamics of seawater approximations: static energy asymptotics (SEA). *Ocean Model.* **188**, 102339.

TAYLOR, J.R. & ZHOU, Q. 2017 A multi-parameter criterion for layer formation in a stratified shear flow using sorted buoyancy coordinates. *J. Fluid Mech.* **823**, R5.

TSENG, Y.-H. & FERZIGER, J.H. 2001 Mixing and available potential energy in stratified flows. *Phys. Fluids* **13**, 1281–1293.

TURNER, J.S. 1985 Multicomponent convection. *Annu. Rev. Fluid Mech.* **17** (1), 11–44.

WINTERS, K.B. & BARKAN, R. 2013 Available potential energy density for Boussinesq fluid flow. *J. Fluid Mech.* **714**, 476–488.

WINTERS, K.B. & D'ASARO, E.A. 1996 Diascalar flux and the rate of fluid mixing. *J. Fluid Mech.* **317**, 179–193.

WINTERS, K.B., LOMBARD, P.N., RILEY, J.J. & D'ASARO, E.A. 1995 Available potential energy and mixing in density stratified fluids. *J. Fluid Mech.* **289**, 115–128.

YOUNG, W.R. 2010 Dynamic enthalpy, conservative temperature, and the seawater Boussinesq approximation. *J. Phys. Oceanogr.* **40**, 394–400.

YOUNG, W.R. 2012 An exact thickness-weighted average formulation of the Boussinesq equations. *J. Phys. Oceanogr.* **42**, 692–707.

ZEMSKOVA, V.E., WHITE, B.L. & SCOTTI, A. 2015 Available potential energy and the general circulation: partitioning wind, buoyancy forcing, and diapycnal mixing. *J. Phys. Oceanogr.* **45**, 1510–1531.