A statistical mechanics approach of mixing in stratified fluids

Antoine Venaille, Louis Gostiaux, Joël Sommeria

A statistical mechanics approach of mixing in stratified fluids

A. VENAILLE[1] †, **L. GOSTIAUX**[2] **AND J. SOMMERIA**[3]

[1] Laboratoire de Physique UMR 5276 CNRS, ENS de Lyon, Université de Lyon, France, [2] LMFA UMR 5509 CNRS, Université de Lyon, France, [3] LEGI, CNRS, Université de Grenoble, France

(Received July 22, 2016)

Predicting how much mixing occurs when a given amount of energy is injected into a Boussinesq fluid is a longstanding problem in stratified turbulence. The huge number of degrees of freedom involved in those processes renders extremely difficult a deterministic approach to the problem. Here we present a statistical mechanics approach yielding prediction for a cumulative, global mixing efficiency as a function of a global Richardson number and the background buoyancy profile. Assuming random evolution through turbulent stirring, the theory predicts that the inviscid, adiabatic dynamics is attracted irreversibly towards an equilibrium state characterised by a smooth, stable buoyancy profile at a coarse-grained level, superimposed with fine-scale fluctuations of velocity and buoyancy. The convergence towards a coarse-grained buoyancy profile different from the initial one corresponds to an irreversible increase of potential energy, and the efficiency of mixing is quantified as the ratio of this potential energy increase to the total energy injected into the system. The remaining part of the energy is literally lost into small scale fluctuations. We show that for sufficiently large Richardson number, there is equipartition between potential and kinetic energy, provided that the background buoyancy profile is strictly monotonic. This yields a mixing efficiency of one quarter, which provides statistical mechanics support for previous predictions based on phenomenological kinematics arguments. In the general case, the cumulative, global mixing efficiency predicted by the equilibrium theory can be computed using an algorithm based on a maximum entropy production principle. It is shown in particular that the variation of mixing efficiency with the Richardson number strongly depends on the background buoyancy profile. We argue that this approach is useful to the understanding of mixing in stratified turbulence in the limit of large Reynolds and Péclet numbers.

1. Introduction

The large-scale stratification and dynamics of the oceans depend crucially on localised turbulent mixing events (Wunsch & Ferrari 2004; Thorpe 2005). These mixing processes occur on temporal and spatial scales much smaller than the current resolutions of general circulation models and must therefore be parameterised (Large *et al.* 1994). It is essential for that purpose to know how much mixing occurs when stratification is stirred by a turbulent flow (Hopfinger 1987; Fernando 1991; Staquet & Sommeria 2002; Peltier & Caulfield 2003; Ivey *et al.* 2008). More precisely, which fraction of the injected energy is lost through a direct turbulent kinetic energy cascade and viscous dissipation, which fraction contributes to modifying the background stratification, and what is the resulting vertical buoyancy profile ? Here we propose to use statistical mechanics as a guideline for the understanding of turbulent stirring and mixing in a stratified fluid.

† antoine.venaille@ens-lyon.fr

Equilibrium statistical mechanics counts the available states of the system with given constraints based on conservation laws. Under random evolution, the system is expected to reach the *macroscopic* state which corresponds to the maximum number of *microscopic* configurations. In this paper, the *macroscopic* quantity to be determined by the theory is the partition between kinetic and potential energy, as well as the corresponding mean (coarse-grained) vertical buoyancy profile. The *microscopic* configurations will be any buoyancy field and non-divergent velocity field, and the constraints will be provided by dynamical invariants of the flow model.

The application of equilibrium statistical mechanics theory to systems described by continuous fields is however problematic; see e.g. Pomeau (1994). Indeed, such systems are characterised by an infinite number of degrees of freedom, which can lead to an accumulation of energy at small scales, whose divergence can be only avoided by an artificial truncation in Fourier space. Kraichnan (1967)

has however explained the energy cascade toward small scales as a trend of the system to approach such equilibrium. By contrast, in two-dimensional turbulence, statistical equilibrium rather accumulates energy at large scale, which Kraichnan has related to the occurrence of an inverse energy cascade. The statistical equilibrium therefore reveals the trend of the evolution for the actual irreversible turbulent system. We here follow a similar idea to study mixing in stratified fluids, using however a quite different statistical mechanics approach.

Instead of considering Galerkin-truncated flows, Onsager (1949) modelled the fluid continuum by a very large but finite set of singular point vortices to explain the self-organisation of two-dimensional turbulent flows as a tendency to reach an equilibrium state, see also Eyink & Sreenivasan (2006). Extensions of those ideas to the continuous two-dimensional Euler and quasi-geostrophic dynamics have been developed independently by Miller (1990) and Robert & Sommeria (1991) (MRS hereafter). This theory is the equivalent of Lynden-Bell's statistical mechanics of Vlasov dynamics (Lynden-Bell 1967), which describes self-organisation in plasma and self-gravitating systems, see e.g. Chavanis (2002). Subsequent work on the theoretical foundation of the approach, as well as on the analytical and numerical computation of equilibrium states is reviewed in Sommeria (2001); Majda & Wang (2006); Bouchet & Venaille (2012). The theory introduces a truncation for the vorticity field, leading to unrealistic vorticity fluctuations at small scale, but it provides remarkable quantitative predictions for the mean velocity field at large scale, as checked by comparisons with direct numerical computations. In the geophysical context, the theory has been used to explain the structure of the Great Red Spot of Jupiter (Turkington *et al.* 2001; Bouchet & Sommeria 2002), oceanic rings and jets (Weichman 2006; Venaille & Bouchet 2011), bottom-trapped oceanic recirculations (Venaille 2012), the stratospheric polar vortex (Prieto & Schubert 2001), the vertical structure of geostrophic turbulence in stratified quasi-geostrophic turbulence (Merryfield 1998; Schecter 2003; Venaille *et al.* 2012) and the structure of the thermocline in global oceanic circulation (Salmon 2012).

Perhaps more surprisingly, this approach has also been shown to be relevant in flow systems that permit the existence of a direct energy cascade, such as three-dimensional axisymmetric Euler flows (Naso *et al.* 2010; Thalabard *et al.* 2014, 2015). Bifurcations involving symmetry breaking between different large-scale flow structures could be explained and quantitatively described by such statistical mechanics theory (Naso *et al.* 2010; Thalabard *et al.* 2015), and energy partition between toroidal and poloidal modes is predicted by the theory (Thalabard *et al.* 2014). Similarly, equilibrium theory has been used to predict the energy partition between inertia-gravity waves and vortical modes in shallow water models (Warn 1986; Weichman & Petrich 2001; Renaud *et al.* 2016). The

success of these previous works provides a strong incentive to apply a similar approach to a non-rotating, density-stratified Boussinesq fluid in order to predict the partition between kinetic and potential energy for a given amount of energy injected into the system. We build for that purpose upon previous work by Tabak & Tal (2004), who computed the most probable buoyancy field of a two-layer fluid with a prescribed total energy, assuming that the kinetic energy is constant at each height. Our contributions are twofold. First, we generalise their result to arbitrary buoyancy profiles, and obtain the kinetic energy profile as the output of the statistical theory. Second, we use these results to obtain predictions for mixing efficiency in decaying configurations.

How to infer the efficiency of mixing in forced-dissipative or decaying experiments has been carefully addressed in previous studies; see e.g. Winters *et al.* (1995); Peltier & Caulfield (2003); Wykes *et al.* (2015) and references therein. The traditional approach involves direct analyses of the diffusive destruction of small scale density variance as the experiment proceeds, which in turn requires a separation of the influence of stirring from that of irreversible mixing through application of the Lorenz concept of available potential energy that can be converted into kinetic energy and a base-state potential energy which can not. It has been demonstrated that the diffusive destruction of small scale density variance may be represented by the time derivative $\mathcal{M}$ of base-state potential energy plus a small correction due to the action of molecular diffusion on the initial density stratification, a correction that becomes negligible in the limit of high Reynolds number (Winters *et al.* 1995). The time dependent efficiency of turbulent mixing may be then computed from the direct numerical simulations as $\eta_t = \mathcal{M}/(\mathcal{M}+\epsilon)$ where ϵ is the rate of viscous kinetic energy dissipation in the fluid domain (Peltier & Caulfield 2003). This definition of mixing efficiency is global in space since the computation of the base-state potential energy requires a rearrangement of the fluid particle at the domain scale. Using a number of additional assumptions, it may be related to a local mixing efficiency that is often used in oceanography to model an effective diffusivity for diapycnal mixing Osborn (1980); Hopfinger (1987); Tailleux (2009). In decaying experiments, it is also convenient to define a cumulative mixing efficiency $\eta_{tot} = \int_0^{+\infty} \mathrm{d}t\mathcal{M} / \int_0^{+\infty} \mathrm{d}t\,(\mathcal{M}+\epsilon)$, which measures how much of the total injected energy has been used to irreversibly raise the potential energy of the system in the experiment. In practice, this quantity can easily be inferred in laboratory experiments by measuring the buoyancy profile once all dissipative effects have died-out, assuming the initial background stratification and the initial injected energy are known. We argue in this paper that the equilibrium statistical mechanics is a tool to estimate the cumulative mixing efficiency in decaying stratified turbulence.

Applying the statistical mechanics programs to Boussinesq dynamics is done in three steps. The first step is to find relevant phase-space variables. These variables must satisfy a Liouville theorem, and we show in this paper (Appendix A) that this is the case of the velocity and buoyancy fields. This ensures that the dynamics is non-divergent in phase-space, so that the probability densities expressed in these variables remain constant during the time evolution of the system. The fundamental postulate of equal probability for each microscopic configuration is then consistent with the dynamical evolution. Second, we need to introduce a discretisation of the continuous fields describing the system. This technical step is classical when computing equilibrium states of systems described by deterministic partial differential equations. Once the discrete approximation of the fields is introduced, one can count the microscopic configurations, and the computation of the equilibrium states is rigorous. The third step is to introduce a macroscopic description of the system, and to find the most probable macrostates among all those that satisfy a set

of constraints provided by dynamical invariants. Using the equilibrium theory to describe the long time behaviour of the system requires finally the assumption of ergodicity, i.e. that the system evenly explores phase space. Even if the ergodic assumption may not be fulfilled in actual turbulent flow, computing the equilibria is at least a useful and necessary first step before more comprehensive studies out-of-equilibrium.

Denoting H the height of the flow domain, Δb the typical variations of the background buoyancy profile, $(U,\ L_t)$ the typical velocity and length scale of turbulence, and $(\nu,\ \kappa)$ the molecular viscosity and diffusivity, the efficiency of mixing depends *a priori* on four non-dimensional parameters in laboratory or numerical experiments on stratified turbulence: a global Richardson number based on the domain scale $Ri = H\Delta b/U^2$, the Reynolds number $Re = UL_t/\nu$, the Péclet number $Pe = UL_t/\kappa$, and the ratio L_t/H which depends on the energy injection mechanism.

The equilibrium statistical mechanics theory applies to the freely evolving inviscid adiabatic dynamics. Considering such an approach to describe actual stratified turbulence amounts to assuming that the Reynolds number Re and the Péclet number Pe are sufficiently large, and that the typical time scale to approach the equilibrium state is smaller than the typical time scale for energy dissipation. Independently from statistical mechanics arguments, neglecting molecular effects is a natural assumption in the large Reynolds number limit, which has been proven useful in previous studies on three-dimensional turbulence (Eyink & Sreenivasan 2006), in which case the observed dissipation rate of energy ϵ becomes independent from viscosity; see e.g. Vassilicos (2015) and references therein. In the presence of vanishingly small viscosity, these fluctuations would eventually be smoothed out, and we assume that the amount of energy which is dissipated by viscosity is governed by the inertial dynamics. Similarly, we assume that the amount of buoyancy fluctuations which are smoothed out by molecular diffusion is controlled by the inertial dynamics. Within the framework of the equilibrium theory, we assume conservation of the total energy, but we show that part of this energy is irreversibly transferred to small scale fluctuations once the equilibrium state is reached. Since the amount of kinetic energy and buoyancy fluctuations that are irreversibly transferred to small scales can be computed explicitly within the equilibrium statistical mechanics framework, the theory makes possible a prediction for the cumulative mixing efficiency, even in the absence of viscosity or molecular dissipation in the model..

Because of the ergodic hypothesis, the mixing efficiency predicted by the equilibrium theory depends only on the global distribution of buoyancy levels (inferred from the background buoyancy profile) and on the injected energy, expressed in non-dimensional form by the global Richardson number Ri. It does not depends on details of turbulence generation, like the ratio L_t/H, and as discussed above, it does not depend on Re and Pe. We here provide an algorithm which predicts how the cumulative mixing efficiency depends on the Richardson number for an arbitrary background buoyancy profile.

The paper is organised as follows. The equilibrium statistical mechanics theory is introduced and discussed in the second section. The actual computation of the equilibrium states is discussed in a third section. Application of the theory to predict mixing efficiency in freely-evolving flow (decaying turbulence) is discussed in a fourth section. We conclude and summarise the main results in the fifth section. Technical results on the Liouville theorem, on the computation of the macrostate entropy and on the numerical algorithm used to compute the equilibria are presented in two appendices.

2. Equilibrium statistical mechanics of non-rotating, density-stratified Boussinesq fluids

2.1. *Dynamical system and invariants*

We consider an inviscid Boussinesq fluid that takes place in a three-dimensional domain $\mathcal{V}_\mathbf{x}$ of volume V, see e.g. Vallis (2006). Spatial coordinates are denoted $\mathbf{x} = (x, y, z)$, with $\mathbf{e}_z$ the vertical unit vector pointing in the upward direction. At each time t the system is described by the buoyancy field $b = g\left(\varrho_0 - \varrho\right)/\varrho_0$, where $\varrho(x, y, z, t)$ is the fluid density, g gravity and ϱ_0 a reference density, and by the velocity field $\mathbf{u} = (u, v, w)$, which is non-divergent:

$$\nabla \cdot \mathbf{u} = 0 \ . \tag{2.1}$$

In the absence of diffusivity, the buoyancy field is purely advected by the velocity field

$$\partial_t b + \mathbf{u} \cdot \nabla b = 0 \ , \tag{2.2}$$

and the dynamics of the velocity field is coupled to the buoyancy field through the momentum equation

$$\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u} = -\frac{1}{\varrho_0} \nabla P + b\mathbf{e}_z \ . \tag{2.3}$$

Equation (2.2) describes the Lagrangian conservation the buoyancy. It implies the conservation of the global distribution (i.e. histogram) of buoyancy levels

$$G(\sigma) = \frac{1}{V} \int_{\mathcal{V}_\mathbf{x}} \mathrm{d}\mathbf{x} \ , \delta(b - \sigma) \tag{2.4}$$

expressed as $\mathrm{dG/dt} = 0$. The conservation of $G(\sigma)$ is equivalent to the conservation of all the Casimir functionals $F[b] = \int \mathrm{d}\mathbf{x} \ f(b)$, with f any arbitrary function; see e.g. Potters *et al.* (2013). This conservation law is also equivalent to the conservation of the background (or sorted) buoyancy profile $b_s(s)$ defined as the buoyancy profile with minimal potential energy using

$$G(b_s)\mathrm{d}b_s = \frac{1}{2H}\mathrm{d}z. \tag{2.5}$$

Similarly, using Eqs. (2.1), (2.2) and (2.3) one can show that the total energy of the flow

$$E = \int_{\mathcal{V}_\mathbf{x}} \mathrm{d}\mathbf{x} \ \left(\frac{1}{2}\mathbf{u}^2 - bz\right) + \int_{\mathcal{V}_\mathbf{x}} \mathrm{d}\mathbf{x} \ zb_s \tag{2.6}$$

is another dynamical invariant : $\mathrm{dE/dt} = 0$. Note that the total energy is defined up to a constant, but we have chosen this constant such that the energy vanishes when there is no motion and when the buoyancy field is sorted ($E = 0$ when $\mathbf{u} = 0$ and $b = b_s$).

The Boussinesq equations are characterised by additional dynamical invariants related to the conservation of Ertel potential vorticity, see e.g. Salmon (1998). These invariants are essential to explain the occurrence of inverse cascade and self-organisation of the velocity field occurring in the presence of sufficiently large rotation. However various theoretical and numerical studies indicate that stratified turbulence in the absence of rotation is not influenced significantly by these invariants (Bartello 1995; Lindborg 2005, 2006; Waite & Bartello 2004; Herbert *et al.* 2014). We will therefore not consider the constraints related to the conservation of Ertel potential vorticity in the remaining of this paper. In the context of equilibrium statistical mechanics, this amounts to assume that the entropy maxima obtained with and without those constraints are the same.

2.2. *Microscopic configurations, macroscopic description and variational problem*

For an isolated system, the fundamental postulate of equilibrium statistical mechanics is the equiprobability of the microscopic configurations corresponding to the same values of the dynamical invariants.

The first step is to define what are the relevant phase-space variables describing these *microscopic configurations*. Those variables must satisfy a Liouville's theorem, which means that the flow in phase space is non-divergent. This ensures that microscopic configurations remain equiprobable during the time evolution of the system. We show in Appendix A that the quadruplet of fields $(b, \mathbf{u})$ satisfy such a Liouvillle theorem, and are therefore relevant phase-space variables.

The second step is to identify the relevant dynamical invariants, which are here the total energy and the global distribution of buoyancy levels, defined in Eq. (2.6) and in Eq. (2.4), respectively. The ensemble of microscopic configurations characterised by the same dynamical invariants is called the microcanonical ensemble. This is the relevant ensemble to consider for an isolated system such as the unforced, inviscid, adiabatic Boussinesq system.

The third step is to identify relevant macrostates, which describe an ensemble of microscopic configurations. We introduce for that purpose the probability $\rho(\mathbf{x}, \sigma, \mathbf{v})$ of finding the buoyancy level σ and the velocity level $\mathbf{v}$ in the vicinity of point $\mathbf{x}$. It is normalised at each point:

$$\forall \mathbf{x} \in \mathcal{V}_{\mathbf{x}},\ \mathcal{N}_{\mathbf{x}}[\rho] = \int_{\mathcal{V}_{\mathbf{v}}} \mathrm{d}\mathbf{v} \int_{\mathcal{V}_\sigma} \mathrm{d}\sigma\ \rho(\mathbf{x}, \sigma, \mathbf{v}) = 1\ , \tag{2.7}$$

where the integral bounds are

$$\mathcal{V}_{\mathbf{v}} = [-\infty,\ +\infty]^3\,, \quad \mathcal{V}_\sigma = [-\infty,\ +\infty]\,. \tag{2.8}$$

Each microscopic state $(b(\mathbf{x}), \mathbf{u}(\mathbf{x}))$ is described at a macroscopic level by the PDF $\rho(\mathbf{x}, \sigma, \mathbf{v})$, and many microscopic configurations are in general associated with a given PDF $\rho(\mathbf{x}, \sigma, \mathbf{v})$. This quantity is relevant to describe fields characterised by wild local fluctuations, including fields that are not differentiable, and is called a Young measure in mathematics; see e.g. Robert & Sommeria (1991).

Let us define more precisely how to compute the macroscopic state $\rho(\mathbf{x}, \sigma, \mathbf{v})$ from a given microscopic configuration $(b(\mathbf{x}), \mathbf{u}(\mathbf{x}))$, which will be useful to count the number of microscopic configurations associated with a given macrostate. For that purpose, we follow a procedure which is standard in the framework of equilibrium statistical mechanics of fluid systems, using a discrete approximation of the continuous fields. We consider a uniform *coarse-grained* grid containing N macrocells, and a *fine-grained* grid obtained by dividing each macrocell of the coarse-grained grid into a uniform grid containing M fluid particles, see Fig. 1. On the one hand, discretisation of the *microscopic* field $b(\mathbf{x})$ and $\mathbf{u}(\mathbf{x})$ are defined on the *fine-grained* grid, which contains MN fluid particles. This procedures also requires a discretisation of the buoyancy and velocity levels carried by the fluid particles, which is further discussed in Appendix B. On the other hand, the discrete approximation of the PDF ρ is defined on the *coarse-grained* grid.

For a given microscopic configuration, one can compute within each macrocell of the *coarse-grained* grid the frequency of occurrence of buoyancy and velocity levels (a normalised histogram). In the limit $M \to +\infty$, for a prescribed value of N, the discrete approximations of the microscopic configurations tend to the continuous ones, and the discrete approximation of the PDF ρ is equivalent to the frequency of occurrence of buoyancy and velocity levels within each macrocell. In other words, the discrete approximation of the PDF ρ can be interpreted as the volume proportion of fluid particles carrying the

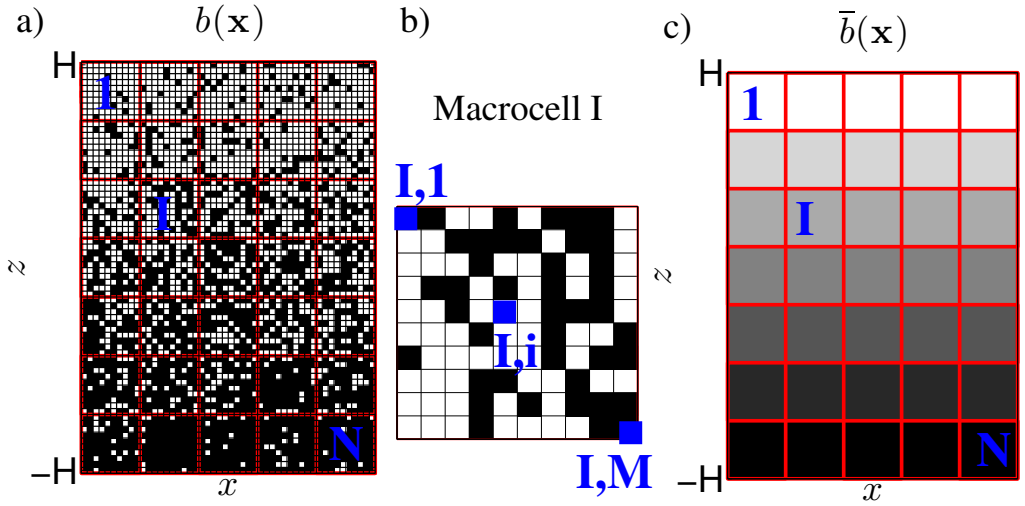

FIGURE 1. a) A microscopic configuration of the discretised buoyancy field $b(\mathbf{x})$. The discretised buoyancy field is defined on a uniform *fine-grained* grid containing $M \times N$ elements, where N is the number of grid points of the uniform *coarse-grained* grid (red color). b) Zoom on a single macrocell, containing M microcells. Each microcell contains one fluid particle. Here we consider the case of a two-level system: the buoyancy carried by each fluid particle is $b = \pm\Delta b/2$. c) The macroscopic buoyancy field $\overline{b}(x)$ is defined on the uniform *coarse-grained* grid (red colour), and is computed in the limit $M \to +\infty$ by averaging the microscopic buoyancy field within each macrocell, see e.g. Miller (1990); Tabak & Tal (2004).

buoyancy level σ and velocity level $\mathbf{v}$ inside each macrocell. The continuous PDF field ρ is then recovered by considering the limit $N \to +\infty$, which corresponds to the limit of infinitesimal macrocells. Several useful macroscopic quantities can be deduced from ρ, such as the macroscopic buoyancy field

$$\overline{b}(\mathbf{x}) = \int_{\mathcal{V}_\sigma} \mathrm{d}\sigma \int_{\mathcal{V}_\mathbf{v}} \mathrm{d}\mathbf{v}\ \rho\sigma\ , \tag{2.9}$$

and the local eddy kinetic energy field

$$\frac{1}{2}\overline{\mathbf{u}^2}(\mathbf{x}) = \int_{\mathcal{V}_\sigma} \mathrm{d}\sigma \int_{\mathcal{V}_\mathbf{v}} \mathrm{d}\mathbf{v}\ \frac{1}{2}\rho\mathbf{v}^2\ . \tag{2.10}$$

Within the framework of the discrete approximation depicted in Fig. 1, those macroscopic quantities correspond to averages over macrocells, i.e. to a spatial coarse-graining at the scale of a macrocell $\sim N^{-1/3}$. Importantly, the small scale fluctuations described by the macroscopic states are confined at spatial scales below this coarse-graining scale, which tends to zero in the limit $N \to +\infty$.

The advantage of considering the probability field ρ rather than only the coarse-grained fields such as $\overline{b}$ for a macroscopic description of the system is that global constraints provided by dynamical invariants can be expressed in terms of ρ. The global constraints are given by the energy and the global distribution of buoyancy levels, which are defined as functional of phase-space variables $(\mathbf{u}, b)$ in Eqs. (2.6) and (2.4), respectively. Considering the discrete approximation described in the previous paragraph, decomposing the spatial integrals appearing in Eqs. (2.6) and (2.4) as a sum of spatial integrals over each macrocells, remembering then that the PDF ρ is the frequency of occurrence of buoyancy and velocity levels within a given macrocell, and taking finally the limit $M \to +\infty$, $N \to +\infty$, the energy and the global distribution of buoyancy levels can be expressed as

functionals of the PDF ρ:

$$\mathcal{E}[\rho] = \int_{\mathcal{V}_\mathbf{x}} \mathrm{d}\mathbf{x} \int_{\mathcal{V}_\mathbf{v}} \mathrm{d}\mathbf{v} \int_{\mathcal{V}_\sigma} \mathrm{d}\sigma \ \rho \left(\frac{\mathbf{v}^2}{2} - \sigma z \right) + \int_{\mathcal{V}_\mathbf{x}} \mathrm{d}\mathbf{x} \ z b_s \ , \tag{2.11}$$

$$\mathcal{G}_\sigma[\rho] = \int_{\mathcal{V}_\mathbf{x}} \mathrm{d}\mathbf{x} \int_{\mathcal{V}_\mathbf{v}} \mathrm{d}\mathbf{v} \ \rho \ . \tag{2.12}$$

The microcanonical ensemble is defined by the ensemble of microstates characterised by the same energy E and global distribution of buoyancy levels $G(\sigma)$. This ensemble contains therefore all the macroscopic states that satisfy the dynamical constraints $\mathcal{E}[\rho] = E$ and $\mathcal{G}_\sigma[\rho] = G(\sigma)$.

The last step is to count how many microscopic configurations are associated with a given macrostate. Considering our discrete approximation of the fields, it is shown in Appendix B that within the microcanonical ensemble, an overwhelming number of the microscopic configurations is concentrated close to the most probable macrostate, which maximises the macrostate entropy

$$\mathcal{S} = - \int_{\mathcal{V}_\mathbf{x}} \mathrm{d}\mathbf{x} \int_{\mathcal{V}_\mathbf{v}} \mathrm{d}\mathbf{v} \int_{\mathcal{V}_\sigma} \mathrm{d}\sigma \ \rho \log \rho \ . \tag{2.13}$$

The expression of the macrostate entropy given in Eq. (2.13) is a classical one, especially in the context of two-dimensional turbulence (Miller 1990; Robert & Sommeria 1991). A rigorous derivation of such macrostate entropy requires the use of large deviation theory; see e.g. Touchette (2009) for an introduction to those tools. A key difficulty in deriving rigorously this macrostate entropy from the usual Boltzmann entropy is that the microstates are continuous fields which contain an infinite number of degrees of freedom, and which are constrained by an infinite number of dynamical invariants. Several discretisation procedures have been proposed to bypass this difficulty, see e.g. Michel & Robert (1994); Boucher *et al.* (2000); Bouchet & Corvellec (2010); Potters *et al.* (2013); Renaud *et al.* (2016). A similar formula has been derived previously by Tabak & Tal (2004) in the context of non-rotating, density stratified Boussinesq fluids, in the particular case of a two-level buoyancy configuration. Here we have generalised this result to arbitrary buoyancy distribution, and more importantly, we have included the velocity field in the description of the microstate, which is essential to account for energy conservation.

2.3. *Computation of the most probable mascrostate, and general properties of the equilibrium states*

The first step to find the equilibrium state is to compute critical points of the variational problem given by the equilibrium theory, i.e. to find the field ρ such that first variations of the macrostate entropy (2.13) around this state vanish, given the constraints of the problem given by $\mathcal{E}[\rho] = E$, $\mathcal{G}_\sigma[\rho] = G(\sigma)$, $\mathcal{N}_\mathbf{x}[\rho] = 1$, where $\mathcal{E}$ is the energy defined in Eq. (2.11), $\mathcal{G}_\sigma$ is the global distribution of buoyancy defined in Eq. (2.12), and $\mathcal{N}_\mathbf{x}$ the local normalization of the PDF expressed in Eq. (2.7). One needs for that purpose to introduce the Lagrange multipliers β_t, $\gamma(\sigma)$, $\xi(\mathbf{x})$ associated with those constraints. Computing first variations with respect to the probability field ρ yields

$$\delta\mathcal{S} - \beta_\mathrm{t}\delta\mathcal{E} + \int_{\mathcal{V}_\sigma} \mathrm{d}\sigma \ \gamma(\sigma) \ \delta\mathcal{G}_\sigma + \int_{\mathcal{V}_\mathbf{x}} \mathrm{d}\mathbf{x} \ \xi(\mathbf{x}) \delta\mathcal{N}_\mathbf{x} = 0 \ . \tag{2.14}$$

Using the expression of the entropy, of the energy, of the global distribution of buoyancy and of the normalisation constraints given respectively in Eqs. (2.13), (2.11), (2.12) and

(2.7), Eq. (2.14) yields

$$-\int_{\mathcal{V}_\mathbf{x}} \mathrm{d}\mathbf{x} \int_{\mathcal{V}_\mathbf{v}} \mathrm{d}\mathbf{v} \int_{\mathcal{V}_\sigma} \mathrm{d}\sigma \ \left((1+\log\rho) + \beta_t \left(\frac{\mathbf{v}^2}{2} - \sigma z \right) - \gamma\left(\sigma\right) - \xi\left(\mathbf{x}\right) \right) \delta\rho = 0. \quad (2.15)$$

This equality is true for any $\delta\rho$, which, using the normalisation constraint in Eq. (2.7), yields the following necessary and sufficient condition for ρ to be a critical point of the variational problem:

$$\rho\left(\mathbf{x},\sigma,\mathbf{v}\right) = \left(\frac{\beta_\mathrm{t}}{2\pi}\right)^{3/2} e^{-\beta_\mathrm{t}\frac{\mathbf{v}^2}{2}} \rho_b(z,\sigma), \quad (2.16)$$

with

$$\rho_b(z,\sigma) \equiv \frac{e^{\beta_\mathrm{t}\sigma z + \gamma(\sigma)}}{\mathcal{Z}(z)}, \quad \mathcal{Z}(z) \equiv \int_{\mathcal{V}_\sigma} \mathrm{d}\sigma \ e^{\beta_\mathrm{t}\sigma z + \gamma(\sigma)} \ . \quad (2.17)$$

The values of the Lagrange multipliers β_t and $\gamma(\sigma)$ are implicitly determined by the expression of the constraints $\mathcal{E}[\rho] = E$ and $\mathcal{G}_\sigma[\rho] = G(\sigma)$, given by Eq. (2.11) and Eq. (2.12), respectively.

The probability density field (2.16) is expressed as a product of the probabilities for buoyancy and velocity, which means that b and $\mathbf{u}$ are two independent quantities at equilibrium. The predicted velocity distribution is Gaussian, with zero mean ($\overline{\mathbf{u}} = 0$), isotropic and homogeneous in space. It is therefore fully characterised by the local eddy kinetic energy

$$e_c \equiv \frac{1}{2}\overline{\mathbf{u}^2} = \frac{3}{2}\frac{1}{\beta_\mathrm{t}} \quad (2.18)$$

The inverse of β_t defines an effective "temperature" of the turbulent field, corresponding to the turbulent agitation of fluid particles. Remarkably, the three-dimensional nature of the flow appears only in this equation, and nowhere else. A two-dimensional case would just have a different relation between kinetic energy and this effective temperature.

The predicted buoyancy distribution ρ_b depends only on the height coordinate z. The equilibrium theory predicts therefore that the local fluctuations of buoyancy are invariant in the horizontal. It means that in the remaining of this paper, the quantities $\bar{\cdot}$ can be interpreted either as a local coarse-graining or as an horizontal average. Similarly, the quantity ρ_b can be interpreted either as a local distribution of buoyancy or as the distribution of buoyancy over an horizontal plane.

Eq. (2.17) relates the mean buoyancy profile and its fluctuations to the effective turbulent temperature. Buoyancy moments are defined at each height in terms of $\rho_b(z,t)$ as

$$\overline{b^n}(\mathbf{x}) \equiv \int_{\mathcal{V}_\sigma} \mathrm{d}\sigma \ \sigma^n \rho_b \ . \quad (2.19)$$

From Eq. (2.17) we get the relations

$$\bar{b} = \frac{1}{\beta_\mathrm{t}} \frac{\mathrm{d}\log\mathcal{Z}}{\mathrm{d}z} \ , \quad \overline{b^2} - \bar{b}^2 = \frac{1}{\beta_\mathrm{t}^2} \frac{\mathrm{d}^2\log\mathcal{Z}}{\mathrm{d}z^2} \ . \quad (2.20)$$

Using those expressions and Eq. (2.18), one gets finally an expression relating the mean buoyancy profile to the ratio of the buoyancy fluctuations to the kinetic energy fluctuations:

$$\frac{\mathrm{d}\bar{b}}{\mathrm{d}z} = 3\frac{\overline{b^2} - \bar{b}^2}{2e_c} \ . \quad (2.21)$$

In the case of a strong stratification, the local variance of buoyancy is proportional to

the small vertical displacement of fluid elements, so this relation can be interpreted as an equipartition between kinetic and potential energy fluctuations, as further discussed in section 4.3.

We stress finally that the equilibrium state has a peculiar spatial structure: the buoyancy field b is characterised by a smooth coarse-grained buoyancy profile $\overline{b}(z)$ superimposed with wild small scale buoyancy fluctuations. More precisely, the theory predicts that when performing a local coarse-graining of the microscopic buoyancy and velocity fields at a scale l (the scale of the macrocell within the framework of our discrete model depicted in fig. 1), the small scale fluctuations are confined at scales smaller than the coarse-graining scale l, no matter how small the coarse-graining length scale l.

3. Computation of mean equilibrium buoyancy profiles

3.1. *The two-level case*

We discussed in the previous subsection the general case with a continuum of buoyancy levels. In the particular case with a finite number of buoyancy levels (say K levels σ_k with $1 \leq k \leq K$), the buoyancy field is described at a macroscopic level by $p_k(\mathbf{x})$, which is the probability of measuring the level σ_k at point $\mathbf{x}$ with $\sum_{k=1}^{K} p_k(\mathbf{x}) = 1$, see Appendix B. The same arguments as in subsection 2.3 for the computation of the equilibrium state then yields

$$p_k(z) \equiv \frac{e^{\beta_t \sigma_k z + \gamma_k}}{\sum_{k=1}^{K} e^{\beta_t \sigma_k z + \gamma_k}} \ , \quad \beta_t = \frac{3}{2e_c}, \tag{3.1}$$

where the values of the Lagrange multipliers β_t and $\{\gamma_k\}_{1\leq k\leq K}$ are implicitly determined by the energy constraint and conservation of the total volume occupied by each buoyancy level σ_k.

Let us restrict ourselves to the case of an initial state composed of two buoyancy levels in equal proportion with

$$\forall \mathbf{x} \in \mathcal{V}_\mathbf{x},\ b(\mathbf{x}) \ \in \ \left\{ -\frac{\Delta b}{2},\ \frac{\Delta b}{2} \right\}. \tag{3.2}$$

The only dimensionless parameter of the problem within the statistical mechanics framework is given by the global Richardson number based on the total height $2H$, buoyancy jump Δb and square of velocity fluctuations $2e_c$:

$$Ri \equiv \frac{H \Delta b}{e_c} \ . \tag{3.3}$$

This global Richardson number based on the domain height H is different from the bulk Richardson number $Ri_b = \Delta b L_t / e_c = (L_t/H) Ri$ based on the turbulent length scale L_t, which is commonly used in the context of turbulent mixing in stratified fluids; see e.g. (Fernando 1991). The statistical mechanics prediction depends only on the total energy, not on its injection scale L_t. This point will be further discussed in section 4.4.

We denote $p_+(z)$ the probability of measuring $\Delta b/2$ at height z. According to the notation used in Eq. (3.1), we get $\sigma_1 = -\Delta b/2$, $\sigma_2 = \Delta b/2$, $p_1 = 1 - p_+$, $p_2 = p_+$, with

$$p_+(z) = \frac{e^{\frac{3Ri}{4}\frac{z}{H}}}{e^{-\frac{3Ri}{4}\frac{z}{H}} + e^{\frac{3Ri}{4}\frac{z}{H}}} \ , \tag{3.4}$$

where we have used the symmetry with respect to $z = 0$ ($p + (z) = -p(-z)$) and the fact that the two buoyancy levels are in equal proportions ($\int_{-H}^{+H} \mathrm{d}z p_+ = \int_{-H}^{+H} \mathrm{d}z p_-$) to eliminate the Lagrange parameters γ_1, γ_2 in Eq. (3.1).

Equation (3.4) is reminiscent of the Fermi-Dirac distribution. Indeed, the conservation of buoyancy plays here the same role as the exclusion principle for the statistics of fermions: within the framework of the discretised model depicted in Fig. 1, the buoyancy carried by a fluid particle at a given grid point can only take one value among $-\Delta b/2$ and $\Delta b/2$. Following this analogy, the buoyancy field is a collection of fluid particles carrying the potential energy $e_p = \pm^1\!/_2 z\Delta b$, with a Fermi level $\varepsilon_f = 0$, in thermal contact with a heat bath characterised by the inverse temperature β_t.

Using Eq. (3.4) and Eq. (2.18), the mean density profile $\overline{b} = \frac{\Delta b}{2} p_+ - \frac{\Delta b}{2}(1-p_+)$ is expressed as

$$\overline{b}(z) = \frac{\Delta b}{2} \tanh\left(\frac{3Ri}{4}\frac{z}{H}\right) . \tag{3.5}$$

Large global Richardson numbers $Ri \gg 1$ correspond to sharp interfaces: the kinetic energy is too small to allow for large excursion of fluid particles away from the rest position. By contrast, small global Richardson numbers $Ri \ll 1$ correspond to a homogenised buoyancy field: the total kinetic energy is much larger than the energy required to mix the buoyancy field. This tanh profile was previously obtained by Tabak & Tal (2004) using similar arguments, but without relating the effective temperature to the kinetic energy of the flow in a consistent theory. Our approach allows us for a direct interpretation of the effective temperature of the flow as the local turbulent kinetic energy, which will make possible quantitive estimate for mixing efficiency.

3.2. *A relaxation equation towards the equilibrium states*

The expression of the equilibrium state given in Eq. (2.17) requires the knowledge of the Lagrange multipliers $\gamma(\sigma)$ and e_c, which depend implicitly on the constraints $G(\sigma)$ and E. This makes analytical computations of those equilibria very challenging. Solutions may be obtained in particular cases, such as for the two-level configuration analysed in subsection 3.1, but more generally it must be determined numerically.

We devise for that purpose an algorithm based on a *maximum entropy production principle*, which was introduced by Robert & Sommeria (1992) in order to compute equilibrium states of two-dimensional Euler flows. The idea of the algorithm is to consider a time dependent probability distribution function

$$\rho\left(\sigma, \mathbf{x}, \mathbf{v}, t\right) = \left(\frac{3}{4\pi e_c(t)}\right)^{3/2} e^{-\frac{3}{2e_c(t)}\frac{\mathbf{v}^2}{2}} \rho_b(z, \sigma, t), \tag{3.6}$$

where the pdf $\rho_b(z,\sigma,t)$ and the local kinetic energy $e_c(t)$ depend on time, and can be different from the pdf and the kinetic energy of the actual equilibrium state. We derive in Appendix C a dynamical equation for ρ_b that conserves the total energy and the global distribution of buoyancy levels, while maximising the entropy production at each time:

$$\partial_t \rho_b = \partial_z \left[D\left(\partial_z \rho_b - \frac{3}{2e_c}\left(\sigma - \overline{b}\right)\rho_b\right)\right] , \tag{3.7}$$

where D is an arbitrary positive diffusion coefficient. The kinetic energy e_c defined in Eq. (2.18) is expressed in term of the total energy E and the buoyancy profile $\overline{b}(z,t)$ by using Eq. (2.11):

$$e_c = \frac{E}{V} + \frac{1}{2H}\int_{-H}^{+H} \mathrm{d}z \ \left(\overline{b} - b_s\right) z, \tag{3.8}$$

with V the volume of the flow domain.

Maximising the entropy production ensures that the system relaxes towards an equilibrium state. Indeed, using Eq. (2.17)-(2.18) and the first equality in Eq. (2.20), the

equilibrium states can be written as

$$\rho_b(\sigma,z)=\rho_b(\sigma,0)e^{\frac{3}{2e_c}\left(\sigma z-\int_0^z \mathrm{d}z'\ \overline{b}(z')\right)}\ , \tag{3.9}$$

which is also the expression of any stationary solution of Eq. (3.7). According to equation (3.7) the equilibrium state can be interpreted as the result of a compensation between usual turbulent diffusion and a drift term corresponding to restratification of buoyancy fluctuations. We stress that the convergence towards equilibrium depends on the parameter D, but that the equilibrium itself does not depend on this parameter. This is why is can be chosen arbitrarily.

Assuming that the initial energy E injected into the system and that the background buoyancy profile $b_s(z)$ are known, one can then use the relaxation algorithm (3.7), starting from the state

$$\rho_b(z,\sigma,0)=\delta(b_s(z)-\sigma),\quad e_c(0)=\frac{E}{V}. \tag{3.10}$$

Equation (3.7) is an integro-differential equation, because the local kinetic energy is a functional of the macroscopic vertical buoyancy profile. Its numerical implementation is much easier assuming that e_c is a constant. One then loses energy conservation, but the equation still conserves the global buoyancy distribution, assuming no buoyancy fluxes at the upper and lower boundaries. It can be shown that this process minimises the free-energy production defined as $\dot{\mathcal{F}}=-\dot{\mathcal{S}}+\beta_t\dot{\mathcal{E}}$, where the upper dot stands for a time derivative, and where $\beta_t=3/(2e_c)$ can be interpreted as the inverse of an effective turbulent temperature. Indeed, assuming constant local kinetic energy amounts to a computation of the equilibrium state within the canonical ensemble where the "heat bath" is provided by turbulent agitation. In order to solve numerically Eq. (3.10) with constant e_c, we first assume a discretisation of the global buoyancy distribution into N_σ buoyancy levels denoted σ_n with $1\leq n\leq\ N_\sigma$. Denoting $\rho_{b,n}(z,t)$ the probability to measure the level σ_n in the vicinity of height z at time t, we obtain a system of one dimensional parabolic partial differential equations for $\{\rho_{b,n}(z,t)\}_{1\leq n\leq\ N_\sigma}$, which can be solved using standard numerical procedures. This dynamical system is integrated in time until a steady state is reached. This steady state if the equilibrium state. Once the equilibrium state associated with a given value of e_c is computed, it is straightforward to compute its total energy E using Eq. (2.11). One can then check that varying e_c from 0 to $+\infty$ amounts to varying E from 0 to $+\infty$. This procedure therefore provides the complete set of equilibria associated with any given background buoyancy profile.

We show in Fig. 2 two examples of equilibrium states computed by this procedure, assuming no buoyancy fluxes at the upper and lower boundaries. Panels a,b corresponds to the two-level configuration. As expected from Eq. (3.5), the mean equilibrium buoyancy profile is characterised by a tanh shape in that case. Panel b confirms that this equilibrium state may be interpreted as the result of a balance between a classical downgradient term $-D\partial_z\overline{b}$ modelling turbulent transport and a term $D(3/2e_c)\left(\overline{b^2}-\overline{b}^2\right)$ modelling restratification.

Panels c,d correspond to the more complicated case of a linear profile for the background buoyancy profile, for which no analytical results exists. Just as in the two-layer case, we see enhanced buoyancy fluctuations in the domain bulk. This numerical method can easily be applied to any background buoyancy profile, and will be applied in next section to the computation of mixing efficiency.

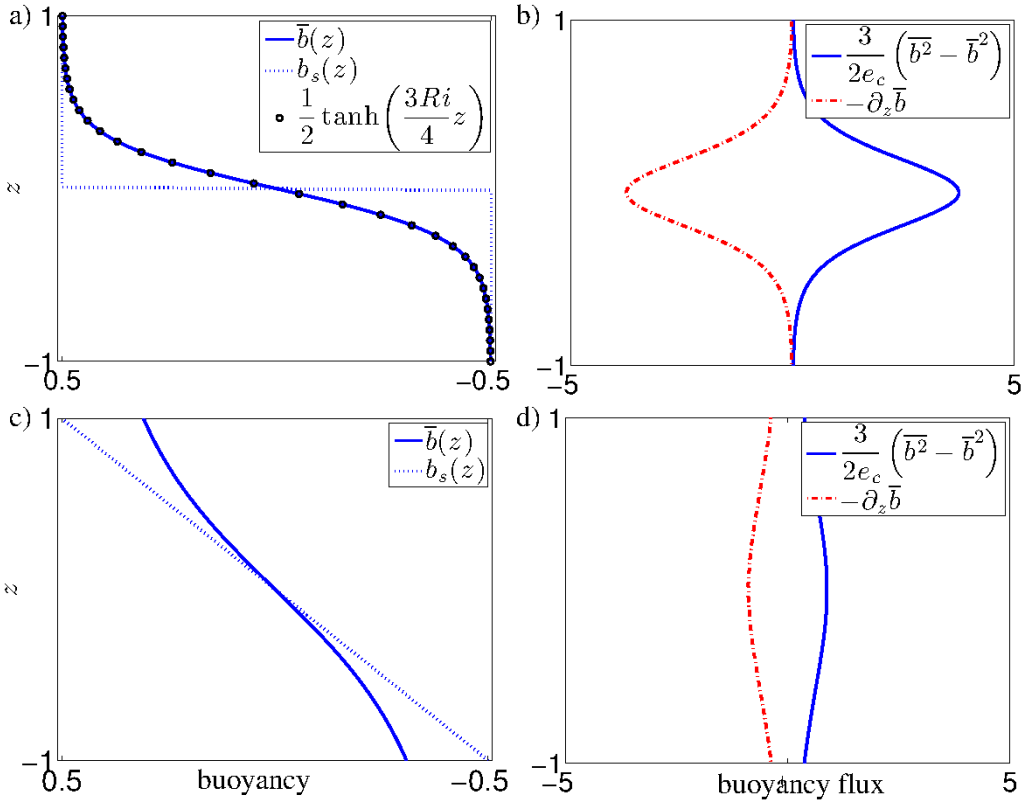

FIGURE 2. a) Plain blue line: equilibrium state $\overline{b}(z)$ computed numerically in the case $Ri = 10$, where $Ri = H\Delta b/e_c$ is the global Richardson number. Here $H = 1$, $\Delta b = 1$. Dot blue line: corresponding background buoyancy profile $b_s(z)$ (here a two-layer case). Black circles : analytical expression from Eq. (3.5) for the equilibrium state of the two-level system. The buoyancy increases from right to left on the horizontal axis. b) Compensation of the downgradient buoyancy flux with the restratification term proportional to buoyancy fluctuations (with $D = 1$). The total buoyancy is the sum of those two terms, which is zero at equilibrium. d) Same as a,b in the case of an initial linear background buoyancy profile (no analytical predictions in that case).

4. Computation of mixing efficiency in decaying flows

4.1. *Irreversibility and mixing efficiency*

We argue in the following that the computation of the equilibrium states for the inviscid, adiabatic system can be used to obtain quantitive predictions for the efficiency of mixing in decaying stratified turbulence.

The first assumption is that molecular viscosity and diffusivity only play a secondary role in the limit of large Reynolds and Péclet numbers. More precisely, we assume that the time scale to reach the equilibrium state of the inviscid, adiabatic dynamics is smaller than the typical time scale of dissipative effects. In other words, inertial dynamics govern the amount of small scale velocity and buoyancy fluctuations that are created on a short time scale, and the only effect of viscosity and diffusivity is to smooth-out these fluctuations on a longer time scale.

The second assumption is that the flow system evenly explores phase space through turbulent stirring, which is necessary to use statistical mechanics predictions. According to the theory, the macroscopic buoyancy profile $\overline{b}$ and the local distribution of small scale

fluctuations do not evolve in time anymore once the equilibrium state is reached: the equilibrium state is an attractor for the dynamics. In that respect, the purely inertial, inviscid and adiabatic dynamics is irreversible. In other words, even if the process described by the equilibrium theory is pure stirring, it implies irreversible mixing of the buoyancy field at a coarse-grained level. Assuming that this stationary property of $\overline{b}$ persists in the presence of weak viscosity and weak dissipation, we see from Eq. (2.21) that the rate of local small scale kinetic energy dissipation $\mathrm{d}\log e_c/\mathrm{d}\ t$ should be equal to the rate of dissipation for the local variance of local buoyancy fluctuations $\mathrm{d}\log\left(\overline{b^2}-\overline{b}^2\right)/\mathrm{d}t$.

Let us assume that a given amount of energy denoted E_{inj} is injected into a fluid initially at rest, characterised by a background buoyancy profile $b_s(z)$. The injected energy may either be purely kinetic (through mechanical stirring) or purely potential (for instance by turning the tank upside down into an unstable configuration). Once the equilibrium state is reached, part of this energy is carried by small scale velocity fluctuations, and the remaining part is used to maintain the potential energy of the system at a higher value than the potential energy of the background state. The coarse-grained buoyancy profile and the small scale fluctuations are decoupled when the equilibrium state is reached. This decoupling is very much similar to the effect of viscosity, which transfers energy from the degrees of freedom of the fluid motion to those of thermal fluctuations.

The total kinetic energy carried by the equilibrium state is denoted $E_c = V e_c$ with e_c the local kinetic energy density, homogeneous in space. This kinetic energy takes the form of small scale fluctuations, that will be eventually dissipated in a decaying experiment with weak viscosity, and the quantity E_c can then be interpreted as the temporal integral of viscous dissipation.

Turbulent stirring implies rearrangements of fluid parcels, and such rearrangements from $b_s(z)$ to $b(x,y,z)$ are necessarily associated with an increase of potential energy

$$E_p = -\int_{\mathcal{V}_\mathbf{x}} \mathrm{d}\mathbf{x}\ \ (b-b_s)\, z. \tag{4.1}$$

At equilibrium, this quantity can be expressed in term of the macroscopic buoyancy profile $\overline{b}$ which depends only on z:

$$E_p = -\frac{V}{2H}\int_{-H}^{+H} \mathrm{d}z\ \ \left(\overline{b}-b_s\right) z\ . \tag{4.2}$$

This definition is equivalent to the classical definition of the available potential energy. However, as explained above, the convergence towards the equilibrium buoyancy profile is irreversible. Once the equilibrium is reached, the available potential energy E_p has been irreversibly transferred to smaller scales, and can not be transferred anymore into another form of energy. It would inescapably result into molecular mixing in the presence of molecular diffusion. In that case, E_p would corresponds to the increase of the background potential energy, which is consistent with Winters *et al.* (1995).

We define the mixing efficiency as

$$\eta \equiv \frac{E_p}{E_p + E_c}, \tag{4.3}$$

where $E_p + E_c = E_{inj}$ is the total energy injected into the system. This definition of mixing efficiency is bounded between 0 and 1. Since E_c is the total amount of kinetic energy lost at small scale, and since E_p corresponds to an irreversible increase of potential energy according to statistical mechanics theory, our definition of η is equivalent to the

long time limit of the *cumulative mixing efficiency* (Peltier & Caulfield 2003), or to the *integrated flux Richardson number* (Linden 1979).

We stress finally that the equilibrium theory does not predict a temporal evolution for the system but just the final outcome of turbulent stirring under the assumption of random evolution without forcing and dissipation. It provides therefore a global (integrated over the whole domain) and cumulative (integrated over sufficiently large time) prediction for the efficiency of mixing. We will discuss in subsection 4.4 how the statistical mechanics predictions may be used as a guide for a parameterisation of the instantaneous, local mixing efficiency.

4.2. *Numerical computation in the general case*

We show in Fig. 3 how the mixing efficiency η varies with the global Richardson number $Ri = H\Delta b/e_c$, with $\Delta b = b_s(H) - b_s(-H)$. We consider two different buoyancy profiles $b_s(z)$: case (a) is the two-level configuration corresponding to a background profile with two homogeneous layers of equal depth, for which an analytical solution exits; case (b) corresponds to a linear background buoyancy profile. Considering those two cases allows us to show very different behaviour for the variations of mixing efficiency as a function of the Richardson number Ri.

The kinetic energy e_c appearing in the Richardson number is not a control parameter, but one can check *a posteriori* that $E_c = Ve_c$ is always of the same order of magnitude as the injected energy E_{inj}, which is a control parameter. In a direct numerical simulations with non-zero viscosity, E_c would be the actual amount of kinetic energy dissipated during the turbulent decay.

We see in Fig. 3 that whatever the background buoyancy profile, the equilibrium buoyancy profile $\overline{b}$ can be considered as almost completely homogenised in the low Richardson number limit $Ri \ll 1$. In that case, most of the injected energy is lost in small-scale velocity fluctuations with $E_c = E_{inj}$ and the fluid is well mixed, so that $\overline{b}$ is a constant, and Eq. (4.2) reduces to $E_p = \frac{V}{2H}\int_{-H}^{+H} b_s z \mathrm{d}z$. The mixing efficiency is then given by

$$\eta =_{Ri\ll 1} Ri\ \Xi[b_s] \quad \text{with} \quad \Xi[b_s] \equiv \frac{1}{2\Delta b H^2}\int_{-H}^{+H} b_s z \mathrm{d}z \quad \text{when } Ri \ll 1. \tag{4.4}$$

The numerical coefficient $\Xi[b_s]$ is bounded in [0 1] and characteristic of the shape of the background buoyancy profile, hence of the distribution of available densities. It is equal to 0 for a homogeneous fluid, 1/6 for a linear stratification and 1/4 for a two-layer system. Whatever this background buoyancy profile, the mixing efficiency scales linearly with the Richardson number is the limit of weak Richardson numbers.

By contrast, we see in Fig. 3 that the large Richardson behaviour of the mixing efficiency depends drastically on the background buoyancy profile b_s: the mixing efficiency decreases to zero with increasing Richardson numbers in the two-level case of Fig. 3a, while it increases to an asymptotic value close to 0.25 in the linearly stratified case of Fig. 3b. We show analytically in the next subsection that an asymptotic value of $\eta = 0.25$ is indeed expected in a low energy limit, as a consequence of energy equipartition, provided that the stratification of the background profile is always strictly positive ($\partial_z b_s > 0$ for $-H \leq z \leq H$).

4.3. *Energy equipartition and mixing efficiency for high Richardson numbers*

The potential energy E_p defined in Eq. (4.1) is a linear functional of $b - b_s$, which is *a priori* sign indefinite. However, the conservation of the global distribution of buoyancy levels (prescribed by b_s) provides a strong constraint on admissible buoyancy levels b,

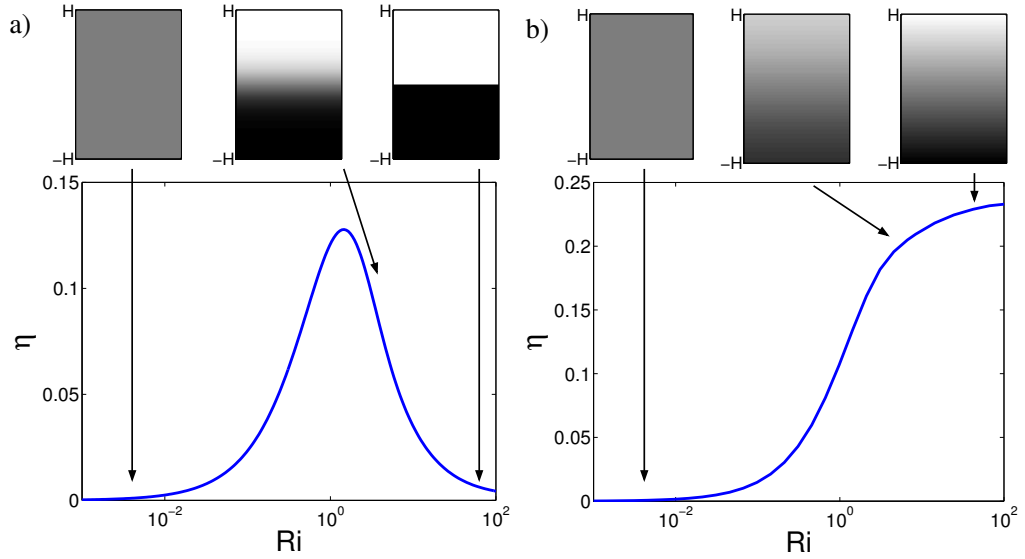

FIGURE 3. Variation of the mixing efficiency $\eta = E_p/E_{inj}$ with the Richardson number $Ri = H\Delta b/e_c$ (a) for a background buoyancy profile with two homogeneous layers, (b) for an initial linear background buoyancy profile. The three insets show the equilibrium buoyancy field $\overline{b}$ for $Ri = 0.07, 7, 70$.

and hence on admissible values for E_p. A direct consequence of these conservation laws it that the potential energy is strictly positive unless $b = b_s$. Denoting $Z_s(b_s)$ the height of fluid particles carrying buoyancy level b_s in the background buoyancy profile, using an asymptotic expansion in term of $b - b_s$ and assuming $\mathrm{db_s/dz} = \mathrm{b_s'} > 0$, one can use the conservation laws related to buoyancy (Casimir functionals) to obtain an explicit quadratic form for the potential energy in a weak energy limit (Shepherd 1993):

$$E_p = \frac{1}{2}\int_{\mathcal{V}_\mathbf{x}} \mathrm{d}\mathbf{x} \frac{(b-b_s)^2}{b_s'} + O\left(Z_s''\,(b-b_s)^3\right) \ . \tag{4.5}$$

The quadratic part is also the classical expression of the potential energy for internal gravity waves, derived for instance in (Gill 1982).

Decomposing the spatial integral of Eq. (4.5) into a sum of integrals over each macrocell of the discrete model depicted in Fig. 1 and taking the limit of an infinite number of macrocells, the potential energy can be expressed in terms of the local variance of buoyancy fluctuations:

$$E_p = \frac{V}{4H}\int_{-H}^{+H} \mathrm{d}z \ \frac{\overline{b^2} - \overline{b}^2}{b_s'} + O\left(Z_s''\overline{(b-b_s)^3}\right) \ , \tag{4.6}$$

The variance of buoyancy fluctuations $\overline{b^2} - \overline{b}^2$ is related to the local kinetic energy e_c and the local buoyancy gradient $\mathrm{d\overline{b}/dz}$ though Eq. (2.21). Inserting this equation into Eq. (4.6), using $\mathrm{d\overline{b}/dz} = \mathrm{b_s'}$ (which is valid for sufficiently large Richardson numbers), and $e_c = E_c/V$:

$$E_p = \frac{E_c}{3}, \tag{4.7}$$

which shows equipartition of the energy between the available potential energy and the three degrees of freedom of the kinetic energy. A direct consequence of energy equiparti-

tion is thus

$$\eta = \frac{1}{4}. \tag{4.8}$$

This result is a direct consequence of the quadratic form of the energy obtained in Eq. (4.6), which relies on the assumptions (i) that b'_s is strictly positive and bounded (ii) that Z''_s is bounded (iii) that b remains sufficiently close to b_s.

Importantly, the hypotheses (i) and (ii) are not satisfied when the background buoyancy profile contains homogeneous layers of fluids, as for instance in the case depicted in Fig. 3a. In order to evaluate when the assumption (iii) is valid, one can estimate the typical value of $b - b_s$ at a given point as the root mean square of local buoyancy fluctuations $\left(\overline{b^2} - \overline{b}^2\right)^{1/2}$. Using Eq. (2.21), $\partial_z \overline{b} = \partial_z b_s$ and $Ri \sim H^2 b'_s / e_c$ yields then $(b - b_s) \sim Ri^{-1/2}$.

We conclude that $\eta = 0.25$ is expected in the limit of large Richardson number, when the background buoyancy profile is strictly increasing with height.

4.4. *Comparison with previous studies of the efficiency of mixing*

Despite the large number of numerical and experimental studies devoted to the understanding of mixing efficiency, there are only few theoretical results yielding predictions for the variations of mixing efficiency with the Richardson number. In the context of shear-stratified turbulence, dimensional analysis is used by Townsend (1958) to model the variation of mixing efficiency with the gradient Richardson number, and upper bounds for the mixing efficiency have been derived rigorously by Caulfield & Kerswell (2001).

A mixing efficiency efficiency $\eta = 0.25$ was obtained in the framework of a phenomenological model due to McEwan (1983*b*), based on purely kinematic arguments. Those predictions were found to be consistent with experimental observations of mixing efficiency following an internal wave-breaking event (McEwan 1983*a*). The argument is the following: take a continuously stratified fluid at rest, and exchange two particle fluids a and b of volume δV with buoyancy difference $\Delta b = b_b - b_a$ and height difference $\Delta z = z_b - z_a$, with $\delta V / V \ll \Delta z / H$. Then consider the small displacement limit $\Delta z \to 0$, which, as explained in previous paragraphs, corresponds to a weak energy limit, or equivalently to a large Richardson number limit, for which $\Delta b = \Delta z \mathrm{d}\overline{b}/\mathrm{dz}$. Given that the injected energy is under the form of available potential energy only, the initial kinetic energy is zero, with $E_{inj} = b'_s (\Delta z)^2 \delta V$. McEwan (1983*b*) then argued that the two displaced fluid particles will be stirred and mix together until homogenisation of their buoyancy, and that the two fluid particles carrying buoyancy $(b_a + b_b)/2$ will "sediment" to their rest position $z = (z_a + z_b)/2$. The available potential energy of the final state is then $E_p = b'_s (\Delta z/2)^2 \delta V$, which corresponds to mixing efficiency $\eta = 0.25$.

Strikingly, several numerical studies have also reported convergence of mixing efficiency towards $\eta = 0.25$ at large Richardson numbers; see e.g. Maffioli *et al.* (2016); Venayagamoorthy & Koseff (2016) and references therein†. It is remarkable that the statistical mechanics theory in the large Richardson number limit also yields $\eta = 0.25$. We stress that the only assumption underlying the equilibrium theory is that the system evenly explores the phase space: there is neither dynamics nor kinematics involved in the derivation of this result. By contrast, the approach of McEwan (1983*b*) relies on the choice of a peculiar kinematic model.

McEwan (1983*b*) also discussed the case of two homogeneous layers separated by a linear pycnocline of thickness δ. He found that mixing efficiency vanishes when consid-

† Maffioli *et al.* (2016) report a mixing coefficient $\Gamma = \eta/(1-\eta) = 0.33$ in the limit of small Froude numbers, which corresponds to $\eta = 0.25$ in the limit of large Richardson numbers.

ering first the limit $\delta \to 0$ and second the limit of Large Richardson numbers $Ri \to +\infty$. This is again fully consistent with the statistical mechanics predictions for the mixing efficiency in the two-level case depicted in Fig. 3a. Indeed, this case corresponds to the case to a background buoyancy profile with an infinity sharp interface. Qualitatively, the fact that mixing efficiency vanishes in the limit of infinite Richardson numbers when the background buoyancy profile is made of two homogeneous layers is due to the fact that kinetic energy is spread equally over the whole domain at equilibrium, while buoyancy mixing is confined to a thin layer surrounding the buoyancy interface, with a thickness that decrease with the Richardson number, as explained in subsection 3.1.

We stress that the statistical theory makes possible predictions for global, cumulative mixing efficiency in decaying turbulence predicts a value for mixing efficiency whatever the Richardson number, and whatever the background buoyancy profile. In particular, it predicts a bell shape for $\eta(Ri)$ in the two-layer case, with a maximum $\eta = 0.15$, and a monotonic increase of $\eta(Ri)$ in the linear case from $\eta = 0$ to $\eta = 0.25$, as shown Fig. 3. The bell shape for $\eta(Ri)$ has been reported in decaying experiments performed by dropping a grid in a two-layer stratified fluid (Linden 1980), but how to estimate the amount of energy injected into the system in such experiment remains debated, see e.g. Huq & Britter (1995). The monotonic increase of cumulative mixing efficiency in the case of a linear background buoyancy profile seems a robust result in laboratory and numerical experiments, see e.g. Stretch *et al.* (2010). However, the equilibrium theory does not account for layering which is often observed in the strongly stratified regime $Ri \gg 1$ (Rehmann & Koseff 2004). In any case, the statistical mechanics prediction that mixing efficiency depends strongly on the global shape of the background buoyancy profile, and not only on the local buoyancy gradient is consistent with observations by Holford & Linden (1999).

There is however one result that does not depend on the shape of the buoyancy profile: according to the equilibrium theory, the mixing efficiency should increase linearly with the Richardson numbers in the limit of weak Richardson numbers. This scaling law can be simply understood as a consequence of the fact that buoyancy behaves as a passive tracer in this limit (Holford & Linden 1999). This linear scaling has been also reported by Maffioli *et al.* (2016) in forced-dissipative numerical experiments, who also provide complementary arguments based on cascade phenomenology.

According to the statistical mechanics theory, the value of mixing efficiency in decaying turbulence depends on the total energy injected into the system, but not on how the energy is injected. However, different values of mixing efficiency have been reported in laboratory and numerical experiments performed with different energy injection mechanism. A value $\eta \approx 0.2$ was reported in decaying sheared-stratified fluids with a Richardson number of order one (**?**). This value is somewhat larger than the cumulative mixing efficiency $\eta = 0.11$ observed in lock-exchange experiments Prastowo *et al.* (2008); Ilıcak (2014), and smaller than the cumulative mixing efficiency $\eta \approx 0.5$ reported in the framework Rayleigh-Taylor experiments (Dalziel *et al.* 2008; Wykes & Dalziel 2014). Importantly, these different values for mixing efficiency do not depend only on the Richardson, Reynolds and Péclet numbers. This suggests that the mechanism of injection plays an important role.

One heuristic way to discuss more precisely the role of the injection mechanism in relation with the ergodicity hypothesis is to consider the parameter L_t/H, i.e. the ratio of the energy injection length scale to the domain scale. In the context of two-dimensional turbulence, this parameter has been proven useful to discuss the relevance of the ergodic hypothesis underlying statistical mechanics theory (Pomeau 1994; Tabeling 2002; Venaille

et al. 2015). Denoting T_{tran} the typical time scale to move a fluid particle from the top to the bottom of the tank through turbulent transport, and calling $T_{diss} = L_t/U$ the typical time scale for the dissipation rates for local buoyancy fluctuations through direct turbulent cascade, the system can explores the phase space only if $T_{trans} < T_{diss}$. Modelling turbulent transport as an effective eddy viscosity or eddy diffusivity UL_t yields $T_{trans} = H^2/(UL_t)$ and then the necessary condition $H < L_t$ for ergodicity. Given that L_t can not be larger than the domain height, we see that the condition for sufficient mixing in phase space will only be marginally satisfied when $H = L_t$, and will not be satisfied when $L_t \ll H$.

The equilibrium theory applies in principle to flow systems in the limit of infinitely large Reynolds and Péclet numbers. Even if it is natural to expect that the dissipation rate of buoyancy and kinetic energy become independent from the value of molecular viscosity and diffusivity when they are sufficiently weak, numerical and laboratory experiments are often performed in intermediate regimes where those parameters may influence the mixing efficiency, see e.g. Shih *et al.* (2005); Lozovatsky & Fernando (2013); Bouffard & Boegman (2013); Salehipour & Peltier (2015). The dependence of mixing efficiency on the parameters (Ri, Re, Pe) is often described in terms of the triplet (Ri, Re_b, Pr), where Pr is the Prandtl (Schmidt) number and Re_b is the buoyancy Reynolds number. In forced-dissipative configurations the buoyancy Reynolds is defined as $Re_b = \epsilon/\nu N^2$ with N the buoyancy frequency and ϵ the kinetic energy dissipation rate. Several studies reported that mixing efficiency η is a highly non-monotonic function of buoyancy Reynolds number and Richardson number in the range of high but finite Reynolds number, with a single extremum found to be close to the Osborn number of 0.2, see e.g. Salehipour & Peltier (2015) and references therein. It is interesting that the statistical mechanics based estimate delivers a result that is close to this extremum.

4.5. *A theory for the Mellor-Yamada parameterisation*

A practical application of the studies of the mixing efficiency in stratified turbulence is the development of parameterisations for the diapycnal eddy diffusivity used in numerical ocean models. Using a number of assumptions (among which stationarity, homogeneity, and a compensation between production and dissipation of local buoyancy variance), Osborn & Cox (1972) proposed to model the diapycnal eddy-diffusivity as $K = \epsilon_b/N^2$, where ϵ_b stands for the mean potential energy dissipation per unit mass and N the local buoyancy frequency. Denoting ϵ the local kinetic energy dissipation rate, $R_f = \epsilon_b/(\epsilon_b+\epsilon)$ the local mixing efficiency, also referred to as the (irreversible) flux Richardson number, and $\Gamma = \epsilon_b/\epsilon = R_f/(1 - R_f)$ the *mixing coefficient*, the diapycnal eddy-diffusivity can be then written as $K = \Gamma\epsilon/N^2$. Following the prescription of Osborn (1980), the mixing coefficient Γ is often modelled as a constant $\Gamma = 0.2$, which corresponds to $R_f = 0.17$.

As explained previously, one may however expect strong variations of the mixing efficiency with the Richardson number. At low and moderate Richardson numbers, stratified turbulence results generally from shear instabilities. These instabilities are characterised by the gradient Richardson number $Ri_g = N^2/(\partial_z U)^2$, with $\partial_z U$ the local mean flow gradient and N the local buoyancy frequency. Empirical formula taking into account the dependence of the flux Richardson number R_f on the gradient Richardson number Ri_g have been proposed (Mellor & Yamada 1982; Nakanish 2001; Karimpour & Venayagamoorthy 2014), and compared with numerical and laboratory experiments Pardyjak *et al.* (2002); Venayagamoorthy & Koseff (2016). Such empirical parameterisations have also been extended to include the effect of Reynolds Buoyancy number and Prandtl number, by considering a large dataset of direct numerical simulations (Salehipour *et al.* 2016). One drawback of approaches based on the analysis of direct numerical simulations

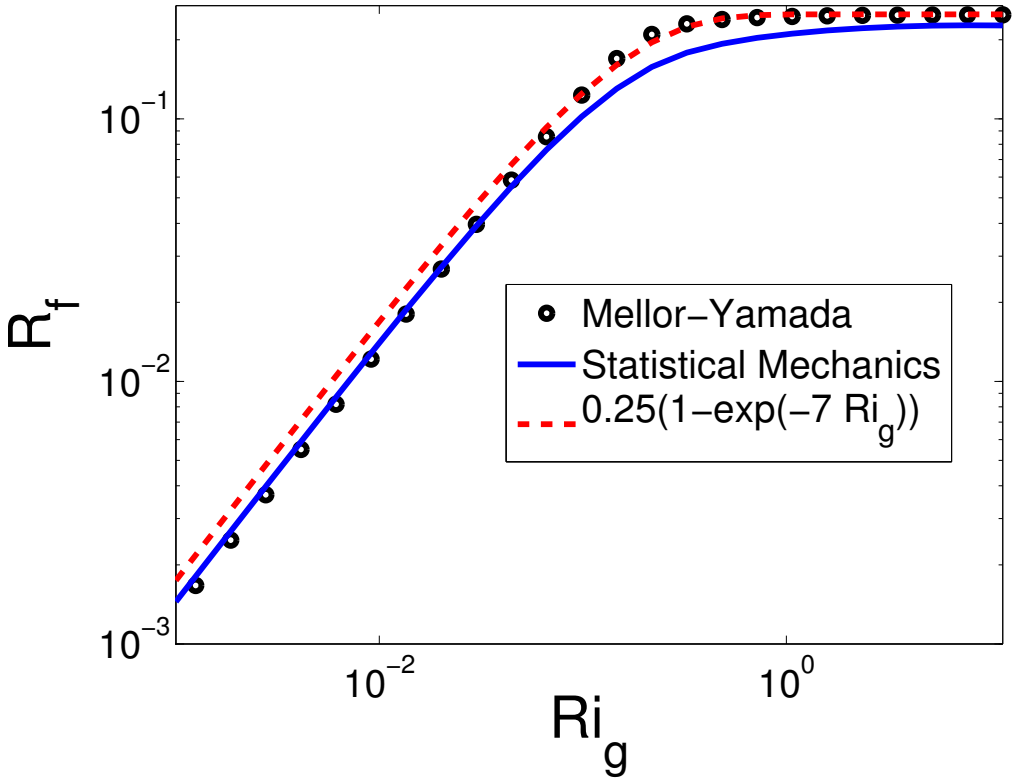

FIGURE 4. Flux Richardson number R_f as a function of the gradient Richardson number Ri_g. The statistical mechanics prediction corresponds to the linear case plotted in Fig 3b, with the assumption $Ri_g = 0.12Ri$. The Mellor-Yamada curve has been plotted using formula B6 in (Karimpour & Venayagamoorthy 2014). The curve $R_f = 0.25\,(1 - \exp(-7Ri_g))$ is used by (Venayagamoorthy & Koseff 2016) to fit the $R_f - Ri_g$ relation observed in the DNS of Shih *et al.* (2005).

is the assumption that the dynamical process leading to mixing in the simulation is the relevant one everywhere in the ocean. The advantage of considering statistical mechanics predictions is to obtain results that are independent from any dynamical mechanism, in the limit of large Reynolds and Péclet numbers.

Several assumptions are necessary to interpret these parameterisations in the framework of the equilibrium statistical mechanics theory. (i) One first needs to assume that the global cumulative mixing efficiency predicted by the equilibrium theory in a tank of height $2H$ can be interpreted as the instantaneous mixing efficiency R_f inside a grid cell of height $2H$ in the ocean model. This assumption is also necessary when decaying numerical experiments are used to infer the dependency of local, instantaneous mixing efficiency on external parameters. In addition, one should keep in mind that the typical spatial and temporal scales of mixing events are much smaller than typical time and spatial scales resolved by the ocean model, so the instantaneous and local quantities described by those models are already coarse-grained quantities. (ii) One then needs to choose a relevant background buoyancy profile, which is an input of the theory. The simpler choice is to consider a linear buoyancy profile interpolating the buoyancy jump Δb between two adjacent layers on the vertical in the ocean model. This buoyancy jump can be expressed in term of the local buoyancy frequency N as $\Delta b = 2HN^2$. (iii) One finally

needs to relate the global Richardson number $Ri = H\Delta b/e_c$ considered in the statistical theory to the gradient Richardson number $Ri_g = N^2/\left(\partial_z U\right))^2$ considered in the empirical parameterisations. Using $N^2 = \Delta b/2H$, we get the relation $Ri_g/Ri = e_c/\left(2((H\partial_z U))^2\right)$. Empirical observations relate the turbulent kinetic energy e_c to the local vertical shear $\partial_z U$, as reviewed in Mellor & Yamada (1982). Here we choose $Ri_g/Ri = 0.12$, this ratio being a fitting parameter consistent with the values reported in Mellor & Yamada (1982).

With those assumptions, the $\eta - Ri$ relation plotted in Fig. 3b can be considered as a parameterisation for the $R_f - Ri_g$ relation, which is plotted in Fig. 4. Even if the statistical mechanics predictions converge less rapidly to $R_f = 0.25$ than the empirical parameterisations. It is remarkable that the statistical mechanics theory yields to a predictions close to the empirical formula used in the level 2 model of Mellor & Yamada (1982) hierarchy, which can be approximated by an exponential fit (Karimpour & Venayagamoorthy 2014), and which has been recently tested with success against direct numerical simulations (Venayagamoorthy & Koseff 2016). We stress that the only fitting parameter is the coefficient Ri_g/Ri. In addition, the fact that the statistical mechanics theory captures both the scaling $R_f \sim Ri_g$ in the weak Richardson number limit and the convergence towards $\eta = 0.25$ in the large Richardson limit is independent from this fitting parameter Ri_g/Ri, which is only shifting the abscissa in figure 4.

5. Conclusion

We have addressed the problem of mixing efficiency from the point of view of equilibrium statistical mechanics. The theory predicts that the unforced, inviscid, adiabatic dynamics is attracted towards a state characterised by small scale velocity fluctuations carrying kinetic energy, and by a smooth, monotonic buoyancy profile superimposed with small scale buoyancy fluctuations. Although the whole dynamics is adiabatic, the buoyancy field is irreversibly mixed at a coarse-grained level, no matter how small the coarse-grained scale. In addition, the coarse-grained fields predicted by the theory are stationary, characterised by a stable buoyancy profile. The theory also predicts velocity fluctuations are Gaussian, isotropic, homogeneous in space, and that the buoyancy fluctuations are homogeneous on horizontal planes.

The input of the theory is the total energy injected initially into the system, and the global distribution of buoyancy levels, or equivalently the background buoyancy profile. The output of the theory is the probability to measure a given buoyancy level at each height. We provide explicit computations of the equilibria in limiting cases, and implement an algorithm based on a maximum entropy production which determines equilibrium state for any background buoyancy profile. This allows us to compute a cumulative mixing efficiency defined as the ratio of the potential energy gained by the system to the total energy injected into the system. Importantly, the potential energy effectively gained by the system is the potential energy of the coarse-grained buoyancy profile at equilibrium minus the potential energy of the background buoyancy profile. The background potential energy remains constant for the adiabatic dynamics, but the irreversible convergence of the system towards the equilibrium state implies an irreversible increase of potential energy for the system. Several important results or predictions on the cumulative mixing efficiency are obtained within this framework:

(a) The cumulative mixing efficiency increases in proportion to the Richardson number in the limit of small Richardson number, whatever the background buoyancy profile.

(b) The cumulative mixing efficiency tends to 0.25 in the limit of infinite Richardson

numbers, provided that the background buoyancy profile is strictly decreasing with height (no homogeneous layer).

(c) The variations of the cumulative mixing efficiency with the Richardson number depends strongly on the background buoyancy profile, and can be non-monotonic. In the particular case of a fluid with two homogeneous layers of different buoyancy, the theory predicts a bell-shape for the cumulative mixing efficiency as a function of the global Richardson number.

(d) When the background buoyancy profile is linear, the variation of the cumulative mixing efficiency with the Richardson number is monotonic, and the shape of the curve predicted by the equilibrium theory is consistent with empirical parameterisations for the variations of the flux Richardson number with the gradient Richardson number, see e.g. Mellor & Yamada (1982); Nakanish (2001); Karimpour & Venayagamoorthy (2014); Venayagamoorthy & Koseff (2016).

To the best of our knowledge, there is so far no other theory that provides such predictions in a unified framework. There remain, however, several caveats for the application of the statistical mechanics theory.

(a) The theory applies to fluids in the limit of infinite Reynolds and Péclet number, and existing laboratory and numerical experiments are usually carried in intermediate regimes where mixing efficiency can be affected by finite values of molecular viscosity and diffusion. Further work will be necessary to find how large the Reynolds and the Péclet numbers should be so that the mixing efficiency becomes independent of those parameters.

(b) Equilibrium statistical mechanics relies on the counting of the available microscopic states, and its predictive power depends on the capability of the system to actually explore those available states. Such an ergodic behaviour is favoured by stirring at the system scale H. By contrast turbulence forced at small scale L_t with $L_t/H < 1$ is expected to produce local mixing before large scale stirring, leading to discrepancies of the statistical mechanics predictions.

(c) Finally, the equilibrium theory does not predict how the system converges towards equilibrium, or what would be the energy fluxes in a forced-dissipative case.

Equilibrium statistical mechanics therefore describes an ideal state of inviscid stirring which is not fully reached in most cases. Turbulent stirring can be however modelled locally as a trend to approach this equilibrium. This can be done by giving a dynamical meaning to the relaxation equations used in this paper as an algorithm to compute the equilibrium state. Indeed, those equations contain a classical term modelling turbulent transport as an effective diffusion, with an additional drift term describing restratification. We believe that this approach will be fruitful to model relaminarisation after a mixing event, or to describe the spontaneous emergence of a sharp interface, see e.g. Venaille & Sommeria (2010). Other models will be needed to account for energy fluxes across scales, which is another important aspect of stratified turbulence; see e.g Godeferd & Cambon (1994); Brethouwer *et al.* (2007); Rorai *et al.* (2014). More generally statistical or stochastic approaches have long been used in the context of combustion (Pope 1985), and adapting those methods to the case of turbulent mixing in stratified fluids has been advocated by Kerstein (1999). We hope the present paper will motivate further studies in those directions.

We warmly thank C. Staquet, T. Dauxois and C. Herbert for useful discussions, and P. Odier for his very detailed comments on the manuscript. LG and AV were partly funded by the ANR-13-JS09-0004-01 (STRATIMIX).

Appendix A. Liouville theorem

We show in this appendix that the quadruplet of fields $(\mathbf{u}, b)$ satisfy a Liouville theorem, i.e. that trajectories of the system are non-divergent in a phase-space described by this quadruplet of fields. The fact that Fourier components of the velocity field in each direction satisfy a detailed Liouville theorem is a classical result for three-dimensional Euler dynamics (Lee 1952). Generalisation of this results to the inviscid, adiabatic Boussinesq system is straightforward, but is reproduced here for completeness. Let us for that purpose decompose both the velocity field and the buoyancy field on Fourier modes:

$$\mathbf{u} = \sum_{\mathbf{k}} \hat{\mathbf{u}}_{\mathbf{k}}(t) e^{i\mathbf{k}\cdot\mathbf{x}}, \quad b = \sum_{\mathbf{k}} \hat{b}_{\mathbf{k}}(t) e^{i\mathbf{k}\cdot\mathbf{x}}. \tag{A 1}$$

Writing $\mathbf{u} = (u_1, u_2, u_3)$, projecting the equations of motion (2.1)-(2.2)-(2.3) on a mode with wavenumber $\mathbf{k}$ yields to

$$\dot{\hat{b}}_{\mathbf{k}} = -\mathrm{i} \sum_{\mathbf{p}+\mathbf{q}=\mathbf{k}} (\hat{\mathbf{u}}_{\mathbf{p}} \cdot \mathbf{q})\, \hat{b}_{\mathbf{q}}, \tag{A 2}$$

$$\dot{\hat{u}}_{i\mathbf{k}} = \sum_{j,l} \left[\left(\delta_{i3} - \frac{k_3 k_i}{k^2} \right) \hat{b}_{\mathbf{k}} + \left(\frac{k_i k_j}{k^2} - \delta_{ij} \right) \sum_{\mathbf{p}+\mathbf{q}=\mathbf{k}} q_l \hat{u}_{l\mathbf{p}} \hat{u}_{j\mathbf{q}} \right]. \tag{A 3}$$

The pressure term has been eliminated from the momentum equation by using the non-divergence condition. Deriving Eq. (A 2) by $b_{\mathbf{k}}$ and Eq. (A 3) by $u_{i\mathbf{k}}$ allows us to show the existence of a detailed Liouville theorem for the Fourier components of the buoyancy field b, and for the Fourier components of the velocity field in each direction:

$$\forall \mathbf{k}, \quad \frac{\partial \dot{b}_{\mathbf{k}}}{\partial \hat{b}_{\mathbf{k}}} + \frac{\partial \dot{\hat{b}}_{-\mathbf{k}}}{\partial \hat{b}_{-\mathbf{k}}} = 0, \quad \text{and } \forall i, \mathbf{k} \quad \frac{\partial \dot{\hat{u}}_{i\mathbf{k}}}{\partial \hat{u}_{i\mathbf{k}}} + \frac{\partial \dot{\hat{u}}_{i-\mathbf{k}}}{\partial \hat{u}_{i-\mathbf{k}}} = 0. \tag{A 4}$$

Using $(u_1, u_2, u_3) = (u, v, w)$, we conclude that the quadruplet of fields (u, v, w, b) satisfies a Liouville theorem:

$$\sum_{\mathbf{k}} \left[\frac{\partial \dot{\hat{b}}_{\mathbf{k}}}{\partial \hat{b}_{\mathbf{k}}} + \frac{\partial \dot{\hat{u}}_{\mathbf{k}}}{\partial \hat{u}_{\mathbf{k}}} + \frac{\partial \dot{\hat{v}}_{\mathbf{k}}}{\partial \hat{v}_{\mathbf{k}}} + \frac{\partial \dot{\hat{w}}_{\mathbf{k}}}{\partial \hat{w}_{\mathbf{k}}} \right] = 0. \tag{A 5}$$

This Liouville theorem expresses the conservation of volume in the space of spectral amplitudes. However the discrete approximation of the fields that we propose in this paper relies on a uniform microscopic grid in physical space, and one needs to show that the Liouville property is not broken by this discrete approximation. We note for that purpose (i) that the Liouville property in Eq. (A 5) remains valid if the sum is truncated at wavenumbers $k_i \leq N/2$ for $1 \leq i \leq 3$, whatever the value of N, and (ii) that for a given truncation of the fields in Fourier space, the spectral amplitudes are related to the values of the fields on a collocation grid uniform in physical space, through a linear transformation that does no depend on the fields. The Jacobian of the transformation is therefore an unimportant constant, as noted in Miller (1990). We conclude that a Liouville theorem holds for the finite-dimensional approximation of the buoyancy and velocity fields on a uniform grid.

Appendix B. From Boltzmann entropy to macrostate entropy

The aim of this appendix is to count the number of microscopic configurations $\mathbf{u}(\mathbf{x}), b(\mathbf{x})$ associated with a given macroscopic state $\rho(\mathbf{x}, \sigma, \mathbf{v})$. In order to simplify the presentation, we show first how to count the number of microscopic configurations $b(\mathbf{x})$ associated with

a given macroscopic state $\rho(\mathbf{x}, \sigma)$. The first step is to introduce a discrete approximation of the fields. The second step is a classical counting arguments within each macrocell of the discrete model. The third step is to consider the limit of an infinite number of grid point within each macrocell, which corresponds to the continuous limit for the microscopic configurations. The last step is to consider the limit of an infinite number of macrocells, which corresponds to the continuous limit for the macroscopic states, or equivalently to the limit of a vanishing coarse-graining length scale.

We assume that the domain $\mathcal{V}_{\mathbf{x}}$ is divided into a uniform grid containing N cubic macrocells indexed by $1 \leq I \leq N$, and that each macrocell is divided into another uniform grid containing M sites, where each site contains one and only one fluid particle indexed by $1 \leq i \leq M$, see Fig. 1. We also assume that the buoyancy $b_{I,i}$ at site (I, i) can only takes a discrete number of values (say K), with $b_{I,i} \in \{\sigma_1, ..., \sigma_K\}$, and that each of the resulting microstates is equiprobable. We note that with this procedure, we count fields that will not be differentiable when taking the continuous limit, and we will see that the equilibrium state is actually generally characterised by wild small scale fluctuations of buoyancy. For a given discretised buoyancy field, we call $M_{I,k}$ the number of fluid particles carrying the buoyancy level σ_k within the macrocell I, and $M_{I,k}/M$ is therefore the frequency of occurrence of the level σ_k at site I for one realisation of the discretised field. The system is described at a macroscopic level by the probability $p_{I,k}$ of measuring the buoyancy level σ_k at site I.

Our aim is to count number of microscopic configurations associated with a prescribed field $p_{I,k}$. We use for that purpose the equivalence between probability and frequency in the large M limit:

$$p_{I,k} = \lim_{M \to +\infty} \frac{M_{I,k}}{M} \ . \tag{B 1}$$

In the large M limit, the number of microscopic discretised buoyancy fields $\{b_{I,i}\}_{1 \leq I \leq N, 1 \leq i \leq M}$ associated with the macroscopic field $\{p_{I,k}\}_{1 \leq I \leq N, 1 \leq k \leq K}$ is

$$\Omega = \prod_{I=1}^{N} \left(\frac{M!}{\prod_{k=1}^{K} (M p_{I,k})!} \right) \ . \tag{B 2}$$

The Boltzmann entropy is defined as

$$S_B = k_B \log \Omega \ , \tag{B 3}$$

where k_B is a constant. In the large M limit, Stirling formula ($\log M! = M \log M$) leads at lowest order to

$$S_B = -k_B M \sum_{I=1}^{N} \sum_{k=1}^{K} p_{I,k} \log p_{I,k}, \tag{B 4}$$

where we have kept only the dominant term, and removed an unimportant constant depending on the grid size M.

It is important to note that for a given macrostate $\{p_{I,k}\}_{1 \leq I \leq N, 1 \leq k \leq K}$, the number of possible microscopic configurations Ω diverges exponentially with M, which a coefficient given by $-\sum_{I=1}^{N} \sum_{k=1}^{K} p_{I,k} \log p_{I,k}$. This means that among a set of different macrostates, there will be an overwhelming number of microstates associated with the one that maximises the coefficient $-\sum_{I=1}^{N} \sum_{k=1}^{K} p_{I,k} \log p_{I,k}$. In other words, a single microscopic configuration picked up at random has a very large probability of being close to the macroscopic equilibrium state. A practical consequence of this *concentration property* is that no particular average procedure is required to observe the actual macroscopic

equilibrium state.

In the limit $N \to +\infty$, the sum over I in Eq. (B 4), can be replaced by an integral over the spatial coordinate $\mathbf{x}$ if the discretised probability field $\{p_{I,k}\}_{1\leq I\leq N, 1\leq k\leq K}$ is also replaced by its continuous counterpart $\{p_K(\mathbf{x})\}_{1\leq k\leq K}$:

$$S_B = -k_B \frac{MN}{V} \int_{\mathcal{V}_\mathbf{x}} \mathrm{d}\mathbf{x} \sum_{k=1}^{K} p_k(\mathbf{x}) \log p_k(\mathbf{x}), \tag{B 5}$$

Note that the quantity $p_k(\mathbf{x})$ is normalised at each point $\mathbf{x}$, with $\sum_{k=1}^{K} p_k(\mathbf{x}) = 1$. It describes the local fluctuations of the (continuous) microscopic field b in the vicinity of point $\mathbf{x}$, and it is called a *Young measure* in mathematics.

A generalisation to the case of a continuum of buoyancy levels $\sigma \in \mathcal{V}_\sigma = [\sigma_{min}\ \sigma_{max}]$ with probability density function $\rho(\sigma, \mathbf{x})$ is less straightforward and requires the use of Sanov's theorem, see e.g. Touchette (2009). However, the result is easily inferred from Eq. (B 5) by decomposing the interval $[\sigma_{min}\ \sigma_{max}]$ into K levels σ_k equally spaced with interval $\Delta\sigma$, and by considering $\rho(\mathbf{x}, \sigma_k) = p_k(\mathbf{x})/\Delta\sigma$. Taking the limit $K \to +\infty$ yields

$$S_B = -k_B \frac{MN}{V} \int_{\mathcal{V}_\mathbf{x}} \mathrm{d}\mathbf{x} \int_{\mathcal{V}_\sigma} \mathrm{d}\sigma\ \rho(\mathbf{x}, \sigma) \log \rho(\mathbf{x}, \sigma), \tag{B 6}$$

up to an unimportant term depending on K. The quantity $\rho(\mathbf{x}, \sigma)$ is now the probability density function of measuring the buoyancy level $b = \sigma$ at height z, with the normalisation constraint $\int_{\mathcal{V}_\sigma} \mathrm{d}\sigma\ \rho(\mathbf{x}, \sigma) = 1$.

We are now ready to generalise this result to the case where a fluid particle at point $\mathbf{x}$ is carrying not only a buoyancy level $b(\mathbf{x}) = \sigma$ with $\sigma \in \mathcal{V}_\sigma$, but also a velocity (vector) level $\mathbf{u}(\mathbf{x}) = \mathbf{v}$ with $\mathbf{v} \in \mathcal{V}_\mathbf{v} = [-v_{max},\ v_{max}]^3$. The same steps leading to Eq. (B 6) can be applied to that case, which yields

$$S_B = -\frac{k_B MN}{V} \int_{\mathcal{V}_\mathbf{x}} \mathrm{d}\mathbf{x} \int_{\mathcal{V}_\sigma} \mathrm{d}\sigma \int_{\mathcal{V}_\mathbf{v}} \mathrm{d}\mathbf{v}\ \rho(\mathbf{x}, \sigma, \mathbf{v}) \log \rho(\mathbf{x}, \sigma, \mathbf{v}) \tag{B 7}$$

where $\rho(\mathbf{x}, \sigma, \mathbf{v})$ is the probability density function for the buoyancy and velocity at point $\mathbf{x}$. Note that we have introduced a cut-off denoted v_{max} for the maximum possible velocity. Anticipating that velocity fluctuations are bounded due to the energy constraint, we expect that the results will not depend on v_{max} if it is chosen much larger than the root mean square velocity of the equilibrium state, and we will consider in the remaining of this paper $v_{max} = +\infty$.

Finally, choosing $k_B = V/(NM)$ in Eq. (B 7), we recover $S_B = \mathcal{S}[\rho]$, where $\mathcal{S}[\rho]$ is the macrostate entropy defined in Eq. (2.13).

Appendix C. Relaxation equations from a maximum entropy production principle

The aim of this appendix is to provide an algorithm that makes possible numerical computations of the equilibrium states for arbitrary energy E and global distribution of buoyancy $G(\sigma)$. We consider for that purpose the ansatz (3.6) for the local distribution of velocity and buoyancy levels, and we propose in the following a dynamical system describing the temporal evolution of the quantities $\rho_b(\mathbf{x}, \sigma, t)$, $e_c(t)$ in such a way that the total energy and the global distribution of buoyancy levels are conserved, just as in the original Boussinesq system, and in such a way that the entropy production is

maximum at each time. This maximum entropy production principle ensures convergence towards an entropy maximum for a given set of constraints E, $G(\sigma)$. Since the effective temperature (i.e. the Lagrange parameter associated with the energy) is positive, the entropy maximum is unique for a given set of constraints, and the dynamical system will therefore relax towards the equilibrium state. We stress that considering the temporal evolution of this dynamical system is a trick to find the equilibrium state. The actual flow dynamics may follow a different path towards equilibrium than the one maximizing the entropy production.

Since the dynamical system is fully described by $\rho_b(\mathbf{x}, \sigma, t)$ and $e_c(t)$, it will be useful in the following to express the conservation of the global buoyancy distribution and of the total energy in terms of those parameters. Inserting Eq. (3.6) in (2.12) and (2.11) yields to

$$\mathcal{G}[\rho_b] = \frac{V}{2H} \int_{-H}^{H} \mathrm{d}z \ \rho_b \ . \tag{C 1}$$

$$\mathcal{E}[\rho_b](e_c) = V e_c + \frac{V}{2H} \int_{-H}^{H} \mathrm{d}z \int_{\mathcal{V}_\sigma} \mathrm{d}\sigma \ \rho_b \sigma z \ . \tag{C 2}$$

If the initial condition $\rho_b(z, \sigma, 0)$ and the initial kinetic energy $e_c(0)$ are known, then the global distribution of buoyancy levels and total energy can be computed using Eq. (C 1) and Eq. (C 2), respectively.

Assuming that there is no source nor sink of density, recalling that the flow is non-divergent, and anticipating that there is no mean flow, the temporal evolution of the pdf ρ_b satisfies the general conservation law

$$\partial_t \rho_b + \partial_z J_b = 0 \ , \tag{C 3}$$

where we have introduced the turbulent flux of probability $J_b(z, \sigma, t)$ directed along z, with $J_b = 0$ at the upper and the lower boundary $z = \pm H$.

The temporal evolution of the system requires a model for the flux J_b and the kinetic energy production $\dot{e}_c = \mathrm{d}e_c/\mathrm{d}t$. The maximum entropy production principle amounts to finding the flux J_b and the kinetic energy production $\dot{e}_c$ that maximise the entropy production while satisfying the constraints of the problem.

Let us first compute the entropy and energy production. Injecting the ansatz (3.6) in Eq. (2.13), the macrostate entropy can be expressed as

$$\mathcal{S} = -\frac{V}{2H} \int_{-H}^{+H} \mathrm{d}z \int_{\mathcal{V}_\sigma} \mathrm{d}\sigma \ \rho_b \log \rho_b - \frac{3}{2} V \log \frac{3}{2 e_c}. \tag{C 4}$$

Taking the time derivative of Eq. (C 4) and using Eq. (C 3), the entropy production can be expressed as

$$\dot{\mathcal{S}} = \frac{V}{2H} \int_{-H}^{+H} \mathrm{d}z \int_{\mathcal{V}_\sigma} \mathrm{d}\sigma \ J_b \partial_z (\log \rho_b) \ - \frac{3}{2} V \frac{\dot{e}_c}{e_c}. \tag{C 5}$$

The constraints of the problem are given by

- the conservation of the local normalisation (2.7), implying

$$\forall z \in [-H \ H], \ \int_{\mathcal{V}_\sigma} \mathrm{d}\sigma \ J_b = 0, \tag{C 6}$$

which ensures the local normalization $\int_{\mathcal{V}_\sigma} \mathrm{d}\sigma \rho_b = 1$,

- the energy conservation, which can be expressed as $\dot{\mathcal{E}} = 0$. Taking the temporal

derivative of Eq. (C 2) yields to

$$\dot{\mathcal{E}} = V\dot{e}_c + \frac{V}{2H}\int_{-H}^{+H} \mathrm{d}z \int_{\mathcal{V}_\sigma} \mathrm{d}\sigma\ \sigma J_b. \tag{C 7}$$

• Finally, the fluxes of probability must be finite to be dynamically relevant. Indeed, an infinite flux would corresponds to an instantaneous rearrangements of the buoyancy field. We impose therefore a bound for the norm of the probability flux J_b, expressed as:

$$\forall z \in [-H\ H], \quad \int_{\mathcal{V}_\sigma} \mathrm{d}\sigma\ \frac{J_b^2}{2\rho_b} \leq C(z)\ . \tag{C 8}$$

The quantity J_b/ρ_b can be interpreted as a diffusion velocity for the probability density field, and the constraint in Eq. (C 8) ensures that this velocity remains finite everywhere and for each buoyancy level during the relaxation process.

The variational problem of the maximum entropy production principle is treated by introducing Lagrange multipliers $\zeta(z)$, β_t and $-/D(z)$ associated with the constraints in Eqs. (C 6), (C 7) and (C 8), respectively. Note that following the Karush-Kuhn-Tucker conditions, an inequality such as (C 8) can be treated as an equality constraint when computing the first order variations in an optimisation problem (Sundaram 1996). The condition

$$\delta\dot{\mathcal{S}} - \beta_t\delta\dot{\mathcal{E}} + \int_{-H}^{+H} \mathrm{d}z\ \int_{\mathcal{V}_\sigma} \mathrm{d}\sigma\ \zeta(z)\delta J_b - \int_{-H}^{+H} \mathrm{d}z\ \int_{\mathcal{V}_\sigma} \mathrm{d}\sigma\ \frac{1}{D}\frac{J_b}{\rho_b}\delta J_b = 0\ , \tag{C 9}$$

must be satisfied for each $\delta\dot{e}_c$ and δJ_b. Using

$$\delta\dot{\mathcal{S}} = \frac{V}{2H}\int_{-H}^{+H} \mathrm{d}z \int_{\mathcal{V}_\sigma} \mathrm{d}\sigma\ \partial_z(\log\rho_b)\delta J_b - \frac{3V}{2e_c}\delta\dot{e}_c,\ \ \delta\dot{\mathcal{E}} = V\delta\dot{e}_c + \frac{V}{2H}\int_{-H}^{+H} \mathrm{d}z \int_{\mathcal{V}_\sigma} \mathrm{d}\sigma\ \sigma\delta J_b, \tag{C 10}$$

Eq. (C 9) yields

$$\beta_t = 3/(2e_c), \quad J_b = -D\left(\partial_z\rho_b - \beta_\mathrm{t}\left(\sigma - \overline{b}\right)\rho_b\right), \tag{C 11}$$

where $\zeta(z)$ has been determined by using the constraint in Eq. (C 6). In addition, the coefficient D must be positive for the entropy production to be positive. As far as the equilibrium state is concerned, the value of D in not important. Indeed, the flux J_b vanishes when equilibrium is reached, which ensures that the equilibrium state does not depend on D.

REFERENCES

Bartello, P. 1995 Geostrophic adjustment and inverse cascades in rotating stratified turbulence. *Journal of the atmospheric sciences* **52** (24), 4410–4428.

Boucher, C., Ellis, R. S. & Turkington, B. 2000 Derivation of maximum entropy principles in two-dimensional turbulence via large deviations. *Journal of Statistical Physics* **98** (5-6), 1235–1278.

Bouchet, F. & Corvellec, M. 2010 Invariant measures of the 2d euler and vlasov equations. *Journal of Statistical Mechanics: Theory and Experiment* **2010** (08), P08021.

Bouchet, F. & Sommeria, J. 2002 Emergence of intense jets and jupiter's great red spot as maximum-entropy structures. *Journal of Fluid Mechanics* **464**, 165–207.

Bouchet, F. & Venaille, A. 2012 Statistical mechanics of two-dimensional and geophysical flows. *Physics reports* **515** (5), 227–295.

Bouffard, D. & Boegman, L. 2013 A diapycnal diffusivity model for stratified environmental flows. *Dynamics of Atmospheres and Oceans* **61**, 14–34.

Brethouwer, G., Billant, P., Lindborg, E. & Chomaz, J.-M. 2007 Scaling analysis and simulation of strongly stratified turbulent flows. *Journal of Fluid Mechanics* **585**, 343–368.

Caulfield, C.P. & Kerswell, R.R. 2001 Maximal mixing rate in turbulent stably stratified couette flow. *Physics of Fluids (1994-present)* **13** (4), 894–900.

Chavanis, P.-H. 2002 Statistical mechanics of two-dimensional vortices and stellar systems. In *Dynamics and Thermodynamics of Systems with Long-range Interactions*, pp. 208–289. Springer.

Dalziel, S. B., Patterson, M. D., Caulfield, C.P. & Coomaraswamy, I. A. 2008 Mixing efficiency in high-aspect-ratio rayleigh–taylor experiments. *Physics of Fluids* **20** (6), 065106.

Eyink, G.L. & Sreenivasan, K.R. 2006 Onsager and the theory of hydrodynamic turbulence. *Reviews of modern physics* **78** (1), 87.

Fernando, H.J.S. 1991 Turbulent mixing in stratified fluids. *Annual Review of Fluid Mechanics* **23**, 455–493.

Gill, A.E. 1982 *Atmosphere-ocean dynamics*, , vol. 30. Academic press.

Godeferd, F. S. & Cambon, C. 1994 Detailed investigation of energy transfers in homogeneous stratified turbulence*. *Physics of Fluids (1994-present)* **6** (6), 2084–2100.

Herbert, C., Pouquet, A. & Marino, R. 2014 Restricted equilibrium and the energy cascade in rotating and stratified flows. *Journal of Fluid Mechanics* **758**, 374–406.

Holford, J.M. & Linden, P.F. 1999 Turbulent mixing in a stratified fluid. *Dynamics of atmospheres and oceans* **30** (2), 173–198.

Hopfinger, E.J. 1987 Turbulence in stratified fluids: A review. *Journal of Geophysical Research: Oceans* **92** (C5), 5287–5303.

Huq, P. & Britter, R.E. 1995 Turbulence evolution and mixing in a two-layer stably stratified fluid. *Journal of Fluid Mechanics* **285**, 41–68.

Ilicak, Mehmet 2014 Energetics and mixing efficiency of lock-exchange flow. *Ocean Modelling* **83**, 1–10.

Ivey, G.N., Winters, K.B. & Koseff, J.R. 2008 Density stratification, turbulence, but how much mixing? *Annual Review of Fluid Mechanics* **40** (1), 169.

Karimpour, F. & Venayagamoorthy, S.K. 2014 A simple turbulence model for stably stratified wall-bounded flows. *Journal of Geophysical Research: Oceans* **119** (2), 870–880.

Kerstein, A.R. 1999 One-dimensional turbulence: model formulation and application to homogeneous turbulence, shear flows, and buoyant stratified flows. *Journal of Fluid Mechanics* **392**, 277–334.

Kraichnan, R. H. 1967 Inertial Ranges in Two-Dimensional Turbulence. *Physics of Fluids* **10**, 1417–1423.

Large, W. G., McWilliams, J. C. & Doney, S. C. 1994 Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization. *Reviews of Geophysics* **32** (4), 363–404.

Lee, T.D. 1952 On some statistical properties of hydrodynamical and magnetohydrodynamical fields. *Q. Appl. Math* **10** (1), 69–74.

Lindborg, E. 2005 The effect of rotation on the mesoscale energy cascade in the free atmosphere. *Geophysical research letters* **32** (1).

Lindborg, E. 2006 The energy cascade in a strongly stratified fluid. *Journal of Fluid Mechanics* **550**, 207–242.

Linden, P.F. 1979 Mixing in stratified fluids. *Geophysical & Astrophysical Fluid Dynamics* **13** (1), 3–23.

Linden, P.F. 1980 Mixing across a density interface produced by grid turbulence. *Journal of Fluid Mechanics* **100** (04), 691–703.

Lozovatsky, I.D. & Fernando, H.J.S. 2013 Mixing efficiency in natural flows. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **371** (1982), 20120213.

Lynden-Bell, D. 1967 Statistical mechanics of violent relaxation in stellar systems. *Monthly Notices of the Royal Astronomical Society* **136**, 101.

Maffioli, A., Brethouwer, G. & Lindborg, E. 2016 Mixing efficiency in stratified turbulence. *Journal of Fluid Mechanics* **794**, R3.

Majda, A. & Wang, X. 2006 *Nonlinear dynamics and statistical theories for basic geophysical flows*. Cambridge University Press.

McEwan, A.D. 1983*a* Internal mixing in stratified fluids. *Journal of Fluid Mechanics* **128**, 59–80.

McEwan, A.D. 1983*b* The kinematics of stratified mixing through internal wavebreaking. *Journal of Fluid Mechanics* **128**, 47–57.

Mellor, G. L. & Yamada, T. 1982 Development of a turbulence closure model for geophysical fluid problems. *Reviews of Geophysics* **20** (4), 851–875.

Merryfield, W. J. 1998 Effects of stratification on quasi-geostrophic inviscid equilibria. *Journal of Fluid Mechanics* **354**, 345–356.

Michel, J. & Robert, R. 1994 Large deviations for young measures and statistical mechanics of infinite dimensional dynamical systems with conservation law. *Communications in Mathematical Physics* **159** (1), 195–215.

Miller, J. 1990 Statistical mechanics of euler equations in two-dimensions. *Physical review letters* **65** (17), 2137.

Nakanish, M. 2001 Improvement of the mellor–yamada turbulence closure model based on large-eddy simulation data. *Boundary-layer meteorology* **99** (3), 349–378.

Naso, A., Monchaux, R., Chavanis, P.-H. & Dubrulle, B. 2010 Statistical mechanics of beltrami flows in axisymmetric geometry: Theory reexamined. *Physical Review E* **81** (6), 066318.

Onsager, L. 1949 Statistical hydrodynamics. *Il Nuovo Cimento (1943-1954)* **6**, 279–287.

Osborn, T.R. 1980 Estimates of the local rate of vertical diffusion from dissipation measurements. *Journal of Physical Oceanography* **10** (1), 83–89.

Osborn, T.R. & Cox, C.S. 1972 Oceanic fine structure. *Geophysical & Astrophysical Fluid Dynamics* **3** (1), 321–345.

Pardyjak, E.R., Monti, P. & Fernando, H.J.S. 2002 Flux richardson number measurements in stable atmospheric shear flows. *Journal of Fluid Mechanics* **459**, 307–316.

Peltier, W.R. & Caulfield, C.P. 2003 Mixing efficiency in stratified shear flows. *Annual review of fluid mechanics* **35** (1), 135–167.

Pomeau, Y. 1994 Statistical approach (to 2D turbulence). In *O. Cardoso, P. Tabeling (Eds) Turbulence: A tentative dictionary* (ed. Plenum Press), pp. 385–447.

Pope, S.B. 1985 Pdf methods for turbulent reactive flows. *Progress in Energy and Combustion Science* **11** (2), 119–192.

Potters, M., Vaillant, T. & Bouchet, F. 2013 Sampling microcanonical measures of the 2d euler equations through creutz's algorithm: a phase transition from disorder to order when energy is increased. *Journal of Statistical Mechanics: Theory and Experiment* **2013** (02), P02017.

Prastowo, T., Griffiths, R. W., Hughes, G.O. & Hogg, A.M. 2008 Mixing efficiency in controlled exchange flows. *Journal of Fluid Mechanics* **600**, 235–244.

Prieto, R. & Schubert, W. H. 2001 Analytical predictions for zonally symmetric equilibrium states of the stratospheric polar vortex. *Journal of the atmospheric sciences* **58** (18), 2709–2728.

Rehmann, C. R. & Koseff, J.R. 2004 Mean potential energy change in stratified grid turbulence. *Dynamics of atmospheres and oceans* **37** (4), 271–294.

Renaud, A., Venaille, A. & Bouchet, F. 2016 Equilibrium statistical mechanics and energy partition for the shallow water model. *Journal of Statistical Physics* **163** (4), 784–843.

Robert, R. & Sommeria, J. 1991 Statistical equilibrium states for two-dimensional flows. *Journal of Fluid Mechanics* **229**, 291–310.

Robert, R. & Sommeria, J. 1992 Relaxation towards a statistical equilibrium state in two-dimensional perfect fluid dynamics. *Physical Review Letters* **69** (19), 2776.

Rorai, C., Mininni, P.D. & Pouquet, A. 2014 Turbulence comes in bursts in stably stratified flows. *Physical Review E* **89** (4), 043002.

Salehipour, H. & Peltier, W.R. 2015 Diapycnal diffusivity, turbulent prandtl number and mixing efficiency in boussinesq stratified turbulence. *Journal of Fluid Mechanics* **775**, 464–500.

Salehipour, H., Peltier, W.R., Whalen, C.B. & MacKinnon, J.A. 2016 A new characterization of the turbulent diapycnal diffusivities of mass and momentum in the ocean. *Geophysical Research Letters* **43** (7), 3370–3379.

Salmon, R. 1998 *Lectures on geophysical fluid dynamics*, , vol. 378. Oxford University Press Oxford.

Salmon, R. 2012 Statistical mechanics and ocean circulation. *Communications in Nonlinear Science and Numerical Simulation* **17** (5), 2144–2152.

Schecter, D.A. 2003 Maximum entropy theory and the rapid relaxation of three-dimensional quasi-geostrophic turbulence. *Physical Review E* **68** (6), 066309.

Shepherd, T. G. 1993 A unified theory of available potential energy 1. *Atmosphere-Ocean* **31** (1), 1–26.

Shih, L. H., Koseff, J. R., Ivey, G. N. & Ferziger, J. H. 2005 Parameterization of turbulent fluxes and scales using homogeneous sheared stably stratified turbulence simulations. *Journal of Fluid Mechanics* **525**, 193–214.

Sommeria, J. 2001 Two-dimensional turbulence. In *New trends in turbulence Turbulence: nouveaux aspects*, pp. 385–447. Springer.

Staquet, C. & Sommeria, J. 2002 Internal gravity waves: from instabilities to turbulence. *Annual Review of Fluid Mechanics* **34** (1), 559–593.

Stretch, D.D., Rottman, J.W., Venayagamoorthy, S. K., Nomura, K.K. & Rehmann, C.R. 2010 Mixing efficiency in decaying stably stratified turbulence. *Dynamics of atmospheres and oceans* **49** (1), 25–36.

Sundaram, R. K. 1996 *A first course in optimization theory*. Cambridge university press.

Tabak, E.G. & Tal, F.A. 2004 Mixing in simple models for turbulent diffusion. *Communications on pure and applied mathematics* **57** (5), 563–589.

Tabeling, P. 2002 Two-dimensional turbulence: a physicist approach. *Physics Reports* **362** (1), 1–62.

Tailleux, R 2009 Understanding mixing efficiency in the oceans: do the nonlinearities of the equation of state for seawater matter? *Ocean Science (OS)* .

Thalabard, S., Dubrulle, B. & Bouchet, F. 2014 Statistical mechanics of the 3d axisymmetric euler equations in a taylor–couette geometry. *Journal of Statistical Mechanics: Theory and Experiment* **2014**, P01005.

Thalabard, S., Saint-Michel, B., Herbert, É., Daviaud, F. & Dubrulle, B. 2015 A statistical mechanics framework for the large-scale structure of turbulent von kármán flows. *New Journal of Physics* **17** (6), 063006.

Thorpe, S.A. 2005 *The turbulent ocean*. Cambridge University Press.

Touchette, H. 2009 The large deviation approach to statistical mechanics. *Physics Reports* **478** (1), 1–69.

Townsend, AA 1958 The effects of radiative transfer on turbulent flow of a stratified fluid. *Journal of Fluid Mechanics* **4** (04), 361–375.

Turkington, B., Majda, A., Haven, K. & DiBattista, M. 2001 Statistical equilibrium predictions of jets and spots on jupiter. *Proceedings of the National Academy of Sciences* **98** (22), 12346–12350.

Vallis, G.K. 2006 *Atmospheric and oceanic fluid dynamics: fundamentals and large-scale circulation*. Cambridge University Press.

Vassilicos, J.C. 2015 Dissipation in turbulent flows. *Annual Review of Fluid Mechanics* **47**, 95–114.

1177 Venaille, A. 2012 Bottom-trapped currents as statistical equilibrium states above topographic
1178 anomalies. *Journal of Fluid Mechanics* **699**, 500–510.
1179 Venaille, A. & Bouchet, F. 2011 Oceanic rings and jets as statistical equilibrium states.
1180 *Journal of Physical Oceanography* **41**, 1860–1873.
1181 Venaille, A., Dauxois, T. & Ruffo, S. 2015 Violent relaxation in two-dimensional flows
1182 with varying interaction range. *Physical Review E* **92** (1), 011001.
1183 Venaille, A. & Sommeria, J. 2010 Modeling mixing in two-dimensional turbulence and
1184 stratified fluids. In *Proceedings of the IUTAM Symposium on Turbulence in the Atmosphere*
1185 *and Oceans*, , vol. 28, p. 155. Springer.
1186 Venaille, A., Vallis, G.K. & Griffies, S.M. 2012 The catalytic role of the beta effect in
1187 barotropization processes. *Journal of Fluid Mechanics* **709**, 490–515.
1188 Venayagamoorthy, S. K. & Koseff, J. R. 2016 On the flux richardson number in stably
1189 stratified turbulence. *Journal of Fluid Mechanics* **798**, R1 (10 pages).
1190 Waite, M. L. & Bartello, P. 2004 Stratified turbulence dominated by vortical motion.
1191 *Journal of Fluid Mechanics* **517**, 281–308.
1192 Warn, T. 1986 Statistical mechanical equilibria of the shallow water equations. *Tellus A* **38** (1).
1193 Weichman, P. B. 2006 Equilibrium theory of coherent vortex and zonal jet formation in a
1194 system of nonlinear rossby waves. *Physical Review E* **73** (3), 036313.
1195 Weichman, P. B. & Petrich, D. M. 2001 Statistical Equilibrium Solutions of the Shallow
1196 Water Equations. *Physical Review Letters* **86**, 1761–1764.
1197 Winters, K.B., Lombard, P.N., Riley, J.J. & D'Asaro, E.A. 1995 Available potential
1198 energy and mixing in density-stratified fluids. *Journal of Fluid Mechanics* **289**, 115–128.
1199 Wunsch, C. & Ferrari, R. 2004 Vertical mixing, energy, and the general circulation of the
1200 oceans. *Annu. Rev. Fluid Mech.* **36**, 281–314.
1201 Wykes, M. S. D. & Dalziel, S. B. 2014 Efficient mixing in stratified flows: experimental study
1202 of a rayleigh–taylor unstable interface within an otherwise stable stratification. *Journal of*
1203 *Fluid Mechanics* **756**, 1027–1057.
1204 Wykes, M. S. D., Hughes, G. O. & Dalziel, S.B. 2015 On the meaning of mixing efficiency
1205 for buoyancy-driven mixing in stratified turbulent flows. *Journal of Fluid Mechanics* **781**,
1206 261–275.