J. Plasma Phys. (2025), *vol*. 91, E35
Published by Cambridge University Press

doi:10.1017/S0022377824001521

Metastability of stratified magnetohydrostatic equilibria and their relaxation

D.N. Hosking[1,2,†], **D. Wasserman**[3] **and S.C. Cowley**[4]

[1]Princeton Center for Theoretical Science, Princeton, NJ 08540, USA

[2]Gonville & Caius College, Trinity Street, Cambridge CB2 1TA, UK

[3]Northeastern University, Boston, MA 02115, USA

[4]Princeton Plasma Physics Laboratory, Princeton, NJ 08540, USA

(Received 1 June 2024; revised 6 November 2024; accepted 8 November 2024)

Motivated by explosive releases of energy in fusion, space and astrophysical plasmas, we consider the nonlinear stability of stratified magnetohydrodynamic equilibria against two-dimensional interchanges of straight magnetic-flux tubes. We demonstrate that, even within this restricted class of dynamics, the linear stability of an equilibrium does not guarantee its nonlinear stability: equilibria can be metastable. We show that the minimum-energy state accessible to a metastable equilibrium under non-diffusive two-dimensional dynamics can be found by solving a combinatorial optimisation problem. These minimum-energy states are, to good approximation, the final states reached by our simulations of destabilised metastable equilibria for which turbulent mixing is suppressed by viscosity. To predict the result of fully turbulent relaxation, we construct a statistical mechanical theory based on the maximisation of Boltzmann's mixing entropy. This theory is analogous to the Lynden-Bell statistical mechanics of collisionless stellar systems and plasma, and to the Robert–Sommeria–Miller theory of two-dimensional vortex turbulence. Our theory reproduces well the results of our numerical simulations for sufficiently large perturbations to the metastable equilibrium.

Keywords: plasma nonlinear phenomena, plasma instabilities

1. Introduction

Both in the laboratory and in nature, plasmas that host strong magnetic fields sometimes exist in slowly evolving quasi-equilibrium states. Such plasmas may have a dynamical time scale $\tau_A \sim L/v_A$ (L is the system's length scale and v_A the Alfvén speed) that is much smaller than the time scales associated with energy injection and transport. In the solar corona, for example, the footpoints of magnetic-flux loops evolve on the photospheric driving time scale of $\sim$10 min, while $\tau_A \sim 10$ s (Cranmer & Winebarger 2019). Likewise, the transport time scale for magnetic-confinement-fusion devices is typically ~ 0.1 s, much larger than $\tau_A \sim 10^{-6}$ s (ITER Physics Basis 1999).

† Email address for correspondence: dhosking@princeton.edu

Occasionally, these plasmas depart suddenly and violently from their quasi-equilibria. Explosive releases of energy – eruptions – involving substantial reconfiguration of the magnetic field happen both in the corona (coronal mass ejections; see, e.g. Chen 2011 for a review) and fusion experiments (disruption events; see, e.g. Hender *et al.* 2007). Evidently, the quasi-equilibria can become unstable during their evolution. When eruptions occur, the system relaxes towards a new state that is a lower minimum of the potential energy in configuration space. For a significant amount of potential energy to be liberated, the new minimum must be distant from the original one. The instability must therefore be nonlinear: eruptions happen when metastable states are destabilised.

Ideal magnetohydrodynamic (MHD) instabilities may be categorised into two types: kink and interchange (or ballooning) instabilities. Kink instabilities are global, occurring at the system scale, and are characterised by significant variation along magnetic-field lines ($k_\parallel \sim k_\perp$, where $k_\parallel$ and $k_\perp$ are characteristic wavenumbers along and across the magnetic field, respectively). In contrast, interchange instabilities (Connor, Hastie & Taylor 1979) are local and scale-independent, with elongation along magnetic-field lines ($k_\parallel \ll k_\perp$). In most known cases, a supercritical bifurcation occurs when an MHD equilibrium crosses the linear threshold for kink instability; two new stable equilibria that are nearby in configuration space become realisable (Friedrichs 1960; Rutherford, Furth & Rosenbluth 1971; White *et al.* 1977; Lorenzini *et al.* 2009). Because these equilibria are nearby, no significant release of potential energy is possible if the system is pushed out of one and into the other. On the other hand, subcritical bifurcation is possible at the linear threshold for interchange instability. In this case, the system is nonlinearly unstable at marginal linear stability. Any stable equilibrium to which the system can relax is distant in configuration space, meaning that a finite amount of potential energy can be liberated.

Previous studies have elucidated certain properties of the relaxation of MHD equilibria from states that are metastable to interchange-type dynamics. Cowley & Artun (1997) studied the case of a stratified equilibrium with initially horizontal magnetic field embedded in conducting walls (fixed field-line endpoints). They showed that gradients of thermal or magnetic pressure (balanced in equilibrium by gravity) can provide sources of free energy for a buoyancy instability that is stabilised by magnetic tension only for linear perturbations – not nonlinear ones. By solving the weakly nonlinear equations of motion numerically, Cowley & Artun (1997) observed a phenomenon that they termed detonation: progressive destabilisation of the metastable equilibrium by erupting finger-like magnetic structures. These results were extended to more general geometry by Hurricane, Fong & Cowley (1997), Wilson & Cowley (2004) and Ham *et al.* (2018). Cowley *et al.* (2015) showed that, with fast thermal conduction along field lines, erupting flux tubes have two possible fates: either they find a new equilibrium position or they reach a singular state with zero magnetic-field strength (flux expulsion).

Despite the successes of these studies, certain fundamental questions remain difficult to answer accurately because of the geometrical complexity that arises from the bending of magnetic field lines. Such questions include: To what state does a metastable equilibrium relax when it is destabilised? What fraction of its energy is available for liberation? Is that energy always liberated in practice, i.e. is relaxation complete?

The aim of this paper is to consider metastability in a simpler setting where these questions can be answered. A key result (which, to our knowledge, has not been recognised previously) is that bending of magnetic field lines, while certainly a feature of eruptions in the corona and fusion devices, is not a requirement for the existence of nonlinear instability. We demonstrate this fact in figures 1 and 2, which visualise numerical simulations that are two-dimensional (2D) with out-of-plane magnetic field only. In

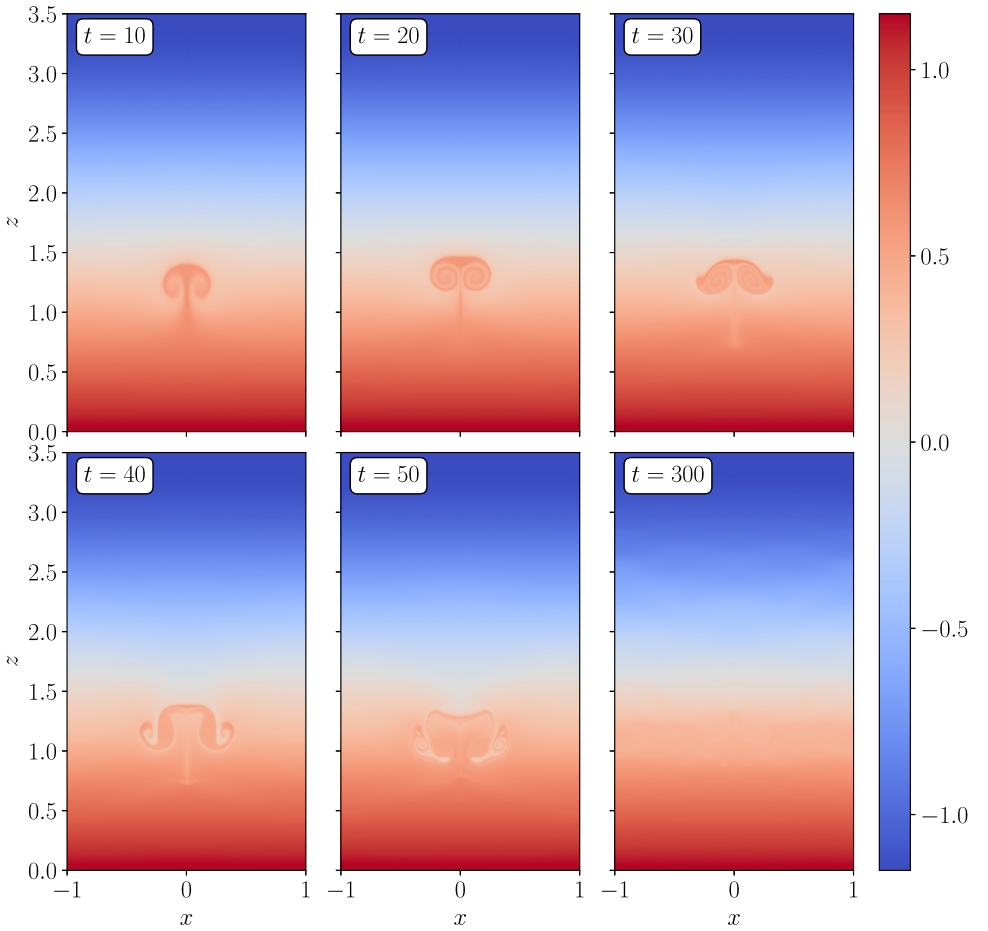

FIGURE 1. Two-dimensional simulation of a MHD atmosphere subjected to an impulse that does not trigger instability. The atmosphere relaxes to a final state that approximates the initial condition. The quantity visualised is the natural logarithm of the ratio of the entropy function (2.20) to the specific magnetic flux (2.19). This quantity is conserved in a Lagrangian sense in the absence of diffusion and controls the compressibility of the fluid, with larger values being more compressible (see § 2.5). The initial velocity field is $\boldsymbol{u} = u_0\hat{\boldsymbol{z}}\exp(-[x^2 + (z-1.0)^2]/0.1^2)$ with $u_0 = 0.2$. The equilibrium is defined by (2.30) with $\epsilon_0 = 10^{-2}$ in (2.34). The co-ordinates x and z are measured in units of the total-pressure scale height at $z = 0$ and the time t is measured in units of the sound-crossing time of the total-pressure scale height at $z = 0$ (see § 2.6 for details). A movie version of this figure is available at https://doi.org/10.1017/S0022377824001521.

figure 1, a small perturbation to a stratified equilibrium or 'atmosphere' (gravity is in the negative-z direction) does not lead to destabilisation: the original state is restored. In figure 2, a larger perturbation destabilises the same equilibrium: a rising plume of material does not return to its original position, but establishes a new equilibrium state (see figure captions for further information about the physical set-up and Appendix A for details of numerical methods).

In the 2D context (upon which we focus exclusively in this paper) metastability is enabled by the fact that less magnetised fluid (i.e. fluid with a larger ratio β of the thermal to magnetic pressures) is more compressible than fluid that is more magnetised

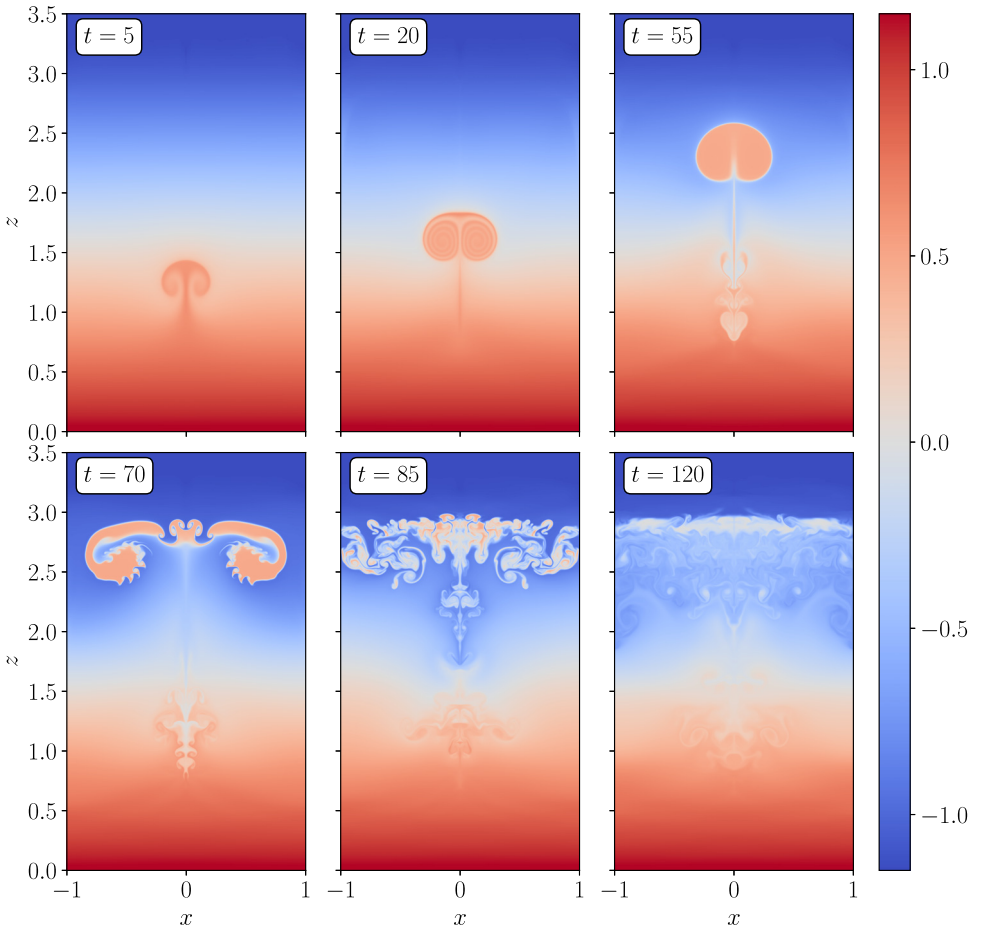

FIGURE 2. As figure 1, but for an initial velocity field that is twice as large ($u_0 = 0.4$). The equilibrium is destabilised. A movie version of this figure is available at https://doi.org/10.1017/S0022377824001521.

(smaller β). A flux tube may therefore experience a greater change in density in response to a large (nonlinear) displacement than the background density profile does over the same distance, provided its β is larger than that of the surrounding flux tubes. This produces a destabilising buoyancy force, even though the equilibrium may be stable linearly. In fact, a precisely analogous phenomenon occurs in the Earth's atmosphere. This is because air that is saturated (i.e. sufficiently cool for water vapour to condense) is more compressible than unsaturated air. As a result, the atmosphere can be linearly stable to convection but unstable nonlinearly to displacement of saturated air through the so-called level of free convection. The resulting updraughts lead to the formation of cumulonimbus clouds and, as a result, thunderstorms (see, e.g. Rogers & Yau 1996). We present a general theoretical treatment of metastability due to composition-dependent compressibility in § 2 and review its application to the terrestrial atmosphere in Appendix B.

The rest of our paper is concerned with the question of how metastable MHD equilibria relax when they are destabilised. In § 3, we show that the available energy under nonlinear interchanges of flux tubes may be determined accurately by solving a combinatorial optimisation problem (linear sum assignment) – although we believe that we are the

first to use this approach for a magnetised system, we note that a similar one has been employed recently by Hieronymus & Nycander (2015), Su & Ingersoll (2016) and Stansifer, O'Gorman & Holt (2017) to calculate available potential energy in atmospheric and oceanic contexts. We show in § 3.3 that the states that are global minima of the potential energy typically turn out to have horizontal structure over a finite range of the vertical coordinate z, even if the initial configurations are one-dimensional in z. This appears to be a new observation. We further find that the amount of energy that can be liberated from a metastable MHD equilibrium in 2D is always a small fraction (at most a few per cent) of its total potential energy (§ 3.1). The reason is that flux tubes exclude each other: when a flux tube moves to the top or bottom of an atmosphere, it prevents others from doing the same.

The smallness of the available potential energy means that the kinetic energy that develops during relaxation is small compared with the internal energy of the fluid. Hence, relaxation is subsonic (for the vast majority of fluid). This has two important consequences. First, the decay of kinetic energy due to viscosity results in negligible heating, so that the thermal entropy of the fluid is unchanged during relaxation, provided that thermal and magnetic diffusion can be neglected. A corollary is that, if viscosity is sufficiently large to suppress turbulent mixing, i.e. the Reynolds number is sufficiently small (we estimate precisely how small in § 4.1), then relaxation terminates in a state close to the one with minimum energy with respect to ideal rearrangements. We confirm this expectation with direct numerical simulations in § 4.2.

A second consequence of relaxation being subsonic is that fluid parcels maintain local pressure balance during relaxation: the fractional variation in pressure at fixed height is small compared with that of density. We utilise this fact in § 5 to construct a statistical mechanical theory of turbulent relaxation at large Reynolds number. We conjecture that the state that develops as a result of turbulent mixing is the 'most mixed' (in the sense of maximising the Boltzmann mixing entropy with respect to non-diffusive rearrangements of flux tubes) locally pressure-balanced state with the given total energy (§ 5.2). This idea is analogous to the Lynden-Bell (1967) statistical mechanics of collisionless stellar systems and plasma, and to the Robert–Sommeria–Miller (RSM) theory of 2D vortex turbulence (Miller 1990; Robert 1991; Robert & Sommeria 1991; Miller, Weichman & Cross 1992), which has found widespread application in geophysical fluid dynamics (see Singh & O'Neill 2022 for a recent review). In our theory, the specific entropy s and magnetic flux χ, each of which is conserved in a fluid-element-wise sense under a non-dissipative dynamics, play the role of the phase-space density or vorticity in the Lynden-Bell or RSM theories, respectively.

Of course, neglect of diffusion is valid only until such time as the turbulent mixing develops sufficiently fine-scale structure in s and χ. In the Lynden-Bell (RSM) theory, diffusion due to collisions (viscosity) is straightforwardly modelled: when diffusion acts, the statistical mechanical probability distribution function for phase-space density (vorticity) collapses onto its expectation value. Our case is different: magnetic diffusion and thermal conduction increase the total thermal entropy (by a much larger amount than viscous heating does). Thus, we extract predictions for diffused states by assuming that our probability distribution function collapses not onto its expectation value but onto the value consistent with flux and energy conservation (§ 5.3). As it turns out, the resulting 1D equilibrium state may itself be unstable, linearly or nonlinearly, because diffusion alters buoyancy.[1] We consider taking the predicted unstable diffused state to be the starting point

[1] A well-known example is the phenomenon of 'buoyancy reversal' in the terrestrial atmosphere: when buoyantly rising dry air mixes with moist air, it may, after diffusion, become more dense than the unmixed moist air and sink again as a result (see, e.g. Stevens 2005).

for a new relaxation (if necessary, this procedure can be repeated until a stable diffused state is reached, although we find that one iteration is often enough to do so). The profiles of s and χ produced by this procedure are typically only slightly modified from the ones derived from the statistical mechanics in the first instance – qualitatively, the subsequent rearrangements and accompanied diffusion tend to produce local plateaus in the profiles of s and χ.

In § 6, we compare our theoretical predictions with the results of direct numerical simulations of relaxing metastable equilibria at large Reynolds number. The agreement turns out to be reasonably good, although less so for small initial perturbations. The reason for this appears to be that the detonation is incomplete in such cases: the system becomes trapped in a new metastable state and does not mix thoroughly. We conclude with a discussion of the possible implications and applications of our study in § 7.

2. Theory of convective metastability

2.1. *Definitions*

In this paper, we shall be concerned with a fluid dynamics defined by the momentum equation

$$\rho\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t} = -\boldsymbol{\nabla}P - \rho g\hat{\boldsymbol{z}}, \tag{2.1}$$

where ρ is density, $\boldsymbol{u}$ fluid velocity, t time, P total pressure, g the constant gravitational acceleration and $\mathrm{d}/\mathrm{d}t \equiv \partial/\partial t + \boldsymbol{u}\cdot\boldsymbol{\nabla}$ the material derivative. Density evolves according to the continuity equation

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = -\rho\boldsymbol{\nabla}\cdot\boldsymbol{u}, \tag{2.2}$$

and is related to P by the equation of state

$$\rho = \rho(P, \boldsymbol{Q}) \implies P = P(\rho, \boldsymbol{Q}). \tag{2.3}$$

The vector $\boldsymbol{Q}$ encodes the conserved material properties of the fluid (i.e. its Lagrangian invariants), which we allow to vary spatially. We shall primarily be concerned with the case of 2D MHD with out-of-plane magnetic field, for which the relevant components of $\boldsymbol{Q}$ are the specific entropy s and specific magnetic flux $\chi \equiv B/\rho$, where B is the magnetic-field strength; we shall specialise to this case in § 2.5. In different contexts, the components of $\boldsymbol{Q}$ might instead (or additionally) include mixing ratios (e.g. salinity or specific humidity) or the entropies of a coupled radiation field or cosmic rays. Over sufficiently small time scales and at sufficiently large spatial scales that diffusion can be neglected, $\boldsymbol{Q}$ is conserved following fluid particles, i.e.

$$\frac{\mathrm{d}\boldsymbol{Q}}{\mathrm{d}t} = 0. \tag{2.4}$$

By neglecting diffusion, we exclude double-diffusive buoyancy instabilities (see, e.g. Garaud 2018, Hughes & Brummell 2021 and references therein) from our analysis. We likewise exclude instabilities relating to anisotropic thermal conductivity (Balbus 2000; Quataert 2008). The application of our methods to the saturation of such instabilities is a topic to which we plan to return in future work (see the discussion in § 7.2).

2.2. *Stability analysis*

We now consider the stability under (2.1)–(2.4) of a 1D static equilibrium state, i.e. one for which $\boldsymbol{u} = 0$ and P, ρ and $\boldsymbol{Q}$ depend on z only, with $\mathrm{d}P/\mathrm{d}z = -\rho g$. The net force $\boldsymbol{F}$

on a small parcel of fluid moved in pressure balance with its surroundings and without diffusion (i.e. satisfying (2.4)) from initial height z_1 to new height z_2 is

$$\boldsymbol{F} = -gV_2\hat{\boldsymbol{z}}[\rho(P_2, \boldsymbol{Q}_1) - \rho(P_2, \boldsymbol{Q}_2)], \tag{2.5}$$

where V_2 is the volume of the parcel at z_2 and $P_2 = P(z_2)$, etc. Writing the difference in densities as an integral in z, we obtain

$$\boldsymbol{F} = gV_2\hat{\boldsymbol{z}}\int_{z_1}^{z_2} \mathrm{d}z \frac{\mathrm{d}\boldsymbol{Q}}{\mathrm{d}z} \cdot \frac{\partial\rho(P_2, \boldsymbol{Q})}{\partial\boldsymbol{Q}}. \tag{2.6}$$

2.2.1. *Linear stability*

The criterion for linear stability follows from taking $\delta z \equiv z_2 - z_1 \to 0$ in (2.6)

$$\mathcal{L} \equiv -\frac{\mathrm{d}\boldsymbol{Q}}{\mathrm{d}z} \cdot \frac{\partial \ln\rho(P, \boldsymbol{Q})}{\partial\boldsymbol{Q}} > 0, \quad \forall z. \tag{2.7}$$

The function $\mathcal{L}$ may alternatively be written as

$$\mathcal{L} = -\frac{\mathrm{d}\ln\rho}{\mathrm{d}z} + \frac{\partial \ln\rho(P, \boldsymbol{Q})}{\partial \ln P}\frac{\mathrm{d}\ln P}{\mathrm{d}z}, \tag{2.8}$$

from which $\mathcal{L} > 0$ is readily interpreted as the condition for the density of a fluid parcel displaced upwards (downwards) infinitesimally while conserving $\boldsymbol{Q}$ to be greater (less) than the density of the background fluid in the new position of the parcel.

2.2.2. *Nonlinear stability*

We can use (2.6) to write the criterion for nonlinear stability as

$$-\delta z\int_{z_1}^{z_2} \mathrm{d}z \frac{\mathrm{d}\boldsymbol{Q}}{\mathrm{d}z} \cdot \frac{\partial\rho(P_2, \boldsymbol{Q})}{\partial\boldsymbol{Q}} > 0, \quad \forall\, z_1, z_2. \tag{2.9}$$

If (2.9) is satisfied, then the buoyancy force is always in the opposite direction to δz. In many cases of interest (including the examples listed above), the components Q_i of $\boldsymbol{Q}$ can each be chosen such that an increase in Q_i at fixed pressure causes expansion

$$\frac{\partial \ln\rho(P, \boldsymbol{Q})}{\partial Q_i} < 0, \quad \forall\, i, P. \tag{2.10}$$

In the case of $Q_i = s$, for example, (2.10) holds for any fluid that expands under heating at fixed pressure. We shall assume in what follows that (2.10) holds. In that case, a sufficient condition for (2.9) to hold is

$$\frac{\mathrm{d}Q_i}{\mathrm{d}z} > 0, \quad \forall\, i, z. \tag{2.11}$$

Importantly, (2.7) and (2.11) are equivalent if the vector $\boldsymbol{Q}$ has only one component (and (2.10) holds). A fluid of this kind cannot exist in a metastable equilibrium, because it is always nonlinearly stable if linearly stable. The archetypical example is ordinary hydrodynamics, for which the only component of $\boldsymbol{Q}$ is s, and therefore an equilibrium is nonlinearly stable to convection if it satisfies the Schwarzschild criterion (Schwarzschild 1906)

$$\frac{\mathrm{d}s}{\mathrm{d}z} > 0, \quad \forall\, z. \tag{2.12}$$

2.2.3. *Metastability*

For fluids for which $\boldsymbol{Q}$ has more than one component, (2.11) guarantees nonlinear stability, but is not necessary for the weaker condition of linear stability: if $\mathrm{d}Q_i/\mathrm{d}z < 0$ for some i, other components of $\boldsymbol{Q}$ can compensate so that the linear-stability criterion (2.7) remains satisfied. In such cases, the equilibrium may be nonlinearly unstable despite being linearly stable, i.e. it may be metastable. In such cases, the condition for metastability may be deduced by recasting (2.5) as the upwards force per unit mass of moved fluid, $\boldsymbol{F}\cdot\hat{\boldsymbol{z}}/\rho(P_2,\boldsymbol{Q}_1)V_2 = (\mathrm{e}^{\mathcal{R}}-1)g$, where

$$\mathcal{R} \equiv \ln\frac{\rho(P_2,\boldsymbol{Q}_2)}{\rho(P_2,\boldsymbol{Q}_1)} = -\int_{z_1}^{z_2}\mathrm{d}z\mathcal{L} + \int_{z_1}^{z_2}\mathrm{d}z\frac{\mathrm{d}\ln P}{\mathrm{d}z}[\kappa(P,\boldsymbol{Q}) - \kappa(P,\boldsymbol{Q}_1)], \tag{2.13}$$

and

$$\kappa(P,\boldsymbol{Q}) \equiv \frac{\partial\ln\rho(P,\boldsymbol{Q})}{\partial\ln P}, \tag{2.14}$$

is the dimensionless compressibility. To obtain the second equality in (2.13), we have used the fundamental theorem of calculus, i.e. we have differentiated $\mathcal{R}$ with respect to z_2 and integrated the result from z_1 to z_2. Equation (2.13) separates the integrated linear buoyancy response (the first integral on the second line) with the nonlinear response (the second integral), revealing the latter to be determined by the path-integrated difference in κ between the moving parcel and its surroundings.[2] If the displaced fluid is more compressible than the fluid through which it moves, i.e. $\kappa(P,\boldsymbol{Q}_1) > \kappa(P,\boldsymbol{Q})$, then the second integral on the right-hand side of (2.13) has the same sign as δz, so its contribution to the buoyancy force is destabilising. If the stabilising effect of the integral involving $\mathcal{L}$ is sufficiently small, then the second integral dominates in (2.13) for δz larger than some critical value, δz_c. Such an equilibrium is metastable.

2.3. *Direction and size of displacement required for nonlinear instability*

Because metastability requires more compressible fluid to be moved through less compressible fluid [see (2.13)], equilibria can be metastable only to perturbations in the direction in which the compressibility of the background fluid decreases.[3] This direction is given by the sign of the compressibility scale height

$$H_\kappa \equiv -\left(\frac{\partial\ln\kappa}{\partial\boldsymbol{Q}}\cdot\frac{\mathrm{d}\boldsymbol{Q}}{\mathrm{d}z}\right)^{-1}. \tag{2.15}$$

We can estimate the local value of δz_c in the limit of $\mathcal{L}\to 0$ by balancing the sizes of the two integrals in (2.13) to find that

$$\delta z_c \simeq \frac{H_P H_\kappa\mathcal{L}}{\kappa}, \tag{2.16}$$

where we have defined the pressure scale height $H_P \equiv |\mathrm{d}\ln P/\mathrm{d}z|^{-1}$, neglected variation in $\mathcal{L}$ on the scale δz_c and used the fact that $-\kappa H_\kappa^{-1}$ is the coefficient of δz in the small-δz

[2]Note that the term in (2.13) that involves the compressibility of the surroundings cancels between the nonlinear response and the integral of (2.8); the nonlinear response may therefore be viewed as a correcting for the fact that the term involving $\mathcal{L}$ uses the 'wrong' compressibility (that of the background, rather than that of the moving parcel) to determine the density change.

[3]The exception is when the compressibility has a local extremum. In that case, the equilibrium may be unstable to both upwards and downwards perturbations (corresponding to a local maximum of the compressibility) or to neither (a local minimum). We provide explicit examples of both cases in § 2.6.

expansion of the term in square brackets in (2.13). According to (2.16), δz_c becomes arbitrarily small compared with any stratification scale height as $\mathcal{L} \to 0$. Nonetheless, we note that, because metastability of stratified equilibria is a nonlinear effect, it is not captured by Boussinesq-like equations that employ linear approximations of equilibrium gradients.

2.4. *Explosive instability*

The equation of motion of a small fluid parcel displaced by a distance $\delta z \sim \delta z_c$ that is much smaller than any scale height H of the stratification is

$$\frac{\mathrm{d}^2 \delta z}{\mathrm{d}t^2} = \left(-\delta z + \frac{\delta z^2}{\delta z_c} \right) \mathcal{L}. \tag{2.17}$$

It follows from (2.17) that, for $0 < \delta z_c \ll \delta z \ll H$ (the case of $\delta z_c < 0$ is analogous), the motion of a fluid parcel is explosive, *viz.*,

$$\delta z \propto \frac{1}{(C - t)^2}, \tag{2.18}$$

where the constant C is determined by initial conditions. Explosive growth of δz persists until either $\delta z \sim H$, whereupon (2.17) is no longer valid, the rising fluid element is shredded by Kelvin–Helmholtz instability or its speed approaches that of sound and thus (2.5) no longer applies.

2.5. *Case of 2D MHD*

In the remainder of this paper (with the exception of Appendix B, where we consider moist hydrodynamics), we focus on the case of an equilibrium supported against gravity both by thermal pressure p and by the magnetic pressure associated with a straight magnetic field with spatially dependent strength B. Thus, $P = p + B^2/2$. We restrict attention to 2D dynamics in the plane perpendicular to the magnetic field (i.e. to 2D interchanges of flux tubes). Then the specific magnetic flux

$$\chi = \frac{B}{\rho}, \tag{2.19}$$

is conserved in a Lagrangian sense. For symmetry with this definition of χ, we take advantage of the fact that any monotonically increasing function of specific entropy constitutes a good choice for the conserved quantity s, and thus let

$$s = \frac{p^{1/\gamma}}{\rho}, \tag{2.20}$$

where γ is the adiabatic index. To avoid confusion with the true specific entropy, which is proportional to $\exp(s/C_v)$ with C_v the specific heat capacity, we hereafter refer to s as the 'entropy function' (akin to potential temperature in atmospheric science, see Appendix B).

Thus, $\boldsymbol{Q} = (s, \chi)$ and

$$P(\rho, \boldsymbol{Q}) = \rho^\gamma s^\gamma + \rho^2 \chi^2/2. \tag{2.21}$$

With these choices, (2.7) yields the linear-stability condition

$$\mathcal{L} = \frac{c_s^2}{c^2}\frac{\mathrm{d}\ln s}{\mathrm{d}z} + \frac{v_A^2}{c^2}\frac{\mathrm{d}\ln \chi}{\mathrm{d}z} > 0, \tag{2.22}$$

where $c_s \equiv \sqrt{\gamma p/\rho}$ is the sound speed, $v_A \equiv B/\sqrt{\rho}$ the Alfvén speed and $c = \sqrt{c_s^2 + v_A^2}$ is the velocity of compressive waves [(2.22) is sometimes called the modified Schwarzschild criterion]. The compressibility κ (2.14) is

$$\kappa(P, s, \chi) = \frac{1+\beta}{2+\gamma\beta}, \tag{2.23}$$

where $\beta \equiv 2p/B^2$ is the plasma beta (the ratio of thermal to magnetic pressures) which is determined from P, s and χ via

$$\frac{1}{\beta}\left(1+\frac{1}{\beta}\right)^{2/\gamma-1} = \frac{1}{2}\left(\frac{\chi}{s}\right)^2 P^{2/\gamma-1}. \tag{2.24}$$

Equation (2.23) reveals that κ increases monotonically with β for $\gamma < 2$, from $\kappa = 1/2$ at $\beta = 0$ to $\kappa \to 1/\gamma$ as $\beta \to \infty$. It follows that the nonlinear buoyancy response in (2.13) is destabilising when fluid with large β moves through ambient fluid with smaller β. According to (2.24), the β of a flux tube with given s and χ depends on pressure and therefore is not constant during its motion.[4] At any given pressure, however, β is a monotonically increasing function of the ratio s/χ. Therefore, an equilibrium that is sufficiently close to marginal linear stability is always nonlinearly unstable in the direction in which s/χ decreases [the same conclusion can be obtained from (2.15) or by expanding (2.5) to quadratic order in δz directly, see Appendix C].

2.6. *Examples of metastable equilibria*

Explicit examples of metastable equilibria may be obtained as follows. First, we change variables from height z to the total mass supported at height z

$$m = \int_z^\infty \rho(z')\,\mathrm{d}z', \tag{2.25}$$

so that the equilibrium condition becomes $P = mg$, which may be expressed as

$$\rho^\gamma s^\gamma + \frac{1}{2}\rho^2\chi^2 = mg, \tag{2.26}$$

or, using (2.24), as

$$\frac{1}{\beta}\left(1+\frac{1}{\beta}\right)^{2/\gamma-1} = \frac{1}{2}\left(\frac{\chi}{s}\right)^2 (mg)^{2/\gamma-1}. \tag{2.27}$$

Throughout this work, we shall find it convenient to use the supported mass m as a proxy for height because the mass of a flux tube is preserved as it moves, while its cross-sectional area is not (note that m decreases with increasing z).

[4] According to (2.24), β increases when a flux tube with fixed s and χ rises while in total-pressure balance with its surroundings. For this reason, an equilibrium with $\beta(z) = \text{const.}$ is metastable to upwards perturbations if it is sufficiently close to marginal linear stability.

We seek an equilibrium that is close to marginal linear stability [so that the second integral in (2.13) stands a chance of dominating the first]. We therefore demand that

$$\mathcal{L} = \frac{\epsilon}{H_P}, \tag{2.28}$$

where $\mathcal{L}$ is given by (2.22), H_P is the total-pressure scale height $[\mathrm{d}\ln P/\mathrm{d}z]^{-1} = m/\rho$ and $\epsilon \ll 1$ is a small number. Equation (2.28) constitutes a differential equation involving s, χ and their gradients in m – to solve it, we specify a relationship between s, χ and m that defines the particular equilibrium under consideration. We choose to specify as a function of m the ratio s/χ, which controls the compressibility of the fluid (§ 2.5). We therefore recast (2.28) [with $\mathcal{L}$ given by (2.22)] as

$$\frac{\mathrm{d}\ln s}{\mathrm{d}m} = \frac{2}{\gamma\beta + 2}\frac{\mathrm{d}}{\mathrm{d}m}\ln\left(\frac{s}{\chi}\right) - \frac{\epsilon}{m}, \tag{2.29}$$

which we integrate numerically for s as a function of m. We determine the dependence of all other quantities on m (or z) via their definitions and (2.25) and (2.27).

Some example cases are visualised in figure 3, where we show s, χ, s/χ and β as functions of z (lower row of panels) together with the force (2.5) per unit mass on a small parcel of fluid moved from height z_1 to z_2 as a function of z_1 and z_2 (upper two rows of panels). Figure 3*a* shows the case of

$$\frac{s}{\chi} = (m_{\mathrm{tot}}g)^{1/\gamma-1/2}\left(3\frac{m}{m_{\mathrm{tot}}} + 0.3\right), \tag{2.30}$$

for which s/χ increases with m. The most compressible material is therefore at the bottom of the atmosphere and the equilibrium is unstable to upwards displacements (towards larger z). This is the equilibrium whose relaxation is visualised in figures 1 and 2. Figure 3*b* shows the case of

$$\frac{s}{\chi} = (m_{\mathrm{tot}}g)^{1/\gamma-1/2}\left(3\frac{m}{m_{\mathrm{tot}}} + 0.3\right)^{-1}, \tag{2.31}$$

for which s/χ decreases with m, so produces an equilibrium unstable to downward displacements (towards smaller z). Figure 3*c* shows the case of

$$\frac{s}{\chi} = (m_{\mathrm{tot}}g)^{1/\gamma-1/2}\left[3\exp\left(-\frac{(m/m_{\mathrm{tot}} - 0.2)^2}{0.1^2}\right) + 0.3\right], \tag{2.32}$$

for which s/χ has a maximum at $m = 0.2m_{\mathrm{tot}}$; the equilibrium is therefore unstable to both upwards and downwards displacements in the vicinity of the maximum. Finally, figure 3*d* shows the case of

$$\frac{s}{\chi} = (m_{\mathrm{tot}}g)^{1/\gamma-1/2}\left[3\left(1 - \exp\left(-\frac{(m/m_{\mathrm{tot}} - 0.2)^2}{0.1^2}\right)\right) + 0.3\right], \tag{2.33}$$

for which s/χ has a minimum at $m = 0.2m_{\mathrm{tot}}$. The most compressible fluid is therefore situated at the top and bottom of the atmosphere, so it is unstable to both downwards and upwards displacements.

In figures 1–3 and in the rest of the paper, we choose units of mass and length such that $\rho = 1$ and $m = m_{\mathrm{tot}} = 1$ at the bottom of the atmosphere ($z = 0$), so the total-pressure scale height $H_P = 1$ there. We choose units of time such that $g = 1$; it follows that the time

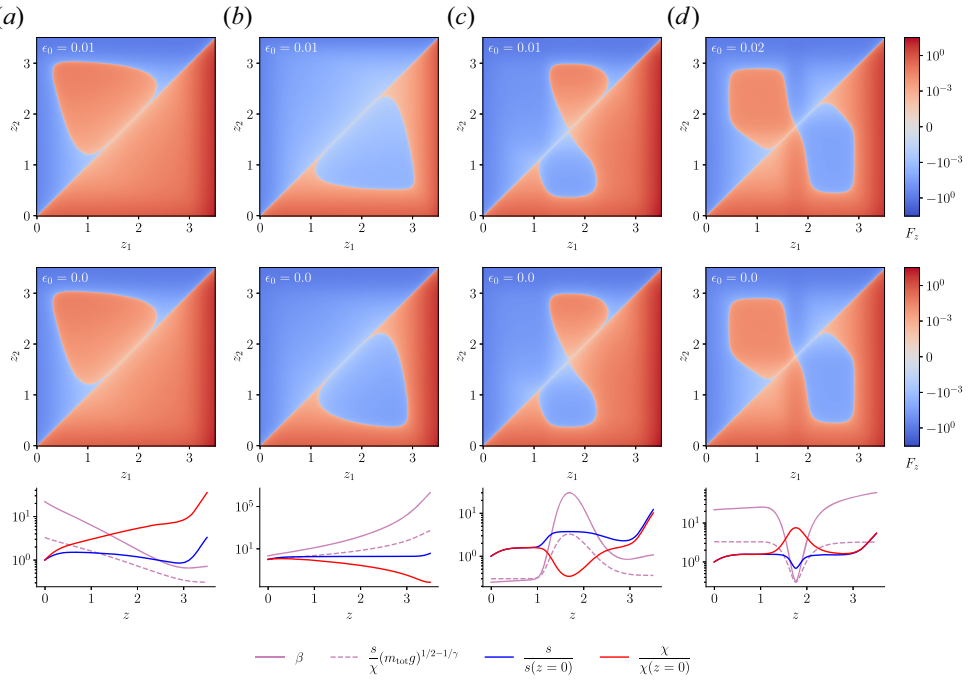

FIGURE 3. Top row: the upward force (2.5) per unit mass on a small fluid parcel moved in pressure balance and without diffusion from height z_1 to z_2 for each of the profiles described in § 2.6. Values of ϵ_0 in (2.34) are indicated on each panel. Middle row: the same as the top row, but with $\epsilon_0 = 0$ (marginal linear stability in the bulk). Bottom row: profiles of the entropy function $s = p^{1/\gamma}/\rho$, specific magnetic flux $\chi = B/\rho$, their ratio and the plasma β (2.27) as a function of height z (these profiles correspond to the top row specifically, but the profiles that correspond to the middle row look essentially the same because the differences in ϵ_0 are very small). Panel (*a*) corresponds to (2.30), (*b*) to (2.31), (*c*) to (2.32) and (*d*) to (2.33).

taken for a compressive wave to traverse a total-pressure scale height $H_P/c \sim \sqrt{H_P/g}$ is 1 at the bottom of the atmosphere. We limit the total height of the atmospheres to $z_{\max} = 3.5$ in these units, where $\rho \lesssim 10^{-2}$ in each of the four cases described above. We stabilise the equilibria near $z = 0$ and $z = z_{\max}$ by choosing

$$\epsilon = \epsilon_0 + \tanh\left(\frac{z - z_u}{\Delta_u}\right) - \tanh\left(\frac{z - z_l}{\Delta_l}\right) + 2, \tag{2.34}$$

in (2.28), with $\Delta_l = \Delta_u = 0.25$, $z_l = 0.25$, $z_u = 3.25$. This ensures that artificial boundary effects do not impact our numerical or theoretical analyses. We choose $\epsilon_0 = 10^{-2}$ for the simulations visualised in figures 1 and 2.

The top row of panels in figure 3 correspond to equilibria with small positive values of ϵ_0; these are linearly stable, and exhibit the various sorts of metastability described above. Unless stated otherwise, we shall in the rest of the paper focus attention on the case of marginal linear stability, i.e. $\epsilon_0 = 0$ in (2.34). This is because we expect nonlinear instability to be triggered for equilibria close to marginal linear instability in practice. The middle row of figure 3 corresponds to the $\epsilon_0 = 0$ case.

3. Available potential energy

In this section, we calculate the energy that can be liberated from a metastable MHD equilibrium by 2D rearrangement of flux tubes. Because the gravitational potential energy depends only on z, we shall assume *a priori* that the state with minimum energy is 1D: all quantities depend only on z (or, equivalently, on m). It will turn out that this assumption can, and often does, fail, but the 2D minimum-energy states will be extractable from the solution to the 1D minimisation.

The total potential energy per unit horizontal length of a 1D state (not necessarily in equilibrium) is

$$E_{\rm tot} = \int_0^\infty \mathrm{d}z \left(\frac{p}{\gamma - 1} + \frac{B^2}{2} + \rho g z \right) = \int_0^{m_{\rm tot}} \mathrm{d}m \left[\mathcal{E}(P, s, \chi) + \frac{mg - P}{\rho(P, s, \chi)} \right], \tag{3.1}$$

where $\rho(P, s, \chi)$ is determined implicitly from the definition of total pressure, $P = \rho^\gamma s^\gamma + \rho^2 \chi^2/2$. In the second equality of (3.1) we have integrated by parts and introduced the specific enthalpy

$$\mathcal{E} \equiv \frac{1}{\rho} \left(\frac{p}{\gamma - 1} + \frac{B^2}{2} + P \right) = \frac{\gamma}{\gamma - 1} \rho^{\gamma - 1} s^\gamma + \rho \chi^2. \tag{3.2}$$

It is readily verified that

$$\left(\frac{\partial \mathcal{E}}{\partial P} \right)_{s,\chi} = \frac{1}{\rho}, \tag{3.3}$$

whence

$$E_{\rm tot} = \int_0^{m_{\rm tot}} \mathrm{d}m \left\{ \mathcal{E}(mg, s, \chi) + \frac{1}{2} \frac{P^2}{\rho^2 c^2} \left(\frac{\delta P}{mg} \right)^2 + \mathcal{O}\left[\left(\frac{\delta P}{mg} \right)^3 \right] \right\}, \tag{3.4}$$

where $\delta P \equiv P - mg$. Evidently, $E_{\rm tot}$ is minimal with respect to P when $P = mg$. It follows that, when looking for the minimum-energy state, we can restrict attention to those states with $P = mg$, i.e. those in static equilibrium at all m. Equation (3.4) then reduces to

$$E_{\rm tot} = \int_0^{m_{\rm tot}} \mathrm{d}m\, \mathcal{E}(mg, s, \chi). \tag{3.5}$$

Following Lorenz (1955), we discretise the integral (3.5) by 'slicing' the atmosphere into thin layers of equal mass Δm that we label by the index i. We consider the 1D equilibria formed by rearranging the slices while conserving the entropy and flux in each slice. Each possible rearrangement is a permutation map $i \to j = \sigma(i)$, under which the slice that initially supports mass m_i now supports mass m_j. The energy of the rearranged equilibrium is

$$E_{\rm tot} \simeq \Delta m \sum_{i=0}^{N} \mathcal{E}(m_{\sigma(i)} g, s_i, \chi_i), \tag{3.6}$$

where $N = m_{\rm tot}/\Delta m$, and s_i, χ_i are, respectively, the entropy function and specific magnetic flux of slice i. We seek the permutation σ that gives the smallest possible value of $E_{\rm tot}$, which we denote $E_{\rm min}$.

Because the energy associated with assigning slice i to support a mass of m_j only depends on i and j and not on the assignments of other slices, minimising (3.6) over

permutations σ is a combinatorial optimisation problem known as linear sum assignment (LSA) (see, e.g. Burkard, Dell'Amico & Martello 2012). The LSA problem is canonically described as one of minimising the total cost associated with assigning a number of 'agents' to the same number of 'tasks' – in our case, the 'agents' are the slices of atmosphere with given s and χ, while their 'tasks' are to occupy discrete positions in the atmosphere corresponding to each possible value of the discretised supported mass. The matrix of costs associated with assigning agent i to task j is $\mathcal{E}(m_j g, s_i, \chi_i)$.

For economy of notation, we hereafter denote the cost matrix $\mathcal{E}(m_j g, s_i, \chi_i)$ by $\mathcal{E}(m_j, m_i)$: this is the energy cost associated with assigning the slice initially at m_i to m_j (in a minor abuse of notation, we use the same symbol, $\mathcal{E}$, for both functions). Despite our discretisation in m, $\mathcal{E}(m_j, m_i)$ remains a continuous function of its arguments and we shall frequently be required to integrate or take derivatives with respect to one or the other in what follows. We shall, therefore, introduce the continuous variable μ to denote the supported mass of a slice in the initial state, and denote the continuous form of $\mathcal{E}(m_j, m_i)$ by $\mathcal{E}(m, \mu) \equiv \mathcal{E}(mg, s(\mu), \chi(\mu))$. Similarly, we shall write $\rho(m, \mu)$ as a shorthand for $\rho(mg, s(\mu), \chi(\mu))$. In this notation, (3.3) becomes

$$\frac{\partial \mathcal{E}(m, \mu)}{\partial m} = \frac{g}{\rho(m, \mu)}. \tag{3.7}$$

3.1. *Estimating the available energy*

Before proceeding to solve the LSA problem, which can only be achieved numerically in most cases, let us try to estimate the outcome analytically: What is the typical available potential energy of a metastable atmosphere? This turns out to be a small fraction of the total potential energy, a fact we that we utilise in § 5 to predict the relaxation of destabilised equilibria.

The change in potential energy δE that results from moving a slice of atmosphere upwards from supported mass m_a to new supported mass m_b, shuffling downwards the slices that it passes on the way, is

$$\begin{aligned}\frac{\delta E}{\Delta m} &= \mathcal{E}(m_b, m_a) - \mathcal{E}(m_a, m_a) + \sum_{i=b}^{a-1} [\mathcal{E}(m_{i+1}, m_i) - \mathcal{E}(m_i, m_i)] \\ &\simeq \int_{m_a}^{m_b} \mathrm{d}m \left[\frac{g}{\rho(m, m_a)} - \frac{g}{\rho(m, m)} \right] \\ &= -g \int_{z_b}^{z_a} \mathrm{d}z \left[\frac{\rho(m(z), m(z))}{\rho(m(z), m_a)} - 1 \right],\end{aligned} \tag{3.8}$$

where $m(z_a) = m_a$ and $m(z_b) = m_b$. We have used (3.7) in moving from the first to the second line of (3.8). The integrand that appears in the last line of (3.8) is straightforwardly recognised as the net buoyancy force (2.5) per unit mass of fluid, so, sensibly, δE is just the work done by this force on the moving slice.[5] Evidently, the energy that can be liberated by this process is maximal when the ratio that appears inside the integrand in the last line of (3.8) – i.e., of the density of the ambient fluid to the density of the moving slice – is maximal. We can evaluate this ratio using (2.13), which yields, in the most optimistic case of (*i*) marginal linear stability, i.e. $\mathcal{L} = 0$, (*ii*) $\kappa = 1/\gamma$ inside the slice

[5]Work is done in moving the slice through its series of adjacent equilibria because interchanging two slices in practice involves the fluid in each of them passing through non-equilibrium states.

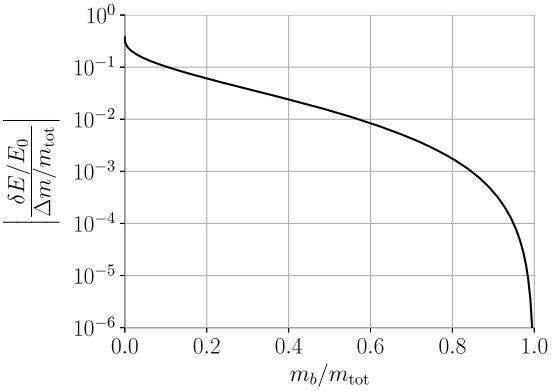

FIGURE 4. The fraction (3.10) of energy liberated when a slice of fluid with mass Δm is moved from the bottom of an atmosphere at marginal linear stability to a new position where the supported mass is m_b, under the most optimistic assumptions about the compressibility of the slice and that of the fluid through which it moves.

(maximally compressible large-β fluid) and (*iii*) $\kappa = 1/2$ for the ambient fluid (minimally compressible small-β fluid)

$$\frac{\rho(m, m_a)}{\rho(m, m)} = \left(\frac{m}{m_a}\right)^{1/\gamma - 1/2}. \tag{3.9}$$

Substituting this into (3.8), choosing $m_a = m_{\rm tot}$ (the slice originates from the bottom of the atmosphere) and evaluating integrals, we find that

$$\frac{\delta E}{E_0} = \frac{3}{4}\frac{\Delta m}{m_{\rm tot}}\left[\frac{\gamma}{\gamma - 1}\left(\left(\frac{m_b}{m_{\rm tot}}\right)^{1-1/\gamma} - 1\right) - 2\left(\left(\frac{m_b}{m_{\rm tot}}\right)^{1/2} - 1\right)\right], \tag{3.10}$$

where E_0 is the initial total energy of the atmosphere (assuming that it is mostly populated with small-β fluid, as is consistent with assumption (*iii*) above).

We plot (3.10) in figure 4 for the case of $\gamma = 5/3$. We observe that the fraction of energy liberated per mass fraction of fluid moved decreases sharply as a function of $m_b/m_{\rm tot}$ (even though its limiting value of $3/8$ as $m_b/m_{\rm tot} \to 0$ is finite). This is significant as, because fluid parcels exclude each other (i.e. different slices cannot all be assigned to the same supported mass), the vast majority of slices that are interchanged in a global rearrangement experience an order-unity change in supported mass. Such reassignments are far less profitable that ones that take a slice all the way to the top of the atmosphere (i.e. $m_b/m_{\rm tot} \to 0$). For example, figure 4 shows that the fraction of energy liberated per mass fraction of fluid moved is approximately 10^{-2} for $m_b = m_{\rm tot}/2$. The reason is that an order-unity change in pressure only results in a small change in density, owing to the smallness of the exponent $1/\gamma - 1/2 = 0.1$ that appears on the right-hand side of (3.9). A more detailed calculation (see Appendix D) reveals that 10^{-2} is indeed a good estimate for the fractional available energy: the maximum fractional available energy of a (marginally) stable atmosphere consisting of small-β fluid above a layer of large-β fluid is 1.75 %.

The above estimates correspond to the most optimistic assumptions: we shall find that the fractional available energies of the equilibria described in § 2.6 and represented in figure 3 are less than 1 % (by roughly an order of magnitude), because (*i*) these equilibria have finite β, and thus do not have the extremal values of the fluid compressibility

considered here; and (*ii*) the stable buffer regions contribute to the potential energy of the equilibrium (and prevent the relaxing fluid from accessing large pressure differences) but do not participate in the reassignment.

3.2. *The Hungarian algorithm*

The discrete energy-minimisation problem (3.6) can be solved without recourse to numerical optimisation only in the limits $\beta \to \infty$ and $\beta \to 0$, i.e. the cases where only one of thermal or magnetic pressure support the atmosphere against gravity. For $\beta \to \infty$, $\mathcal{E} \to \gamma s(mg)^{1-1/\gamma}/(\gamma - 1)$, which increases monotonically with both m and s. The arrangement with least total energy is therefore the one for which the slice with largest s has smallest m, the slice with the next largest s has the next smallest m, and so on. It follows that the profile with the smallest energy is the unique rearrangement for which s is a monotonically increasing function of height. Indeed, we already know this state to be nonlinearly stable by (2.12). A similar conclusion is obtained for $\beta \to 0$, for which $\mathcal{E} \to 2(mg)^{1/2}\chi$ and so the nonlinearly stable atmosphere is the one with χ increasing monotonically with height. Analogous simple constructions for the minimum-energy configuration do not exist in the finite-β case: the optimal solution that balances the competing imperatives of 'entropy should increase upwards' and 'flux should increase upwards' is non-trivial.

Solution of the LSA problem in general relies on the observation that the modified cost matrix $\tilde{\mathcal{E}}(m_j, m_i)$, where

$$\tilde{\mathcal{E}}(m_j, m_i) = \mathcal{E}(m_j, m_i) - a(m_j) - b(m_i), \tag{3.11}$$

has the same optimal assignment of agents to tasks as does the original cost matrix $\mathcal{E}(m_j, m_i)$. This is intuitive: if the cost of assigning a given agent (slice) to each task (position) decreases by the same amount, their optimal assignment will not change (although the total cost of the solution will decrease). Likewise, if a particular task becomes more expensive by the same amount for all agents, the optimal choice of agent for that task will remain the same. $\tilde{\mathcal{E}}$ is called the 'normal form' of the cost matrix $\mathcal{E}$ if it satisfies the following properties (Burkard *et al.* 2012):

(i) $\tilde{\mathcal{E}}(m_j, m_i) \geq 0$,
(ii) There exists at least one bijection σ such that $\tilde{\mathcal{E}}(m_{\sigma(i)}, m_i) = 0, \ \forall\, i$.

A proof that an $\tilde{\mathcal{E}}(m_j, m_i)$ with these properties can always be found with suitable choices of $a(m_j)$ and $b(m_i)$ is provided by the Hungarian algorithm (Kuhn 1955; Munkres 1957), which consists of a series of row and column operations on the cost matrix that are guaranteed to reduce it to normal form in polynomial (in N) time. The utility of the normal form $\tilde{\mathcal{E}}(m_j, m_i)$ is that each of the bijections σ constitutes an optimal assignment.[6]

3.3. *Energy minimisation: numerical results and interpretation*

In this section, we present the outcomes of energy minimisation using the Hungarian algorithm (§ 3.2) for each of the equilibria described in § 2.6 and represented in figure 3.

[6]Typically, the optimal assignment is unique. This is true for each of the equilibria introduced in § 2.6. One can construct counter-examples, however: the optimal assignment is non-unique when multiple slices have the same values of s and χ, for example.

3.3.1. *Metastable upwards,* (2.30)

We first consider the equilibrium defined by (2.30), which is nonlinearly unstable to upward displacements (figure 3*a*). The solution of the LSA problem is visualised in figure 5, which compares the initial and minimum-energy assignments of the slices. In the minimum-energy assignment, material initially from the bottom of the metastable part of the atmosphere (i.e. $z \gtrsim z_l$; see (2.34)) is reassigned to the top $z \lesssim z_u$, as is intuitive. The order in which the slices that are re-assigned are stacked also reverses. This happens because, as the material moves to smaller m, its β increases (2.27), and therefore the contribution of s to the linear-stability criterion becomes more important relative to χ. As a consequence, the stacking order reverses so that s increases upwards.

A striking feature of the minimum-energy state is that slices that were adjacent at the bottom of the atmosphere do not remain adjacent in the minimum-energy state. Instead, slices from the bottom become foliated with those at the top. The scale of the foliation is set by Δm, and therefore is arbitrarily small as $\Delta m \to 0$. This is illustrated by figure 6, which shows, for three different values of Δm, the normal-form cost matrix $\tilde{\mathcal{E}}(m_j, m_i)$ [defined by (3.11)], with its zeros (which indicate the optimal assignment $j = \sigma(i)$) marked with white circles. Because the scale of foliation depends on Δm, the optimal assignment of slices does not converge as $\Delta m \to 0$, although it does converge in a coarse-grained sense: the proportion of slices assigned to any small finite range of m that originated from any similarly small given range of μ converges as $\Delta m \to 0$ (provided that $\tilde{\mathcal{E}}$ converges as $\Delta m \to 0$).[7]

Foliation occurs when the material properties of the fluid vary with m at a different rate in the part of the atmosphere from which a slice originates than in the part to which it is reassigned. We demonstrate this fact by considering the motions of two slices from the bottom of the atmosphere to the top, each displacing downwards the slices through which they pass (as in § 3.1). Let the first slice have initial assignment m_a and new assignment m_b. Because the new assignment is optimal, the total energy has a local minimum as a function of displacement of this slice when its density in its new location, $\rho(m_b, m_a)$, is equal to the density $\rho(m_b, m_b)$ of the slices that surround it there (neutral buoyancy); this makes the integrand in the second line of (3.13) zero at $m = m_b$.[8] Now let us consider a second slice of fluid that initially neighbours the first, i.e. that originates from supported mass $m_a + \Delta m$. This slice reaches neutral buoyancy at a different supported mass $m_b + \delta m$. If the density of the background equilibrium changes more slowly with supported mass at m_b than at m_a, then $\delta m > \Delta m$. Setting the density of the second slice, $\rho(m_b + \delta m, m_a + \Delta m)$, equal to that of the ambient fluid at supported mass $m_b + \delta m$, i.e. $\rho(m_b + \delta m, m_b + \delta m)$ (because we are concerned with the motion of only two slices of infinitesimal thickness, we neglect the fact that the reassignment of the first slice might have changed the identity

[7] In the continuous limit, these proportions can be determined from the gradient of the locus of points for which $\tilde{\mathcal{E}} = 0$. For example, denoting this locus, i.e. the curve to which the white circles in figure 6 converge as $\Delta m \to 0$, by $m_{\mathrm{opt}}(\mu)$, then, provided $m_{\mathrm{opt}}(\mu)$ is single valued (as is the case in figure 6), the proportion of slices assigned to the vicinity of m_2 that originally had supported mass m_1 is $|\mathrm{d}m_{\mathrm{opt}}/\mathrm{d}\mu|$ evaluated at m_1 (by the conservation of mass). In the absence of foliation, $|\mathrm{d}m_{\mathrm{opt}}/\mathrm{d}\mu|^{-1} = 1$, but, where there is foliation, it must be the case that $|\mathrm{d}m_{\mathrm{opt}}/\mathrm{d}\mu| > 1$, so $m_{\mathrm{opt}}(\mu)$ steepens. Both cases may be observed in figure 6. The generalisation to the case where m_{opt} is multi-valued is somewhat more complex, but it remains true in that case that the fractional assignments are determined by gradients of the optimal-solution curve (see Appendix I).

[8] Intuitively, if the densities were different, the new equilibrium would be Rayleigh–Taylor unstable either at the upper or lower surface of the slice.

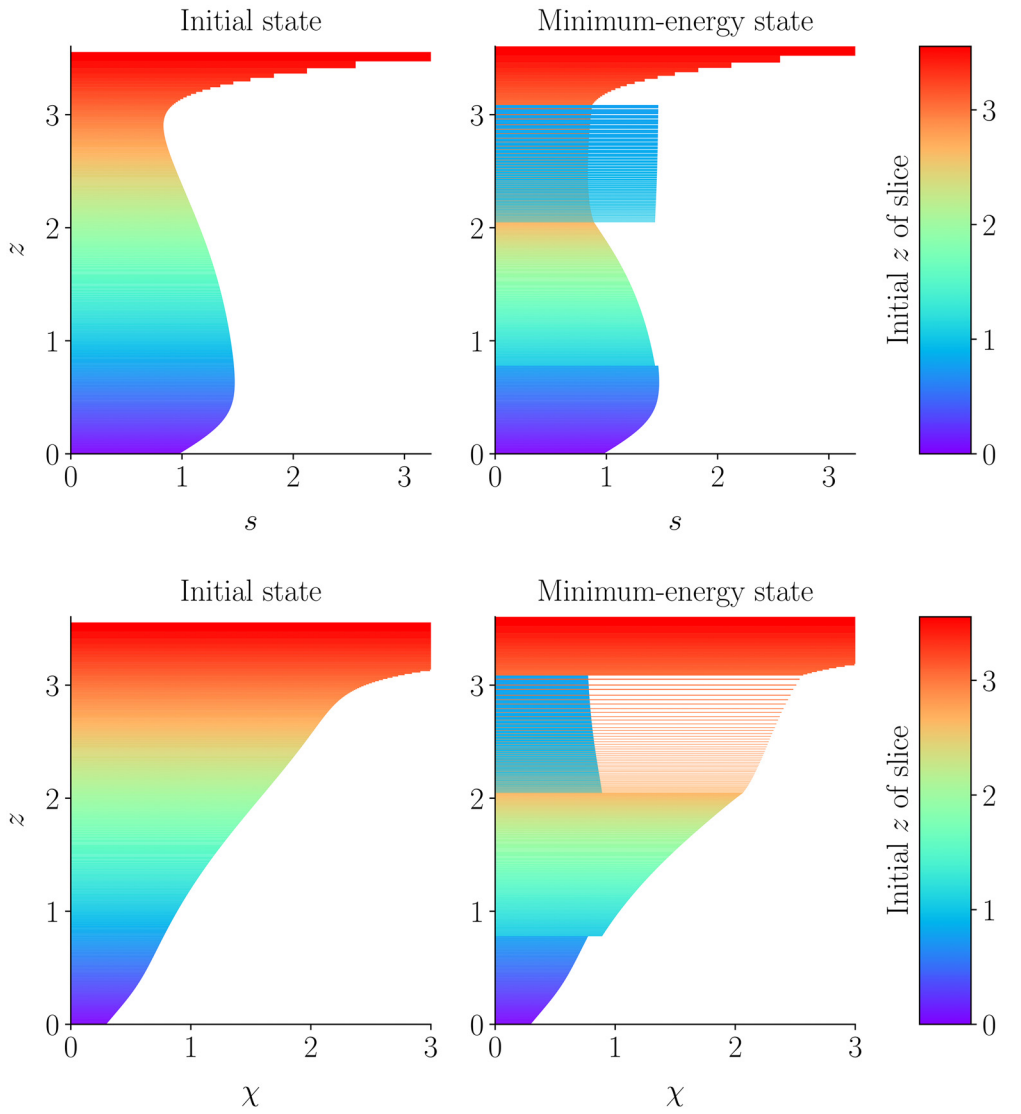

FIGURE 5. Visualisation of the assignment of 1D slices that minimises the total energy, (3.6), with $\Delta m = 5 \times 10^{-4} m_{\text{tot}}$, for the upwards-unstable profile defined by (2.30). Panels on the left show the initial profiles of s and χ as functions of height z, while panels on the right show the minimum-energy assignment. The slices are coloured by their height z in the initial state to aid comparison. Blue slices from $0.8 < z < 1.0$ are moved to $2.1 < z < 3.1$, reversing order and foliating with red slices originally from $2.7 < z < 3.1$. Each slice has vertical extent $\Delta z = \Delta m/\rho$ with ρ given by (2.26).

of the fluid at m_b), we discover that

$$\frac{\delta m}{\Delta m} = \left.\frac{\partial \rho(m_b, \mu)}{\partial \mu}\right|_{\mu=m_a} \Bigg/ \left.\frac{\partial \rho(m_b, \mu)}{\partial \mu}\right|_{\mu=m_b}. \tag{3.12}$$

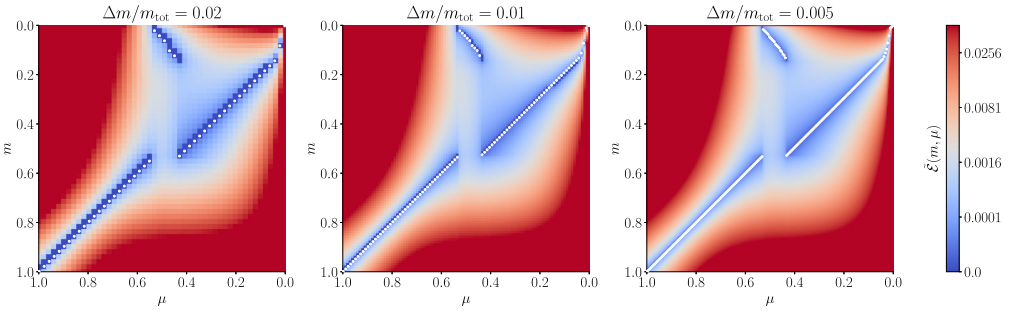

FIGURE 6. The normal form of the cost matrix $\tilde{\mathcal{E}}(m, \mu)$ [see (3.11) for its definition] for the unstable-upwards profile defined by (2.30), for three different choices of the discretisation scale Δm. White dots show the optimal assignment.

Thus, the adjacency of the slices if not preserved ($\delta m \neq \Delta m$) if the rate of change with supported mass of the density that a slice would have if moved to m_b is different for slices originating at m_a than at m_b.

Although (3.12) reveals the physical reason for foliation, it fails if the fraction on the right-hand side is small ($\delta m \lesssim \Delta m$ is not allowed for the discrete problem because slices exclude each other). It also does not apply in the case where a substantial mass of fluid is reassigned, in which case the background equilibrium through which the first slice moves is different from that through which the last slice does. A general treatment of such cases is as follows. Let us suppose that we are somehow given all the optimal assignments $\sigma(i)$, except those to some small range of m, i.e. those with $m_{\sigma(i)} = m_b + \delta m_{\sigma(i)}$, and seek the condition under which the optimal choice of the remaining assignments will be a foliated state. The contribution to the total energy of the slices that remain to be assigned is

$$\begin{aligned}\delta E &= \Delta m \sum_i \mathcal{E}(m_b + \delta m_{\sigma(i)}, m_i) \\ &= \Delta m \sum_i \left[\mathcal{E}(m_b, m_i) + \frac{\delta m_{\sigma(i)} g}{\rho(m_b, m_i)} + \mathcal{O}(\delta m_{\sigma(i)}^2) \right], \end{aligned} \tag{3.13}$$

where sums are over all indices i of the slices that remain to be assigned, and we have used (3.7) to obtain the second equality. The first term inside the square bracket in the second line of (3.13) is independent of the assignment σ. The second term takes the form of the differential supported mass $\delta m_{\sigma(i)}$ multiplied by a quantity that does not depend on $\sigma(i)$, *viz.*, $1/\rho(m_b, m_i)$. This yields a local stacking rule: δE is minimised when the slice with largest $\rho(m_b, m_i)$ is assigned to the largest supported mass, the next largest $\rho(m_b, m_i)$ is assigned to the next largest supported mass, and so on. This is physically intuitive: if more dense slices were situated above less dense ones, the equilibrium would be Rayleigh–Taylor unstable. We deduce that in order for slices from different initial locations to become foliated in the final state they must have

$$\rho(m_b, m_{i'}) - \rho(m_b, m_i) = \mathcal{O}(\Delta m). \tag{3.14}$$

As $\Delta m \to 0$, slices can be foliated in the vicinity of some m only if they have the same density at that m.

A foliated minimum-energy state may be considered a 1D representation of a minimum-energy state that is actually 2D, as follows. Because the scale of foliation is

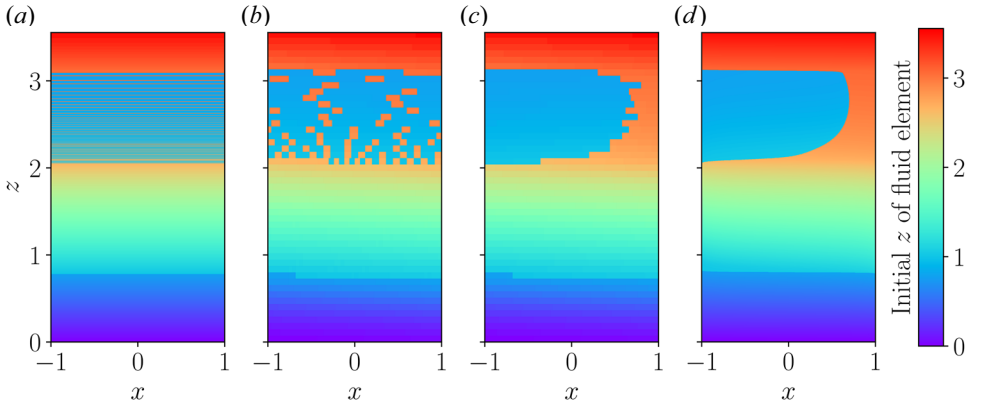

FIGURE 7. 2D minimum-energy states. Panel (*a*) shows the 1D stacking with minimum energy in the $x-z$-plane. Panel (*b*) shows a discrete approximation to the 2D ground state, obtained by arranging vertical slices sequentially into fixed vertical bins of fixed extent Δz, while preserving the P, s and χ of each slice. Each slice occupies the same area as before (because its mass and density each remain constant) but now the slices are arranged horizontally within the bin, with left to right corresponding to increasing height in panel (*a*). Panel (*c*) shows the state in panel (*b*) but sorted horizontally, which does not change the energy and removes the imprint of the foliation. Panel (*d*) shows the expected equivalent of Panel (*c*) at very large resolution, but is obtained differently, by taking the small-thermodynamic-temperature limit of Lynden-Bell statistical mechanics (§ 5.2).

set by Δm, we can reconfigure the foliated state (figure 7*a*) into a 2D state (figure 7*b*) by rearranging fluid parcels locally in z, i.e. over a small vertical distance $\delta z \sim \Delta m/\rho$, and then sorting globally in x (with P, s and χ fixed for each parcel) to remove small-scale variation (figure 7*c*,*d*). Because the foliated state is a stable equilibrium with respect to local rearrangements in z, the force acting on each fluid parcel in the new state will be proportional to δz, and therefore vanishes as $\Delta m \to 0$. Thus, the 2D state is also an equilibrium state with the same energy as the foliated one.[9] In § 4, we shall show with direct numerical simulations that minimum-energy 2D states are the result of the nonlinear relaxation of a destabilised metastable state in certain regimes.

To conclude our discussion of minimum-energy states of equilibria defined by (2.30), we present in figure 8 a comparison of energy-minimising assignments for different values of the parameter ϵ_0, which controls the distance from marginal stability [see (2.34)]. First, we note that $\epsilon_0 = 0$, i.e. marginal linear stability, is not a special point as far as the minimum-energy assignments are concerned: the assignments with $\epsilon_0 = 0.0075$ (linearly stable), $\epsilon_0 = 0.0$ (marginal) and $\epsilon_0 = -0.015$ (linearly unstable) are qualitatively similar, although, in the unstable cases of $\epsilon_0 = -0.015$ and $\epsilon_0 = -0.075$, foliation occurs over a much wider range of supported masses (and material from three different initial locations are foliated together near the top of the atmosphere).

Figure 9 shows the ratio of the available energy $E_{\mathrm{avail}} = E_0 - E_{\mathrm{min}}$ to the original potential energy E_0 as a function of $\epsilon_c - \epsilon_0$, where $\epsilon_c \simeq 1.7 \times 10^{-2}$ is the largest value

[9]More formally, the difference in energy between the two states under the operation described is $E_{2\mathrm{D}} - E_{1\mathrm{D}} = \int \mathrm{d}z\mathrm{d}x\, \rho g \delta z$. Because $\delta z = \mathcal{O}(\Delta m)$ as $\Delta m \to 0$, we would have that $E_{2\mathrm{D}} - E_{1\mathrm{D}} = \mathcal{O}(\Delta m)$ if there existed a finite density difference between neighbouring slices. However, if the difference in density between neighbouring slices is $\mathcal{O}(\Delta m)$ [as required by (3.14)], then δz is the only rapidly varying function of x and z in the integral and therefore can be replaced by its coarse-grained average. This average is zero, because there is no net displacement of fluid parcels in each horizontal band. Thus $E_{2\mathrm{D}} - E_{1\mathrm{D}} = \mathcal{O}(\Delta m^2)$, so the difference in energy per fluid parcel vanishes.

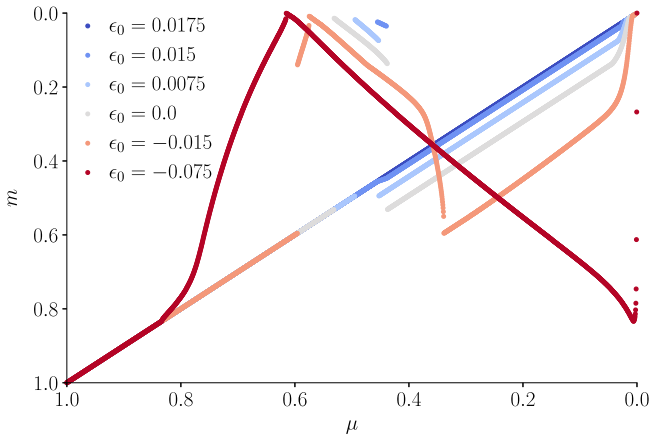

FIGURE 8. The energy-minimising assignments of slices from initial supported mass μ to new supported mass m for the initial profile (2.30) with different values of the parameter ϵ_0, which controls linear stability via (2.34).

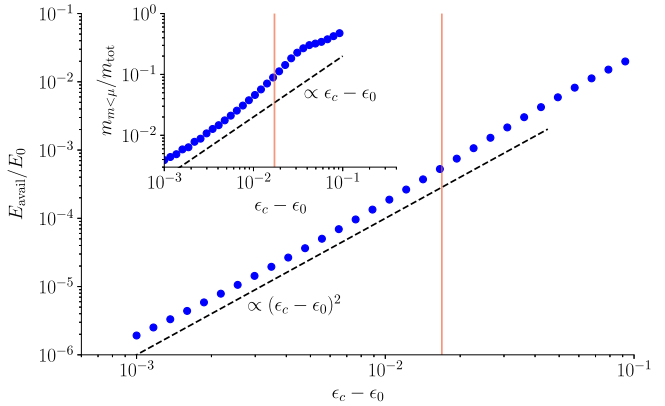

FIGURE 9. The available energy $E_{\mathrm{avail}} = E_0 - E_{\mathrm{min}}$ as a fraction of the initial potential energy E_0, plotted as a function of $\epsilon_c - \epsilon_0$, where $\epsilon_c \simeq 1.7 \times 10^{-2}$ is the largest value of ϵ_0 in (2.34) for which the initial state is metastable. The inset shows the fraction of fluid that is assigned to a smaller supported mass than its initial one under optimal reassignment. Red lines correspond to $\epsilon_0 = 0$.

of ϵ_0 for which the atmosphere has a restacking with smaller energy. We see that E_{avail}/E_0 is small: it is around 10^{-3} for $\epsilon_0 = 0$, which is the value that corresponds to figures 5 and 6. This is despite the fact that the minimum-energy assignment involves significant rearrangement of the atmosphere (around 10 % by mass of the atmosphere is reassigned upwards for $\epsilon_0 = 0$, see inset to figure 9). As explained in § 3.1, the reason for the smallness of E_{avail}/E_0 is the fact that fluid slices exclude each other and so only very few of them can experience a significant change in total pressure as a result of reassignment.

We observe from figure 9 that

$$\frac{E_{\mathrm{avail}}}{E_0} \propto (\epsilon_c - \epsilon_0)^2 \text{ as } \epsilon_c - \epsilon_0 \to 0; \tag{3.15}$$

this scaling is readily interpreted as a result of the fact that both (*i*) the typical amount of energy liberated when a slice is reassigned from the bottom of the atmosphere to the top, and (*ii*) the number of slices that are reassigned in this way, are proportional to $\epsilon_c - \epsilon_0$ when the latter is small [see inset to figure 9; s, χ and, therefore, the buoyancy force on a displaced fluid element, (2.5), depend linearly on ϵ_0, by (2.29)]. This argument being independent of the particular profile under consideration, we expect a quadratic dependence of E_{avail}/E_0 on $\epsilon_c - \epsilon_0$ for any metastable profile equilibrium with ϵ_0 sufficiently close to ϵ_c. We shall find in § 3.3.2 that a quadratic scaling is indeed reproduced for the profile represented by (2.31).

3.3.2. *Metastable downwards,* (2.31)

We now turn to the second example case introduced in § 2, (2.31), which describes an initial state that is metastable to downwards displacements. The 1D minimum-energy state associated with this profile [with $\epsilon_0 = 0$ in (2.34)] is shown figure 10, which is the analogue for (2.31) of figure 5. The minimum-energy assignment is similar qualitatively to the one examined in § 3.3.1: in this case, material from the top of the atmosphere moves to the bottom, reverses stacking order, and becomes foliated with the material already there (in fact, material from three different initial heights becomes foliated).

Figure 12 shows the dependence of the optimal assignment on the value of ϵ_0 in (2.34). A qualitatively new feature appears in the cases of $\epsilon_0 = -0.025$ and $\epsilon_0 = -0.069$: for these unstable equilibria, there exists a range of the initial supported mass coordinate (in the vicinity of $\mu \simeq 0.25$ for the former case and $\mu \simeq 0.4$ for the latter) for which slices that are neighbouring in the initial state are alternately assigned to two different final locations. Like foliation, this phenomenon can be interpreted as a consequence of our seeking a 1D optimisation when, in fact, the true optimal assignment is higher-dimensional in the continuous limit $\Delta m \to 0$. In this case, the optimal assignment involves splitting the fluid at given height in a horizontal sense, and reassigning it to multiple new locations. In Appendix E, we derive a necessary condition for this kind of one-to-many assignment and prove that two is, in fact, the largest possible value of 'many'.

3.3.3. *Bi-directional metastability,* (2.32)

We visualise in figure 13 the optimal assignment for the profile defined by (2.32), which has a local maximum in its profile of s/χ at $z \simeq 1.75$ (see figure 3*c*) and therefore the fluid is metastable both to upward and downward displacements there. It is intuitive, therefore, that the minimum-energy state should be obtained by reassignment of slices from the vicinity of $z \simeq 1.75$ to both the top and the bottom of the region that is at marginal linear stability. This is indeed the case in figure 13 (with foliation between moved and ambient fluid).

Figure 14 shows the optimal assignments for different values of ϵ_0. The available energy associated with this profile is a fraction 6×10^{-4} of the initial total energy for $\epsilon_0 = 0$, which is comparable to the equilibria considered in §§ 3.3.1 and 3.3.2.

3.3.4. *Overturning metastability,* (2.33)

In figure 15, we visualise the optimal assignment for the profile defined by (2.33), which, in contrast to (2.32), has a local minimum in its profile of s/χ at $z \simeq 1.75$ (see figure 3*d*). The fluid at the top of the atmosphere is therefore nonlinearly unstable to downwards motions, while the fluid at the bottom is nonlinearly unstable to upwards motions – we expect therefore that the minimum-energy state will be reached by an 'overturning' of the atmosphere. This is roughly what we observe in figure 15, although the precise optimal assignment is remarkably complex (see figure 16 for a visualisation of the optimal

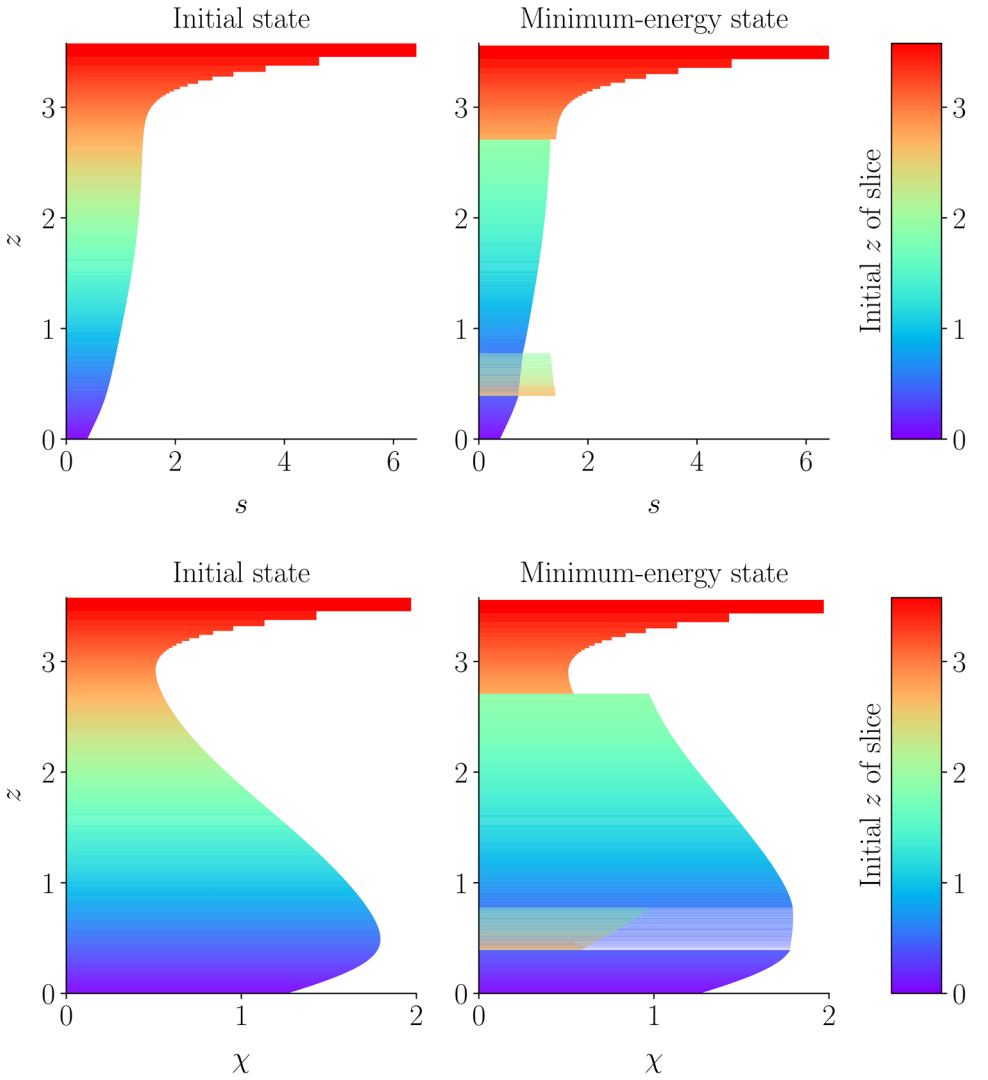

FIGURE 10. Visualisation of the minimum-energy assignment for the equilibrium defined by (2.31), with $\epsilon_0 = 0$ in (2.34). Details are the same as for figure 5.

assignments for different values of ϵ_0). The available energy associated with this profile at $\epsilon_0 = 0$ is 2×10^{-3} of the initial total, which is somewhat more than that of the equilibria considered in §§ 3.3.1–3.3.3.

4. Relaxation without diffusion

In the remainder of this paper, we consider the problem of predicting the state to which a metastable equilibrium relaxes when destabilised. We shall focus on the case where relaxation is complete, i.e. the destabilisation of the initial equilibrium is sufficiently violent to liberate the system from its metastable equilibrium completely. We assume that the system thereafter explores its configuration space freely, only subject to the constraints imposed by conservation laws. We shall assess with numerical simulations whether this is indeed a good assumption in §§ 4.2 and 6.

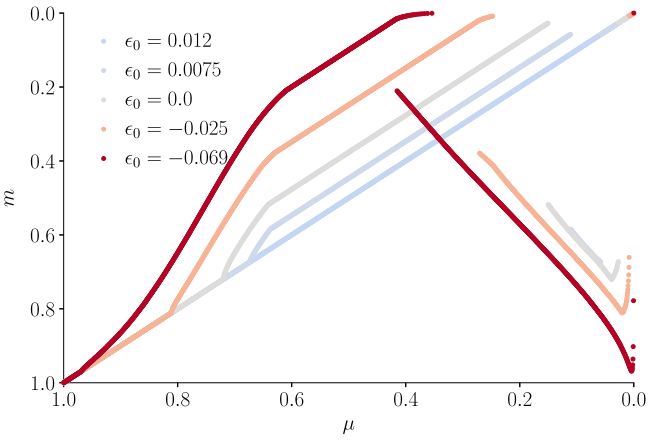

FIGURE 11. Minimum-energy assignments for the profile (2.31). We observe that the optimal assignment is one to two over certain ranges of m_1 in the cases with $\epsilon_0 < 0$.

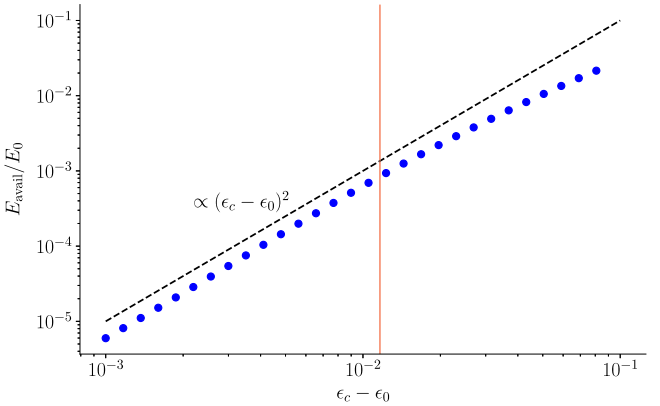

FIGURE 12. Fractional available energy as a function of $\epsilon_c - \epsilon_0$ for the profile (2.31), where $\epsilon_c = 1.2 \times 10^{-2}$ is the largest value of ϵ_0 in (2.34) for which the equilibrium is metastable.

The relevant conservation laws are those of total energy and, to the extent that non-ideal processes (i.e. thermal conduction and resistive and viscous heating; see Appendix A for a statement of the MHD equations including these effects) can be neglected, of s and χ for each fluid element. We shall assume in what follows that viscous heating is negligible because the kinetic energy that develops during relaxation is limited by the available energy of the initial equilibrium, which is small compared with the internal energy (§ 3.1). It follows that the ultimate deposition of kinetic energy as heat does not change the internal energy by very much (even locally). On the other hand, diffusion of s and χ by thermal conduction and resistivity in a well-mixed state may change their values significantly, since s and χ can vary by an order-unity fraction between fluid elements. Two qualitatively different types of relaxation may therefore be distinguished: the one in which s and χ do not diffuse during relaxation and the one in which they do. In this section, we consider the former case, which is realised when turbulent mixing is suppressed by viscosity. We shall argue that relaxation is then to the minimum-energy state calculated in § 3. We address the case of fully turbulent relaxation in § 5.

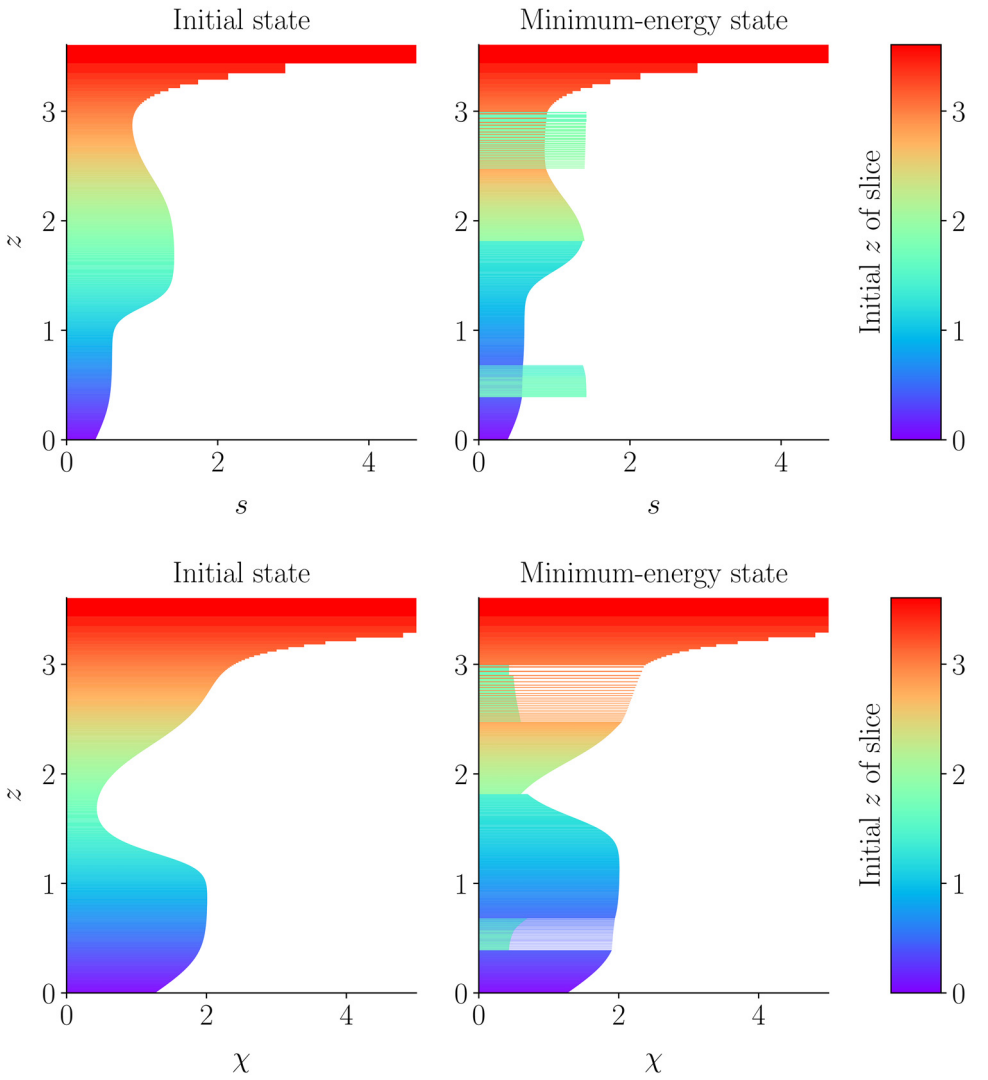

FIGURE 13. Visualisation of the minimum-energy assignment for the equilibrium defined by (2.32), with $\epsilon_0 = 0$ in (2.34). Details are the same as for figure 5.

4.1. *Conditions for diffusion to be absent*

In order for diffusion of s and χ to be negligible during the whole period of relaxation, the flow generated during relaxation must decay before it is able to mix s and χ to scales l such that their diffusion time scales (l^2/K and l^2/η, respectively, where K and η are the thermal and magnetic diffusivities) become comparable to the flow's decay time scale. The latter is H^2/ν (ν is the kinematic viscosity) independently of whether the flow is laminar or turbulent, because 2D turbulence does not cascade energy to smaller scales. A simple regime in which diffusion may be negligible is the one in which the flow is laminar. This requires a Reynolds number $\mathrm{Re} \equiv UH/\nu \lesssim 1$, where U is the characteristic velocity developed during relaxation and H the stratification height, which we take to be the characteristic outer scale of the flow (this being necessary for complete relaxation). By (3.8), $\delta\rho \sim \rho E_{\mathrm{avail}}/E_0$, so the net gravitational force on a fluid element is

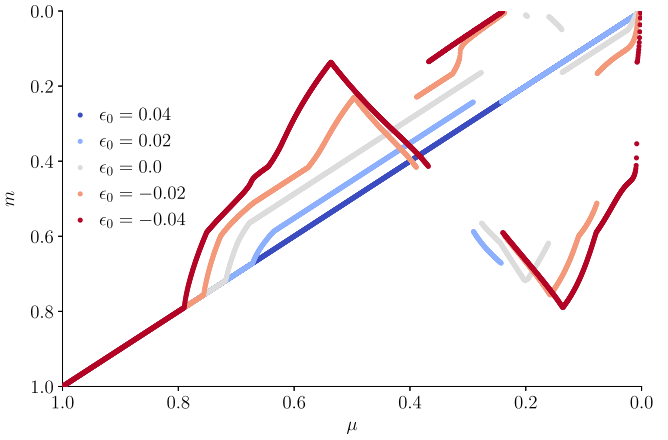

FIGURE 14. Minimum-energy assignments for the profile (2.32) for different values of ϵ_0 in (2.34).

$g\delta\rho \sim \rho c^2 E_{\mathrm{avail}}/HE_0$, where c is the characteristic speed of compressive waves (assumed constant here for simplicity). Balancing this with the viscous force $\rho\nu U/H^2$ to estimate $U \sim Hc^2 E_{\mathrm{avail}}/\nu E_0$, we find that relaxation is laminar if

$$\nu \gtrsim Hc\sqrt{\frac{E_{\mathrm{avail}}}{E_0}}. \tag{4.1}$$

If (4.1) is satisfied, the flow turns over at most once before it decays, so s and χ are not mixed to smaller scales. They are therefore well conserved provided that

$$\mathrm{Pr}_t,\ \mathrm{Pr}_m \gg 1, \tag{4.2}$$

where $\mathrm{Pr}_t \equiv \nu/K$ and $\mathrm{Pr}_m \equiv \nu/\eta$ are the thermal and magnetic Prandtl numbers, respectively.

For $\mathrm{Re} \gg 1$, the outer-scale flow has velocity $U \sim (E_{\mathrm{avail}}/E_0)^{1/2}c$ and is turbulent. It turns over Re times before decaying (in two dimensions), so s and χ are mixed to the scale $l \sim H\exp(-\mathrm{Re})$. Diffusion at this scale can be neglected if its time scale is longer than the decay time of the turbulence, i.e. if

$$\ln \mathrm{Pr}_t,\ \ln \mathrm{Pr}_m \gg \mathrm{Re} \sim \frac{cH}{\nu}\sqrt{\frac{E_{\mathrm{avail}}}{E_0}}. \tag{4.3}$$

In addition to mixing by the flow at scale H, Rayleigh–Taylor instability at interfaces of fluid with different densities may generate small-scale vortices that mix the fluid (this effect was evident in the $t = 70$ panel of figure 2). The fastest-growing Rayleigh–Taylor mode, which develops at scale L_{RT}, is limited by viscosity: its growth rate is $\gamma_{\mathrm{RT}} \sim (gL_{\mathrm{RT}}\delta\rho/\rho)^{1/2} \sim \nu/L_{\mathrm{RT}}^2$, whence $L_{\mathrm{RT}} \sim \mathrm{Re}^{-2/3}H$ and $\gamma_{\mathrm{RT}} \sim \mathrm{Re}^{1/3}U/H$. Taking the nonlinear turnover rate of the developed vortex to be equal to γ_{RT}, this mode turns over $\mathrm{Re}^{4/3}$ times before the outer-scale turbulence decays (at which point we assume that no Rayleigh–Taylor-unstable interfaces remain). An argument similar to the one that led to (4.1) indicates that the Rayleigh–Taylor vortices will not establish diffusion-scale structure

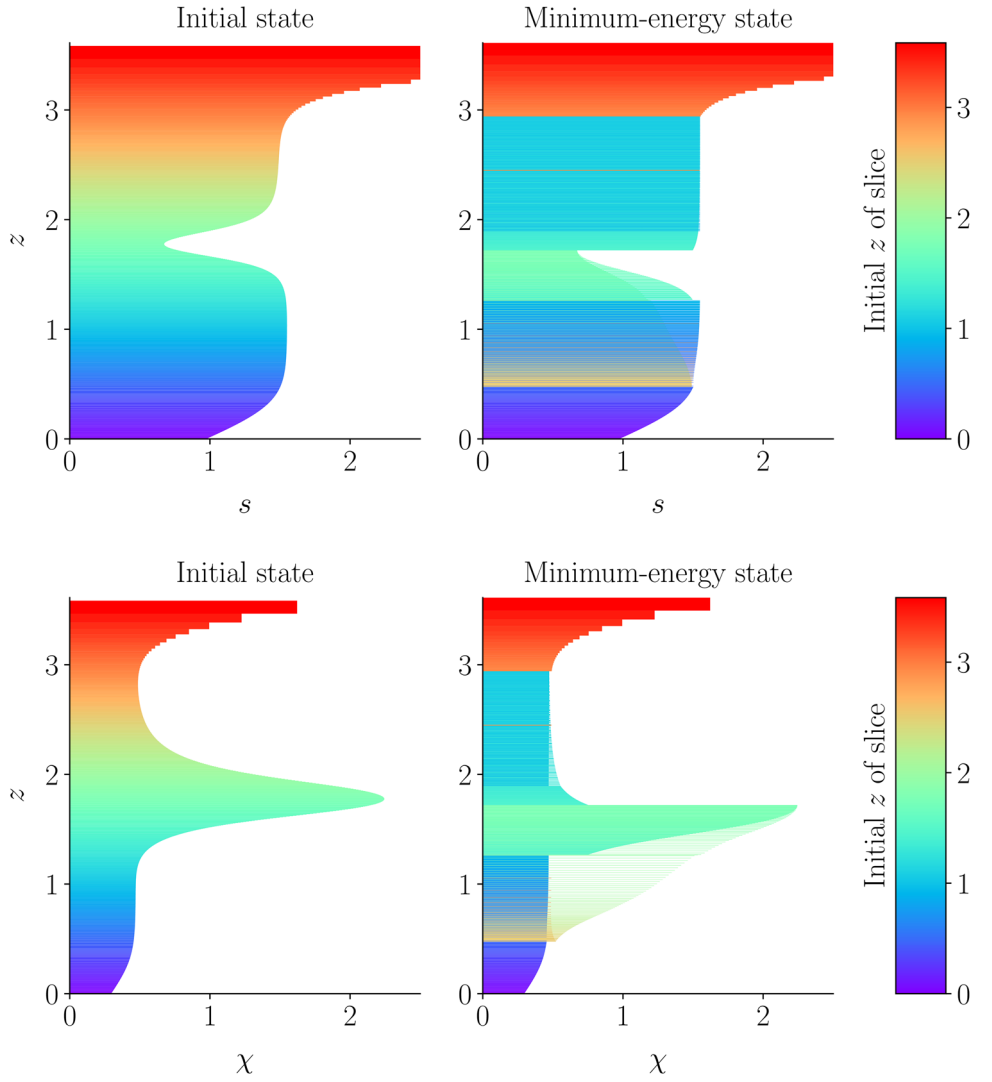

FIGURE 15. Visualisation of the minimum-energy assignment for the equilibrium defined by (2.33), with $\epsilon_0 = 0$ in (2.34). Details are the same as for figure 5.

in s and χ provided that

$$\ln \mathrm{Pr}_t,\ \ln \mathrm{Pr}_m \gg \mathrm{Re}^{4/3} \sim \left(\frac{cH}{\nu}\right)^{4/3} \left(\frac{E_{\mathrm{avail}}}{E_0}\right)^{2/3}. \tag{4.4}$$

Equation (4.4) is a stricter criterion than (4.3); because of their faster turnover rate than the flow at scale H, the Rayleigh–Taylor-generated vortices are more effective at mixing.

If (4.4) is satisfied [or if (4.1) and (4.2) are, for the case of $\mathrm{Re} \lesssim 1$], the relaxation flow decays before s and χ diffuse via thermal conduction or ohmic heating. Provided that the initial destabilisation was sufficiently thorough (so that the system does not become trapped in a new metastable state), we expect the final static state of the system to be the one with smallest potential energy subject to fluid-element-wise conservation of s and χ.

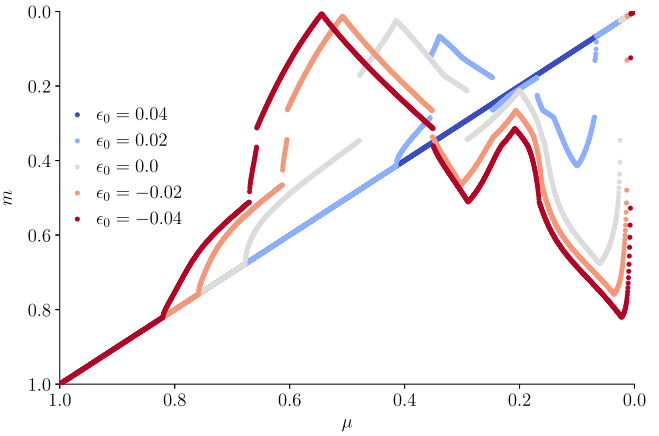

FIGURE 16. Minimum-energy assignments for the profile (2.33) for different values of ϵ_0 in (2.34).

Because the amount of heat per unit mass generated by viscosity is $U^2 \sim c^2 E_{\mathrm{avail}}/E_0 \ll c^2$, the fractional change in s of each fluid element due to viscous heating during relaxation is small. Thus, the final state reached by the system ought to be, to first approximation, the states of minimum energy calculated in § 3. In the next section, we verify with numerical simulations that this is indeed the case.

4.2. *Numerical results*

Figure 17 visualises the relaxation of the equilibrium defined by (2.30) after the application of an impulsive force that accelerates the fluid to a velocity

$$\boldsymbol{u} = u_0 \hat{\boldsymbol{z}} \sin\left(\frac{2\pi x}{L_x}\right) \exp\left(-\frac{(z - z_0)^2}{\Delta z^2}\right), \tag{4.5}$$

where $L_x = 2$ is the size of the simulation domain in the x direction, $u_0 = 0.1$, $z_0 = 1.0$ and $\Delta z = 0.5$ (see § 2.6 for an explanation of our system of units). This corresponds to an initial kinetic energy $E_{\mathrm{kin},0} \simeq 0.3 E_{\mathrm{avail}}$, where the available energy $E_{\mathrm{avail}} \sim 10^{-3} E_0$. The kinematic viscosity is 1.6×10^{-3} in these units, so the Reynolds number of the flow at the initial time is $\mathrm{Re} \sim u_0 L_x/\nu \sim 10^2$. The magnetic and thermal Prandtl numbers are $\mathrm{Pr}_m = 400$ and $\mathrm{Pr}_t = 670$, respectively. We provide further details of the numerical set-up in Appendix A.

Although $\mathrm{Re} \sim 100 > 1$, figure 17 shows that Re is insufficiently large for turbulence to develop, either at the outer scale or driven by Rayleigh–Taylor instability (which does nonetheless lead to the development of structure at scales smaller than $L_x \sim H$). Thus, relaxation takes place without significant diffusion of s and χ because the Prandtl numbers are large (4.2). We observe that the upwards plume generated by the initial impulse forms a long-lived 2D state (upper panels of figure 17), which is indeed consistent with the minimum-energy state obtained in § 3.3.1 (see lower panels of figure 17).

In order to assess the sensitivity of the final state to the initial perturbation, we visualise in figure 18 the late-time state developed by simulations identical to the one shown in figure 17, but with different values of the initial kinetic energy. Specifically, we choose $u_0 = 0.05$ (centre panel) and $u_0 = 0.025$ (left panel), so that $E_{\mathrm{kin},0} \simeq 0.1 E_{\mathrm{avail}}$ and $E_{\mathrm{kin},0} \simeq 0.02 E_{\mathrm{avail}}$, respectively. We observe that there is some sensitivity to the amplitude of the initial perturbation: somewhat more material is displaced upwards at $E_{\mathrm{kin},0} \simeq 0.3 E_{\mathrm{avail}}$ than

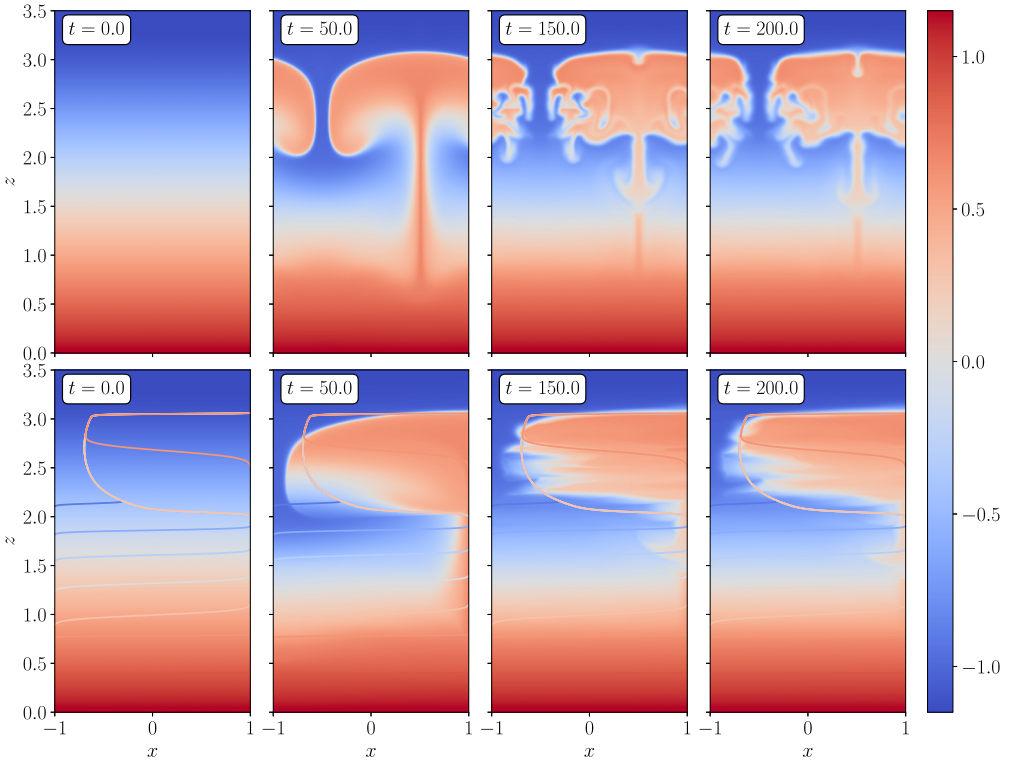

FIGURE 17. The relaxation of the equilibrium defined by (2.30) at Re $\sim 10^2$. The initial velocity field is given by (4.5). Upper panels show the evolution of $\ln(s/\chi)$ in x–z space. Lower panels show the same quantity but sorted horizontally at each z, with contours of the theoretical minimum-energy state (figure 7*d*) overlaid.

$E_{\mathrm{kin},0} \simeq 0.02E_{\mathrm{avail}}$. This weak, but measurable, dependence of the final state on initial conditions is despite the fact that the equilibrium is initially at marginal linear stability, so there is no potential barrier to be overcome in order to trigger instability. However, partial relaxation stabilises the atmosphere (see Appendix D), so that, while the first magnetic-flux tube to move upwards experiences no potential barrier, later ones do.

In Appendix F, we present analogous simulations of viscous relaxation for the equilibrium defined by (2.31) (i.e. metastable to downwards perturbations). The results are qualitatively similar to those presented in this section.

5. Statistical theory of relaxation at large Reynolds number

In this section, we consider relaxation for which the Reynolds number is sufficiently large that inequality (4.4) no longer holds. In this case, turbulent mixing generates sufficiently fine-scale structure in s and χ to enable diffusion. Consequently, the equilibrium reached once the velocity field has decayed cannot be obtained by an ideal rearrangement of the initial state and so will, in general, differ from the minimum-energy states discussed in § 3.

Given that diffusion is to occur, the key question is 'which fluid parcels are brought into contact by the turbulent flow?' If we could identify in advance which fluid parcels were to diffuse with which others and thus become locally homogenised, then we could predict their new values of the ideal Lagrangian invariants s and χ from the conservation of net

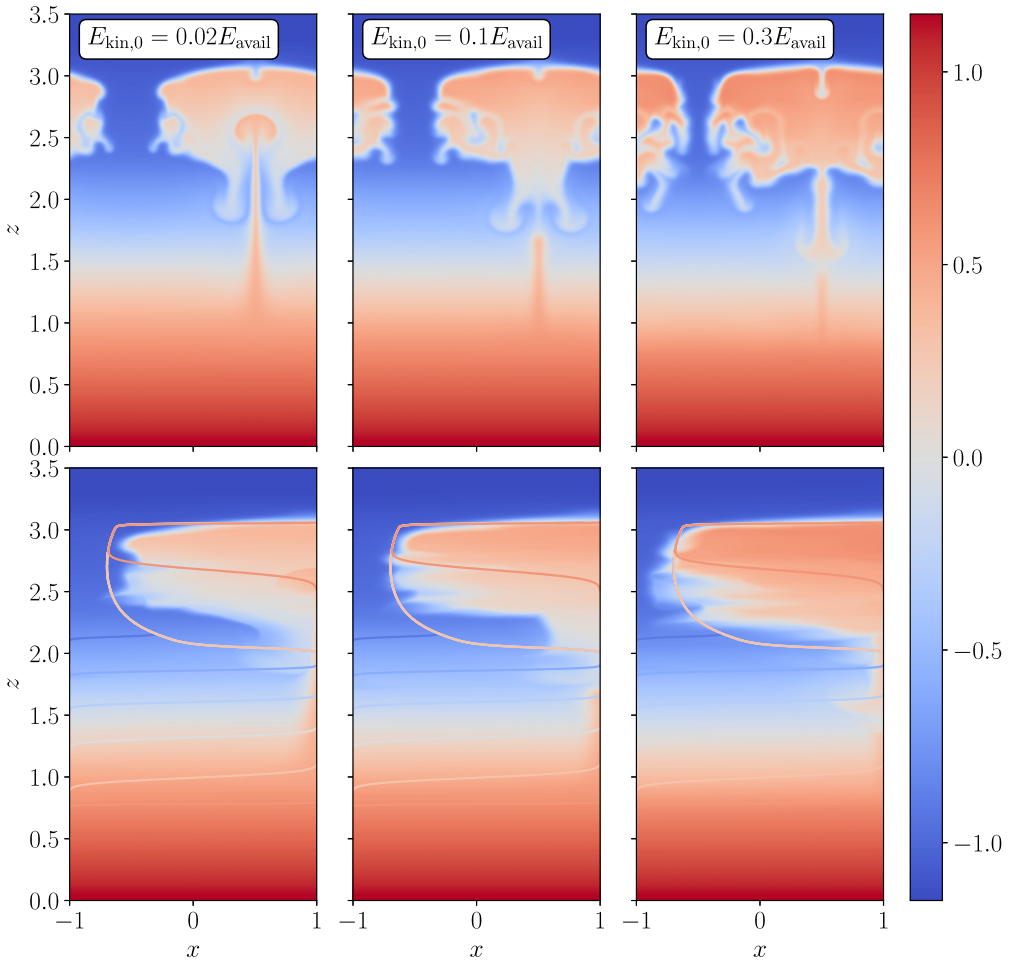

FIGURE 18. The distribution of $\ln(s/\chi)$ at $t = 300$ for simulations analogous to the one visualised in figure 17 but for three different values of u_0. As in figure 17, upper panels visualise the state of the simulation, while lower panels are sorted horizontally and overlaid with the 2D minimum-energy state (figure 7*d*).

magnetic flux and enthalpy during diffusion. The relaxed state would then be the one with minimum energy subject to ideal rearrangements of this post-diffusion state.

Motivated by the chaotic nature of turbulent mixing, we shall treat this problem probabilistically, i.e. with statistical mechanics. We assume that the time scale for thorough mixing is much shorter than that for the onset of diffusion, so that these processes can be treated separately. Specifically, we assume that, before diffusion acts, turbulent mixing causes the system to explore all possible 2D distributions of s and χ (microstates) consistent with its potential energy.[10] We seek the probability-distribution function (the macrostate) for fluid at a given spatial position to have originated from a different given position. We obtain it by maximising the number of microstates for which it gives correct

[10]For simplicity, we take the potential energy to be constant and equal to the total initial energy of the system (where we make comparison with numerical simulations in § 6, this includes the kinetic energy of the perturbation).

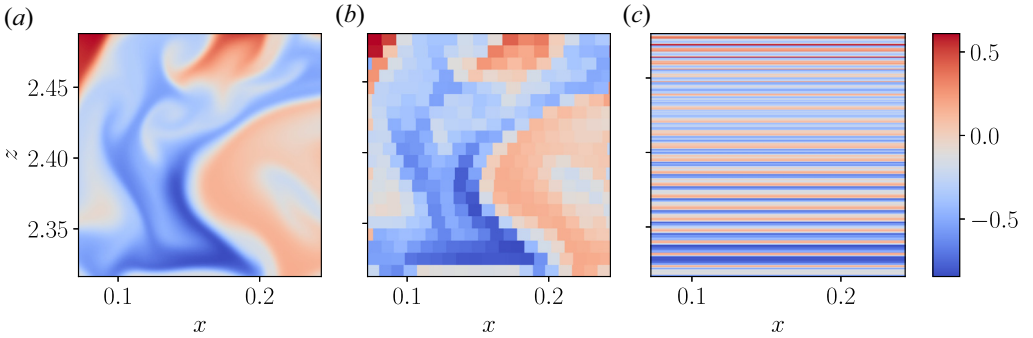

FIGURE 19. At Re $\gg 1$, turbulence mixes the advected scalars s and χ [(*a*); this is a subsection of the state visualised in the $t = 70$ panel of figure 23]. Panel (*b*) is a copy of (*a*) but with the fluid discretised into parcels of equal mass. The 2D non-equilibrium states obtainable by shuffling these parcels while maintaining horizontal pressure balance constitute microstates in our theory. As explained in the main text, there exists a bijection between such 2D states and 1D static equilibria with the same energy (*c*). We may therefore take microstates to be 1D equilibria.

coarse-grained statistics, i.e. by maximising mixing entropy. Finally, we shall determine the diffused states by taking moments of this distribution, as we explain in § 5.3.

5.1. *Equivalence of 1D and 2D microstates*

Even through the relaxing system explores 2D non-equilibrium microstates, maximising the number of them that are consistent with a given macrostate (i.e. the multiplicity of the macrostate) is formally equivalent to doing so for the equilibria obtainable by 1D rearrangements (similar to those we used to find minimum-energy states in § 3). This is because, for every 2D non-equilibrium state in horizontal pressure balance, there exists a distinct 1D equilibrium state with the same total energy and mass of fluid with each value of s and χ, as we now demonstrate.

We define the set of 2D microstates as follows. We partition the initial equilibrium into horizontal slices of fixed width Δz, which we further subdivide into parcels of equal mass Δm. The number of parcels in each horizontal slice is not fixed, and parcels may 'spill' over into adjacent slices if the total mass in the slice is not an integer multiple of Δm, although the fraction of parcels that do vanishes as $\Delta m \to 0$. We consider labelling each parcel by an index that increases along each slice from left to right, starting from the lowest slice and then moving upwards. Then, the full set of microstates is the set of possible permutations of these indices, i.e. the set of 1D shuffles of parcels with fixed s and χ, assembled in 2D space as described (see figure 19*a*,*b*).

A given permutation of parcel indices is not a complete description of the microstate, as the pressure P in each parcel remains to be specified. As noted above, the 2D microstates are not equilibria: horizontal density variations induce baroclinic torques. On the other hand, we do not expect significant horizontal variation in pressure: because $E_{\mathrm{avail}}/E_0 \ll 1$, the flow that develops during relaxation is subsonic, i.e. $U/c \ll 1$. Fluctuations δP of the total pressure P about its horizontal mean P_0 are therefore small, $\delta P/P_0 \sim U^2/c^2 \ll 1$. We can evaluate $P_0(z)$ by integrating the z-component of the momentum equation (2.1) over x and from the given z to $z = \infty$, neglecting inertial terms (which correspond to the pressure fluctuations). This yields $P_0 = mg$, where m is the total mass of fluid above height z per

unit horizontal length, given by

$$m = \frac{1}{L_x}\int_0^{L_x} \mathrm{d}x \int_z^{\infty} \mathrm{d}z' \rho(x, z'), \tag{5.1}$$

where L_x is the horizontal extent of the system.

To leading order in Δz, (5.1) evaluated at a given parcel is the total mass of parcels that have greater indices. Consequently, if we were to rearrange the parcels to be stacked vertically in order of their indices (preserving the P, s and χ from the 2D state), P_0 is the pressure they would have in equilibrium. Restacking in this way produces no change in the total energy as Δz, $\Delta m \to 0$, so the total energy of both states can be evaluated using (3.4). The leading-order term in δP in the integrand is $\sim c^2 \delta P^2 / P_0^2 \sim (U^2/c^2)U^2 \ll U^2$, so the contribution of δP to the energy of the state can be neglected. Thus, there exists a correspondence between 2D non-equilibrium (but horizontally pressure-balanced) states and 1D equilibria: for every 2D state, we can find a 1D equilibrium state with the same energy, by the process described above (visualised in figure 19*c*). Thus, in what follows, we consider microstates to be 1D equilibrium states, with energy given by (3.6).

5.2. *Lynden-Bell statistical mechanics of MHD atmospheres*

As explained in § 5.1, we may construct our statistical mechanics on the space of 1D equilibria, this being fully equivalent to doing so on the space of 2D non-equilibria in horizontal pressure balance. We follow the formulation of Lynden-Bell's (1967) statistical mechanics of distinguishable particles with an exclusion principle – originally derived for collisionless stellar systems and plasma – by Chavanis (2003).

We introduce $\mathcal{P}(m, \mu)\,\mathrm{d}\mu$ as the probability of finding the material with initial supported mass in the range $[\mu, \mu + \mathrm{d}\mu]$ to have a supported mass of m in the final state. We obtain $\mathcal{P}(m, \mu)$ by maximising the number of microstates with which it is consistent after coarse graining. This corresponds to maximising the mixing entropy (Robert & Sommeria 1991)

$$S = -\int \mathrm{d}m \int \mathrm{d}\mu \mathcal{P}(m, \mu) \ln \mathcal{P}(m, \mu). \tag{5.2}$$

We maximise S subject to the constraints of fixed total probability (i.e. the normalisation of P)

$$\int \mathrm{d}\mu \mathcal{P}(m, \mu) = 1, \quad \forall\, m; \tag{5.3}$$

fixed potential energy E_{pot}

$$\int \mathrm{d}m \int \mathrm{d}\mu \mathcal{E}(m, \mu)\mathcal{P}(m, \mu) = E_{\mathrm{pot}}; \tag{5.4}$$

and fixed mass of fluid with each value of μ

$$\int \mathrm{d}m \mathcal{P}(m, \mu) = 1, \quad \forall\, \mu. \tag{5.5}$$

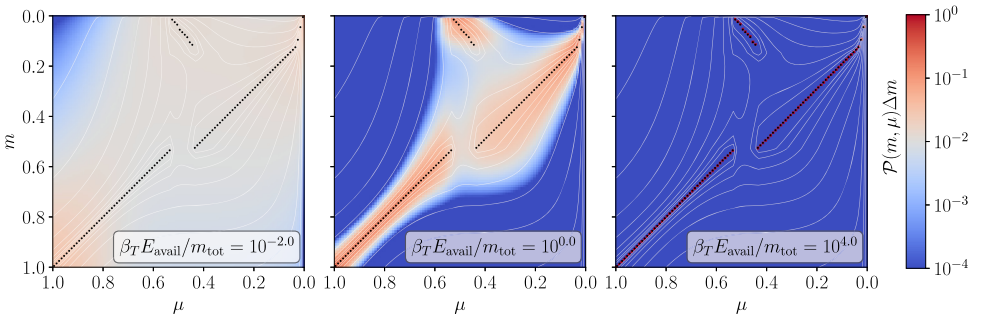

FIGURE 20. Convergence of $\mathcal{P}(m, \mu)$ ((5.7)) to the solution of the LSA problem (black circles; see § 3.2) as $\beta_T E_{\rm avail} \to \infty$ for the case of the unstable-upwards profile (2.30) with all integrals discretised at scale $\Delta m = 0.01$. Contours of the normal from of the cost matrix $\tilde{\mathcal{E}}_{ij}$ visualised in figure 6 are plotted in white.

This constrained maximisation of S is equivalent to unconstrained maximisation of

$$-\int \mathrm{d}m \int \mathrm{d}\mu \mathcal{P}(m,\mu) \ln \mathcal{P}(m,\mu) - \beta_T \int \mathrm{d}m \lambda(m) \left[\int \mathrm{d}\mu \mathcal{P}(m,\mu) - 1\right] - \beta_T \left[\int \mathrm{d}m \int \mathrm{d}\mu\, \mathcal{E}(m,\mu)\mathcal{P}(m,\mu) - E_{\rm pot}\right] - \beta_T \int \mathrm{d}\mu \psi(\mu) \left[\int \mathrm{d}m \mathcal{P}(m,\mu) - 1\right], \tag{5.6}$$

over the probability $\mathcal{P}(m, \mu)$ and the Lagrange multipliers β_T, $\lambda(m)$ and $\psi(\mu)$. The solution is

$$\mathcal{P}(m,\mu) = \mathrm{e}^{-\beta_T[\mathcal{E}(m,\mu) - \psi(\mu) - \lambda(m)]}, \tag{5.7}$$

where the Lagrange multipliers β_T (the thermodynamic beta, to be identified with the inverse of the statistical mechanical temperature), $\psi(\mu)$ and $\lambda(m)$ are determined from the constraints (5.3)–(5.5).[11]

In Appendix H, we prove that, as $\beta_T \to \infty$, $\mathcal{P}(m, \mu)$ becomes increasingly sharply peaked around the solution to the LSA problem of § 3 (see figure 20). Thus, the minimum-energy states are the $\beta_T \to \infty$ limit of our statistical mechanics. Mathematically, this happens because the exponent of (5.7) has the form of a modified cost matrix (see (3.11)) multiplied by β_T; as $\beta_T \to \infty$, $\mathcal{P}(m, \mu)$ vanishes except in the vicinity of the zeros of the normal form of the cost matrix (see § 3.2), which represent the optimal assignment.

The algorithm we use to evaluate $\mathcal{P}(m, \mu)$ numerically for given $E_{\rm pot}$ is analogous to the one proposed by Ewart, Nastac & Schekochihin (2023). A brief summary is as follows.

[11]Let us make explicit the analogy between these formulae and their equivalents in the Lynden-Bell theory of collisionless stellar systems and plasma. In those contexts, $\mathcal{P}(m, \mu)$ is replaced by $\mathcal{P}(\boldsymbol{x}, \boldsymbol{v}, \eta)$, where position $\boldsymbol{x}$ and velocity $\boldsymbol{v}$ are the phase space coordinates (analogous to m) and η is the phase-space density, which, analogously to μ, is conserved under rearrangements of phase space (Liouville's theorem). The energy density $\mathcal{E}(m, \mu)$ is replaced by the energy associated with η particles occupying the $(\boldsymbol{x}, \boldsymbol{v})$ coordinates of phase space, $\eta[v^2/2 + \Phi(\boldsymbol{x})]$ (or appropriate generalisations), where Φ is potential energy. Finally, on the right-hand side of the constraint (5.5), 1 is replaced by a function of η, sometimes called the 'waterbag content', which gives the total volume of phase space with density η. The equivalent object is a constant in our formalism because the mass of fluid in the range $[\mu, \mu + \mathrm{d}\mu]$ is $\mathrm{d}\mu$, independently of μ. In Appendix G we present an alternative formulation of our statistical mechanics, with μ replaced by s and χ, which is also preserved under rearrangement. In that formulation, a function $M(s, \chi)$ appears on the right-hand side of the equation analogous to (5.5); this is the total mass of fluid with given s and χ.

First, we choose a trial value of β_T and, with suitable discretisation for m and μ, calculate

$$\mathcal{P}(m,\mu) = \frac{\mathrm{e}^{-\beta_T[\mathcal{E}(m,\mu)-\psi(\mu)]}}{\int \mathrm{d}\mu'\,\mathrm{e}^{-\beta_T[\mathcal{E}(m,\mu')-\psi(\mu')]}}, \tag{5.8}$$

where we obtain $\psi(\mu)$ from the iterative formula

$$\mathrm{e}^{-\beta_T\psi_{n+1}(\mu)} = \int \mathrm{d}m \frac{\mathrm{e}^{-\beta_T\mathcal{E}(m,\mu)}}{\int \mathrm{d}\mu'\,\mathrm{e}^{-\beta_T[\mathcal{E}(m,\mu')-\psi_n(\mu')]}}. \tag{5.9}$$

Equation (5.9) is the result of integrating (5.8) over m and using (5.5). Equation (5.8) satisfies the constraints (5.5) and (5.3), but does not necessarily correspond to the correct energy E_{pot}; in this case, we increment β_T and repeat the procedure described until the desired energy is obtained.

5.3. *Diffusion*

The function $\mathcal{P}(m,\mu)$ gives the fractional abundances of parcels from supported mass μ at new supported mass m in the 'most mixed' state accessible by ideal rearrangements. We shall use these abundances to determine the result of diffusion: when sufficiently small scales are developed by mixing, diffusion homogenises nearby fluid parcels. Thus, diffused states may be obtained by taking suitable moments of $\mathcal{P}(m,\mu)$.

The standard method for deriving predictions from Lynden-Bell probability distribution functions is to argue that, due to the presumed stochastic nature of the underlying microstate, physically measurable quantities correspond to expectation values. In our case, these are

$$\langle s\rangle \equiv \int \mathrm{d}\mu s(\mu)\mathcal{P}(m,\mu), \tag{5.10}$$

$$\langle \chi\rangle \equiv \int \mathrm{d}\mu \chi(\mu)\mathcal{P}(m,\mu). \tag{5.11}$$

Unlike in the traditional contexts, however, expectation values – in particular, (5.10) – are unsuitable as a model of the state that develops after diffusion acts. This is because diffusive processes do not preserve the mean value of s; instead, thermal conduction and ohmic heating increase thermal entropy.[12] It is readily verified that energy is not conserved under a collapse of the distributions of s and χ onto their expectation values

$$\int \mathrm{d}m\mathcal{E}(mg,\langle s\rangle,\langle\chi\rangle) \neq E_{\mathrm{pot}}, \tag{5.12}$$

because $\mathcal{E}(m,\mu)$ is nonlinear in s and χ. In the terminology of Chavanis (2003), energy is a 'fragile integral' – its value is different depending on whether it is computed using

[12]The relevant analogue of (5.10) and (5.11) for the collisionless-relaxation problem, i.e. $\langle\eta\rangle = \int \mathrm{d}\eta\,\eta P(\boldsymbol{x},\boldsymbol{v},\eta)$, does constitute a plausible prediction for the particle-distribution function in the presence of small collision frequency (collisions act as diffusion in velocity space, smoothing the stochastic variation in the local value of η to its mean). This is because collisions preserve the contribution of each patch of phase space to the total energy, momentum and number of particles, as these quantities are each linear in η. In the language of Chavanis (2003), total energy, momentum and number of particles are 'robust integrals', being the same for both the coarse- and fine-grained distributions of η (although, see the discussion in § 7.2).

the coarse- or fine-grained distributions of s and χ. The difference between the left- and right-hand sides of (5.12) is typically much greater than the available energy of the original equilibrium – around ten times greater in the case of the unstable-upwards profile defined by (2.30).

Equation (5.10) being unsuitable, we instead determine the entropy function after diffusion, $\bar{s}$, from the conservation of energy, i.e. from

$$\mathcal{E}(mg, \bar{s}, \bar{\chi}) = \int \mathrm{d}\mu \mathcal{E}(mg, s(\mu), \chi(\mu)) \mathcal{P}(m, \mu), \tag{5.13}$$

with the diffused magnetic flux $\bar{\chi}$ given by

$$\bar{\chi} = \langle \chi \rangle; \tag{5.14}$$

(the diffusion of straight magnetic-field lines does conserve magnetic flux). We plot $\bar{s}$ and $\bar{\chi}$ against z (with density $\rho(m, \bar{s}, \bar{\chi})$) with a solid cyan line in figure 21. For comparison, we plot $\langle s \rangle$ and $\langle \chi \rangle$ against z (with density $\rho(m, \langle s \rangle, \langle \chi \rangle)$) with a gold dashed line – these profiles are appreciably different, and we will find in § 6 that the former is indeed a better predictor of numerical simulations.

5.4. *Secondary relaxation*

Equations (5.13) and (5.14) need not, in general, represent the final state reached by relaxation, because the turbulent mixing flow remains present even after diffusion. This flow can cause further reorganisation if the post-diffusion state of the system has more than one accessible microstate, i.e. if it is not a nonlinearly stable minimum-energy state.[13] Remarkably, this often turns out to be the case: diffusion [in the sense of (5.13) and (5.14)] of the state predicted by Lynden-Bell statistical mechanics tends to produce new states that are unstable to further (ideal) dynamics. This phenomenon has no analogue in the relaxation of collisionless stellar systems and plasma, for which the coarse-grained probability-distribution function is always a state of minimum energy with respect to rearrangements of phase space. That it is possible for MHD atmospheres is consequence of the fact that diffusion produces changes in buoyancy. A well-known example of this is the phenomenon of 'buoyancy reversal' in the terrestrial atmosphere: the nonlinear dependence of density on the advected Lagrangian invariants (in that context, the mixing ratio of water and potential temperature; see Appendix B) means that a buoyant parcel of fluid that rises and mixes with denser ambient fluid can become denser than the ambient fluid, and sink as a result (see, e.g. Stevens 2005).

The 1D equilibrium states given by (5.13) and (5.14) may be linearly or nonlinearly unstable (metastable). Examples of each case are illustrated in figure 22. Panel (*a*) plots the force (2.5) per unit mass on a small parcel of fluid displaced from height z_1 to z_2 for the post-diffusion state of the metastable-upwards equilibrium (2.30). This state is linearly unstable between $z \simeq 1.4$ and $z \simeq 2.3$, and also close to $z = 3$. Panel (*b*) is analogous, but for a slightly larger energy of $E = E_0 + 0.3E_{\mathrm{avail}}$ in (5.4) (this is the initial energy of the numerical simulation to be presented in figures 23 and 24 in § 6.1). In this case, the new profile is linearly stable, but is unstable nonlinearly (metastable), and turns out to have states with lower energy (as can be confirmed by solving its LSA problem). Likewise, the post-diffusion states corresponding to the 'bi-directional' ((2.32)) and 'overturning' ((2.33)) profiles are not minimum-energy states. On the other hand, it turns out that

[13] Because we assume that the energy of the flow is small, we neglect the possibility of it exciting the system into states with greater potential energy.

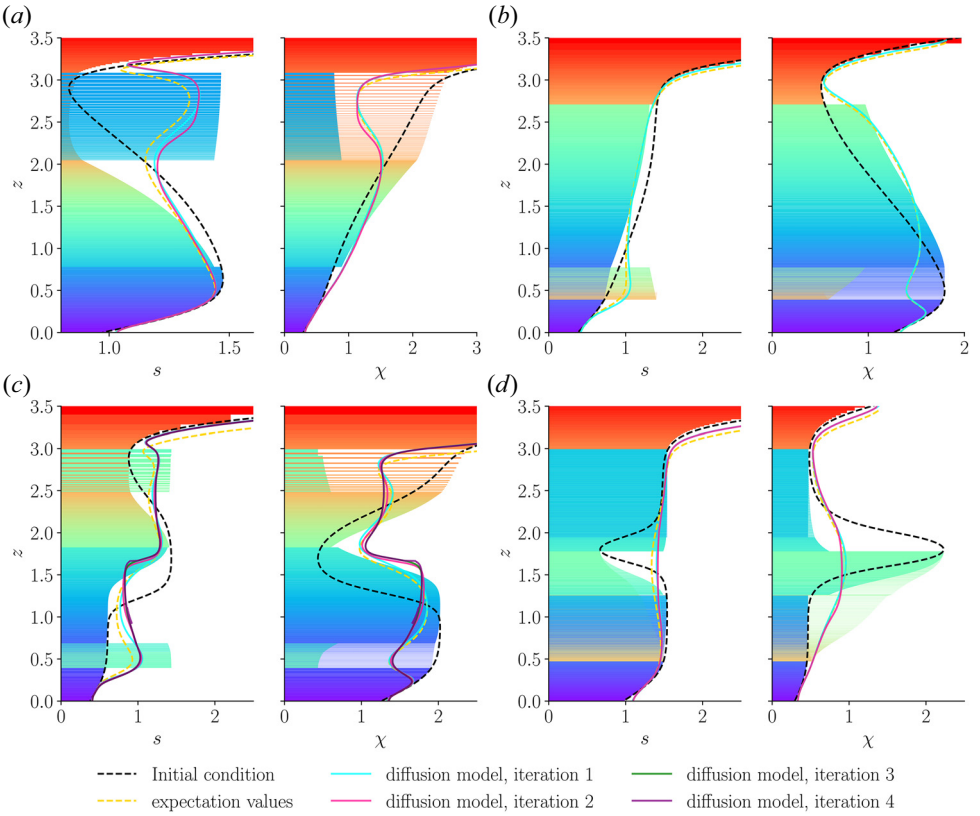

FIGURE 21. The predictions of the Lynden-Bell statistical mechanics (§ 5.2) for each of the profiles described in § 2.6. Panel (*a*) corresponds to (2.30), (*b*) to (2.31), (*c*) to (2.32) and (*d*) to (2.33). In each case, the black dashed line corresponds to the initial profile, the gold dashed line to $\langle s \rangle$ or $\langle \chi \rangle$ [see (5.10) and (5.11)] and the cyan solid line to the predictions $\bar{s}$ and $\bar{\chi}$ for the result of diffusion [see (5.13) and (5.14)]. Other coloured lines correspond to iterations of the statistical mechanical calculation, as described in § 5.4.

diffusion does yield a minimum-energy state in the case of the unstable-downwards profile (2.31).

As noted above, in cases where the new state given by (5.13) and (5.14) has more than one accessible microstate with the same energy, continued turbulent mixing can reshuffle fluid parcels and produce further diffusion until a minimum-energy state is reached. A simple model of this process is to apply the procedure described in §§ 5.2 and 5.3 iteratively. This corresponds to the diffused system exploring the full space of states that are energetically accessible under ideal arrangements before diffusing again. We show in figure 21 the profiles of $\bar{s}$ and $\bar{\chi}$ at the second, third and fourth iterations, plotted in pink, green and purple, respectively. The procedure terminates (i.e. reaches a minimum-energy state) after two iterations in the unstable-upwards (2.30) and 'overturning' (2.32) cases [see panels (*a*) and (*d*)], and after four iterations in the unstable-'bi-directional' case (2.33). [Note that, in the 'overturning' case, we find at the fourth iteration that the state is sufficiently close to the ground state to become sensitive to the discretisation that we employ to compute the integrals in (5.11) and (5.13): see the jagged structure of the purple

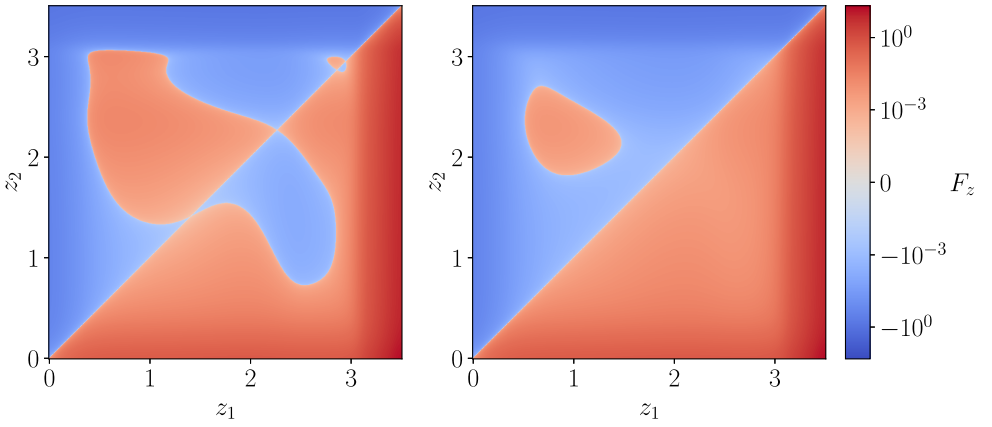

FIGURE 22. The force (2.5) per unit mass of a small fluid parcel moved in pressure balance and without diffusion from height z_1 to z_2 for: left panel, the state that corresponds to the cyan lines in figure 21*a*; and, right panel, the analogue of this profile for $E = E_0 + 0.3E_{\text{avail}}$. The left panel exhibits linear instability, the right panel nonlinear instability (metastability).

line in figure 21*c*. At the fifth iteration, β_T becomes so large as to preclude accurate computation of the relevant integrals, so we terminate the process at the fourth iteration.]

The difference in the profiles of $\bar{s}$ and $\bar{\chi}$ between the first and last stages of the iterative procedure turn out to be slight. Therefore, if, as in reality, secondary relaxations are incomplete or diffusion occurs concurrently with them, the final state reached ought not to be very different. We shall see in § 6 that accounting for these diffusive rearrangements is, in practice, a precision overkill – greater discrepancies between numerical experiment and the theoretical prediction arise which appear to be a result of the tendency of relaxing profiles to become 'stuck' in other metastable states. Nonetheless, it is interesting to note that, on a qualitative level, the chief outcome of the iterative procedure is the formation of plateaus (corresponding to thorough mixing and diffusion in regions where the first relaxed state is close to marginal linear stability). In the case of the unstable-upwards profile, for example, we see from figure 21*a* that, between the first and second iterations, material at $z \lesssim 1.75$ moves upwards to settle in the range $1.75 \lesssim z \lesssim 2.75$, producing a flatter region between $2.5 \lesssim z \lesssim 3.0$ and in the vicinity of $z \simeq 2$. Similar plateaus are observed in panels (*c*) and (*d*), with panel (*c*) resembling a staircase. The intriguing possibility of modelling the formation of staircases (which are observed in myriad diffusing systems in geo- and astro-physical contexts) statistical mechanically is a topic to which we shall return in future work (see § 7.2).

6. Numerical simulations of relaxation at large Reynolds number

In this section, we present a comparison of the theoretical predictions obtained in § 5.3 with the results of numerical simulations with $\text{Re} \gg 1$. The numerical set-up is the same as in § 4.2, but with the kinematic viscosity ν smaller by a factor of 400, such that the Reynolds number based on the initial velocity field is $u_0 L_x/\nu \sim 10^5$.

6.1. *Metastable upwards,* (2.30)

We first consider the case of the unstable-upwards profile defined by (2.30). Figure 23 visualises the distribution of s/χ (the advected scalar that controls the fluid compressibility – see § 2) during the relaxation that follows perturbation by a velocity field given by (4.5) with $u_0 = 0.1$ and $z_0 = 1.0$ (this corresponds to an initial kinetic energy

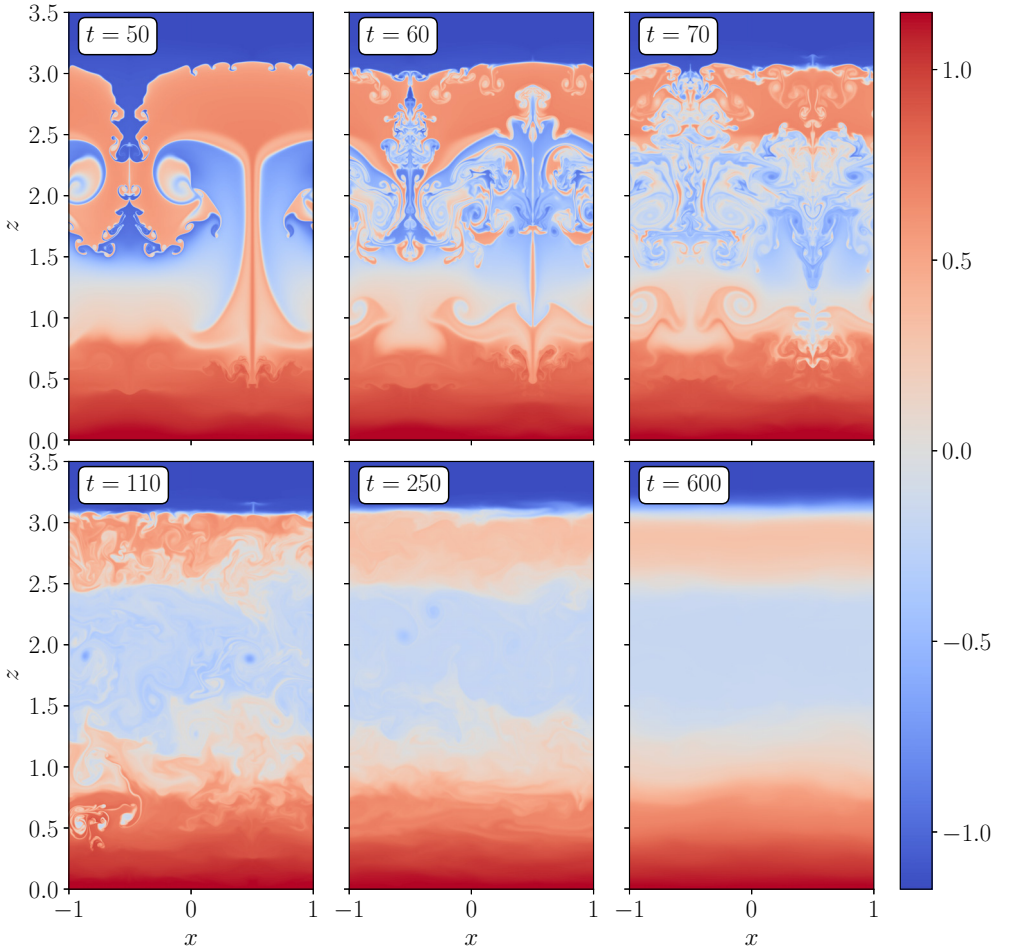

FIGURE 23. Numerical simulation of the relaxation of the equilibrium defined by (2.30) at $\mathrm{Re} \sim 10^5$. The quantity plotted is $\ln(s/\chi)$. The initial velocity field is given by (4.5) with $u_0 = 0.1$, $z_0 = 1.0$ and $\Delta z = 0.5$, which corresponds to $E_{\mathrm{kin},0} \simeq 0.3E_{\mathrm{avail}}$. A movie version of this figure is available at https://doi.org/10.1017/S0022377824001521.

$E_{\mathrm{kin},0} \simeq 0.3E_{\mathrm{avail}}$). We observe that a substantial plume of material rises until reaching the stable region at the top of the simulation domain, where it overturns and develops Rayleigh–Taylor instabilities (upper left panel). Under advection by the increasingly chaotic flow, small-scale structures are developed (upper middle and right panels). These structures diffuse, ultimately leading to a one-dimensional but non-homogeneous state with no fine-scale structure (lower three panels).

Figure 24 compares the horizontally averaged instantaneous profiles of s and χ developed in the simulation with both $\langle s \rangle$ and $\langle \chi \rangle$ [as defined in (5.10) and (5.11)] and $\bar{s}$ and $\bar{\chi}$ [as defined in (5.13) and (5.14)]. The quantities are computed from $\mathcal{P}(m, \mu)$ calculated with $E_{\mathrm{pot}} = E_0 + E_{\mathrm{kin},0}$ in (5.4).[14]

[14]The Mach number of the simulation, $\mathrm{Ma} \equiv u_0/\sqrt{v_A^2 + c_s^2} \sim 0.1$ is small, so that the compressible part of the initial velocity field rapidly propagates away as compressive waves. The energy associated with these waves (i.e. the

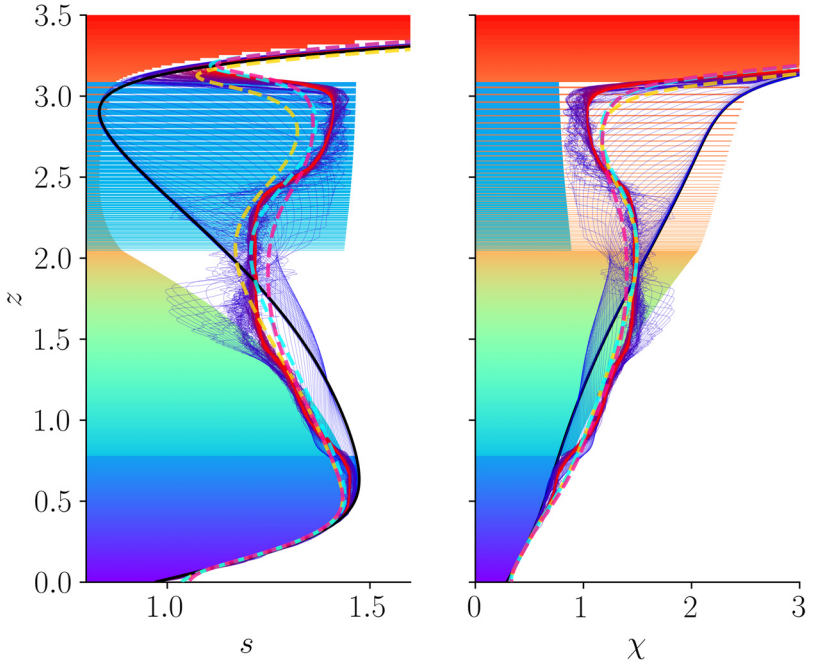

FIGURE 24. Horizontally averaged profiles of s and χ plotted at intervals of 1 code time unit, between $t = 0$ (blue) to $t = 600$ (red), for the simulation visualised in figure 23. Also plotted are the profiles that correspond to expectation values of $\mathcal{P}(m, \mu)$ (gold dashed line; which we claim is not a suitable model of diffusion), from (5.13) and (5.14) (cyan dashed line; these model energy- and flux-conserving diffusion), and after iterating the statistical mechanical prediction from the profile based on the cyan dashed line (pink dashed line). For reference, we also plot the minimum-energy state: this is as shown in figure 5.

We make the following observations. First, the late-time profile of s is almost everywhere larger than $\langle s \rangle$ ((5.10); gold dashed line in figure 24). This validates the reasoning we used to reject (5.10) as a predictor of the final state – evidently, dissipation causes the entropy of the fluid to grow. Secondly, $\bar{s}$ and $\bar{\chi}$ computed from (5.13) and (5.14) (cyan dashed line in figure 24) constitute a very reasonable prediction of the relaxed state. The chief discrepancies are in the range $2.5 \lesssim z \lesssim 3.0$, where $\bar{s}$ and $\bar{\chi}$ are, respectively, somewhat smaller and larger than in the late-time profiles developed by the simulation. We interpret this as a consequence of the system not 'exploring' the full surface of constant energy in configuration space, owing to the fluid that rises from the bottom of the equilibrium becoming trapped in a metastable state at the top. In support of this interpretation, we note that, in the range $0.9 \lesssim z \lesssim 1.5$, the cyan lines somewhat over- and under-predict the simulation result, respectively, indicating that 'too much' material rose in the initial plume (and became stuck).

A second discrepancy between the cyan line and the simulation result is that the latter exhibits a clear plateau in the range $1.5 \lesssim z \lesssim 2.3$. On the other hand, the cyan line is nonlinearly unstable [see figure 3(*b*) for its force diagram] – the dashed pink line in figure 24 shows the result of taking it as the initial state for a secondary relaxation (see § 5.4). The pink line features a plateau over roughly the same range of z as the one that

initial energy of the non-solenoidal part of $\boldsymbol{u}$) is, we assume, irrelevant to the otherwise quasi-incompressible dynamics, so we exclude it from the energy we use for E_{pot}.

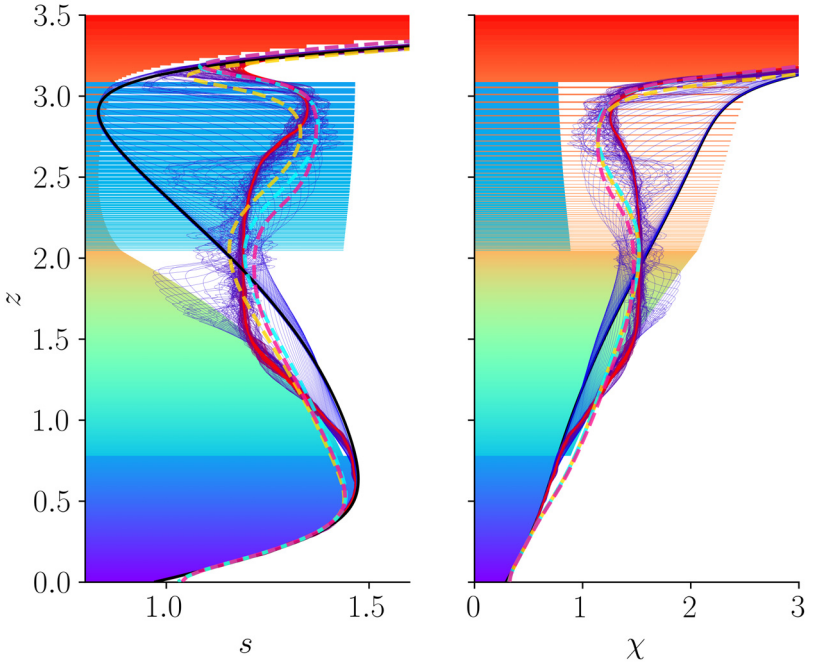

FIGURE 25. As in figure 24, but for a simulation initialised with $u_0 = 0.05$ in (4.5) ($E_{\mathrm{kin},0} \simeq 0.1E_{\mathrm{avail}}$).

forms in the simulation, although the predicted plateau has slightly larger s (smaller χ). This might be interpreted as a consequence of the fact that the large-s fluid that would have risen upwards under the secondary relaxation to form the plateau was, in the simulation, already displaced to the top of the atmosphere. As a consequence, the plateau forms with somewhat smaller s than it would otherwise have had.

Figures 25 and 26 are analogous to figure 24 but for simulations with $u_0 = 0.05$ ($E_{\mathrm{kin},0} \simeq 0.1E_{\mathrm{avail}}$) and $u_0 = 0.025$ ($E_{\mathrm{kin},0} \simeq 0.02E_{\mathrm{avail}}$), respectively. In these cases, the quantitative agreement between simulation and theory is less good than in figure 24: with a smaller initial impulse, relaxation is incomplete. Figure 27, shows that, at peak, between 50 % and 60 % of the available kinetic plus potential energy is in the form of kinetic energy for all three simulations, showing that the liberation of available potential energy during relaxation is fairly efficient.

6.2. *Metastable downwards,* (2.31)

We now report the results of analogous simulations to those in § 6.2, but for the unstable-downwards profile (2.31) and with $z_0 = 2.25$ in (4.5).

Figure 28 visualises the evolution in x–z space for $u_0 = 0.14$ ($E_{\mathrm{kin},0} \simeq 0.2E_{\mathrm{avail}}$), following the tracer s/χ. Similarly to figure 23, we observe that a descending plume reaches the stable buffer region at the bottom of the simulation domain, develops small-scale structure due to Rayleigh–Taylor instability and advection by chaotic motions, and ultimately diffuses. The evolution of the horizontally averaged profiles is displayed in figures 29–31 for the cases of $u_0 = 0.14$ ($E_{\mathrm{kin},0} \simeq 0.2E_{\mathrm{avail}}$), $u_0 = 0.1$ ($E_{\mathrm{kin},0} \simeq 0.1E_{\mathrm{avail}}$), and $u_0 = 0.05$ ($E_{\mathrm{kin},0} \simeq 0.03E_{\mathrm{avail}}$), respectively. Again, we observe reasonable agreement between the statistical mechanical prediction and the late-time limit of the simulations, although the predictions do somewhat under-predict s and over-predict χ for $z \gtrsim 1.5$.

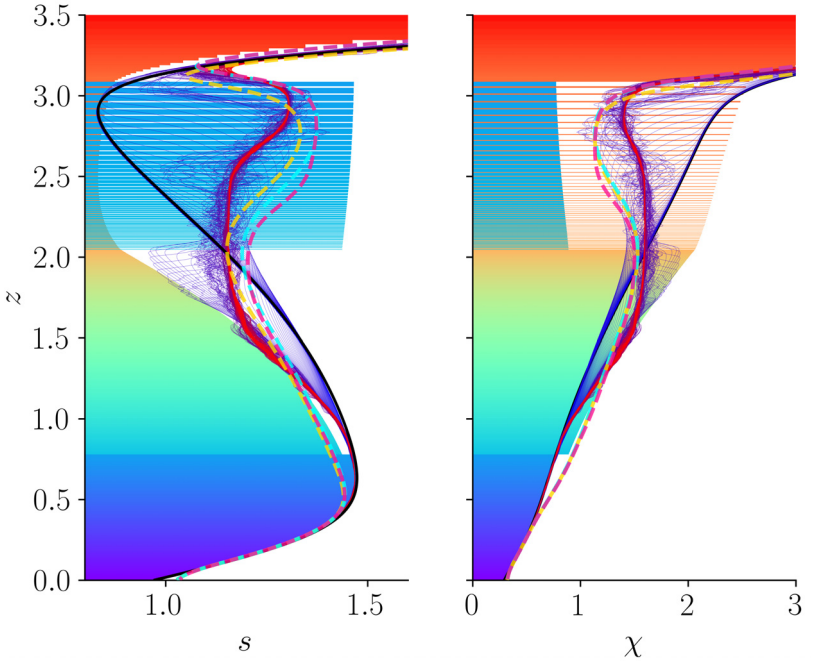

FIGURE 26. As in figure 24, but for a simulation initialised with $u_0 = 0.025$ in (4.5) ($E_{\mathrm{kin},0} \simeq 0.02 E_{\mathrm{avail}}$).

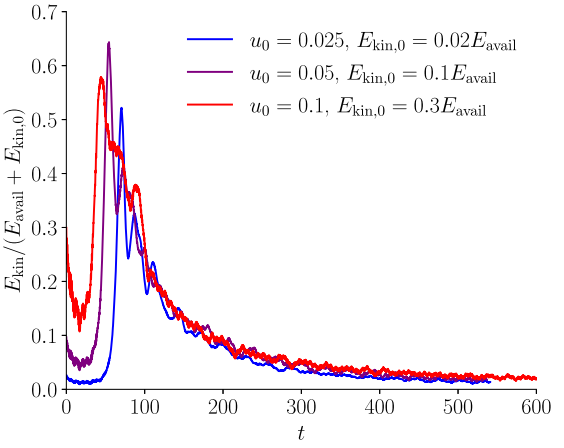

FIGURE 27. Evolution of the kinetic energy as a fraction of the total available energy, which is the kinetic energy plus the available potential energy of the initial state, for the simulations visualised in figures 24–26.

Again, the reason appears to be partial relaxation: the degree of inaccuracy is greater in the simulations with smaller initial velocity fields. We also note that the theory fails to predict the plateau that forms in the vicinity of $z \simeq 0.5$ (no secondary relaxation is possible from the diffused state that corresponds to the cyan line, as it is nonlinearly stable, see § 5.4). Nonetheless, the theoretical predictions agree reasonably well in this region with all three simulation profiles, the flatness notwithstanding. Finally, we show the

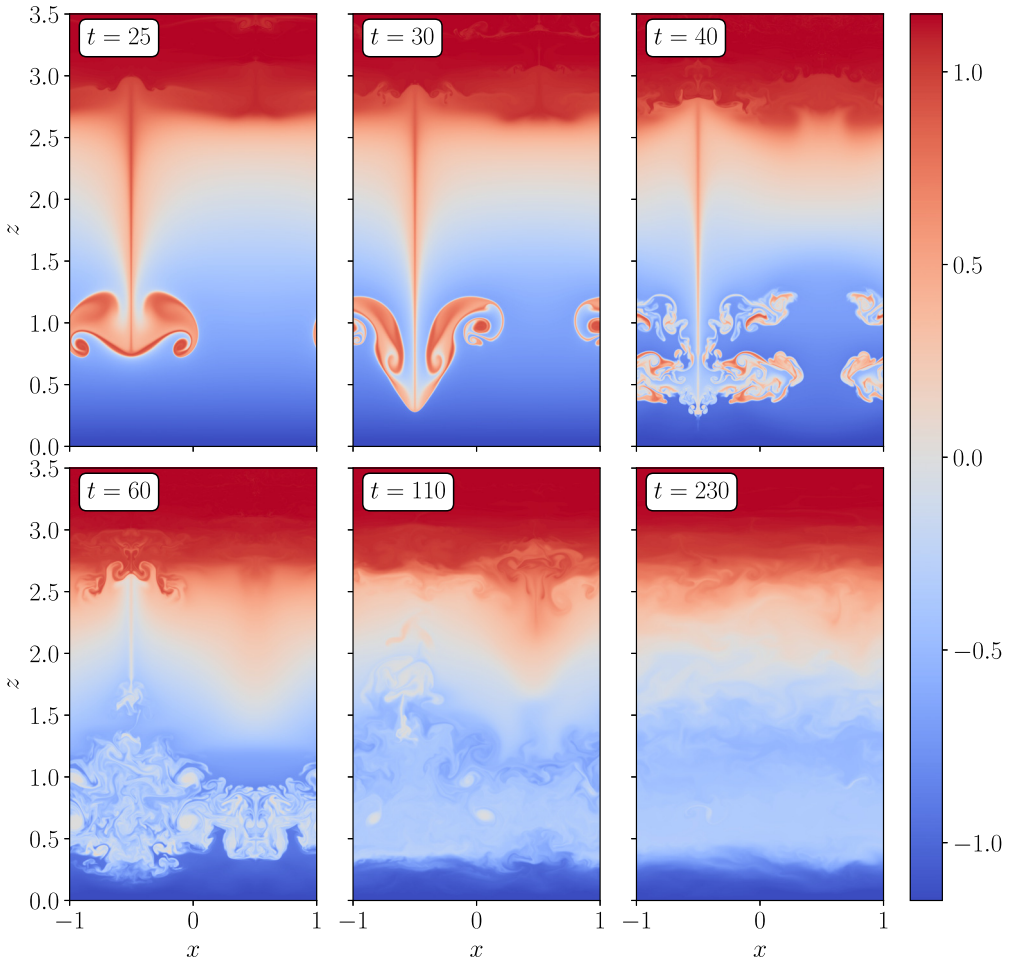

FIGURE 28. Numerical simulation of the relaxation of the equilibrium defined by (2.31) at $\mathrm{Re} \sim 10^5$. The quantity plotted is $\ln(s/\chi)$. The initial velocity field is given by (4.5) with $u_0 = 0.14$, $z_0 = 2.25$ and $\Delta z = 0.5$; this corresponds to an initial kinetic energy $\simeq 0.2$ times the available potential energy. A movie version of this figure is available at https://doi.org/10.1017/S0022377824001521.

evolution of the kinetic energy as a fraction of the total available energy in figure 32. As in figure 27, the relaxation is fairly efficient at liberating available potential energy. Large initial perturbations are not required to do so: over a factor $\simeq 30$ difference in the initial kinetic energy, the peak ratio of kinetic energy to available energy varies only between around 30 % and 45 %.

7. Conclusion

7.1. *Summary*

In this work, we have demonstrated that MHD equilibria with straight magnetic-field lines can be metastable to 2D interchange-type motions in the plane perpendicular to the magnetic field (§ 2). This phenomenon occurs because fluid with large plasma β is more compressible than fluid with small β. As a result, when displaced upwards (for example) by a large distance and thus exposed to a large change in pressure, the density of large-β

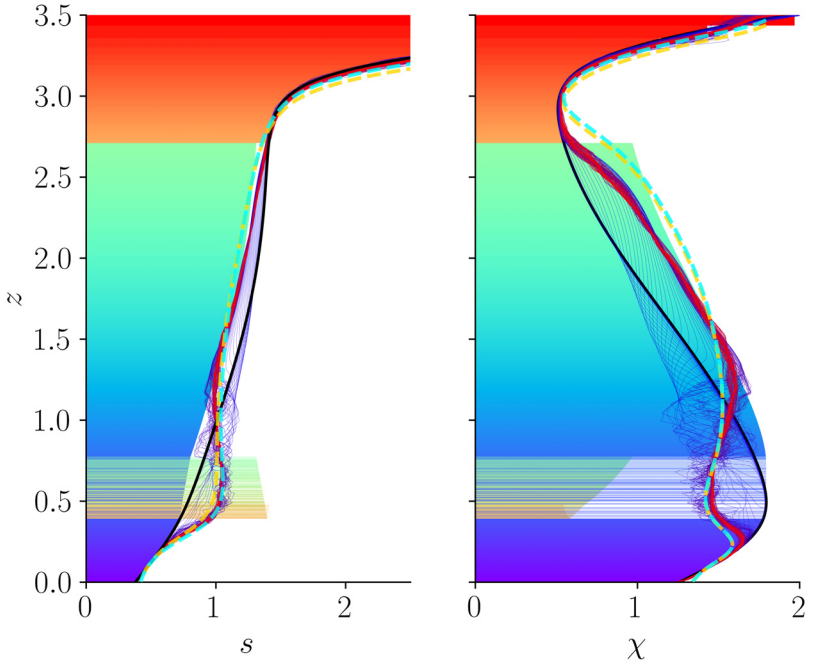

FIGURE 29. Horizontally averaged profiles of s and χ plotted at intervals of 1 code time unit, between $t = 0$ (blue) to $t = 600$ (red), for the simulation visualised in figure 28. Also plotted are the profiles obtained by taking expectation values of $\mathcal{P}(m, \mu)$ (gold dashed line) and from (5.13) and (5.14) (cyan dashed line).

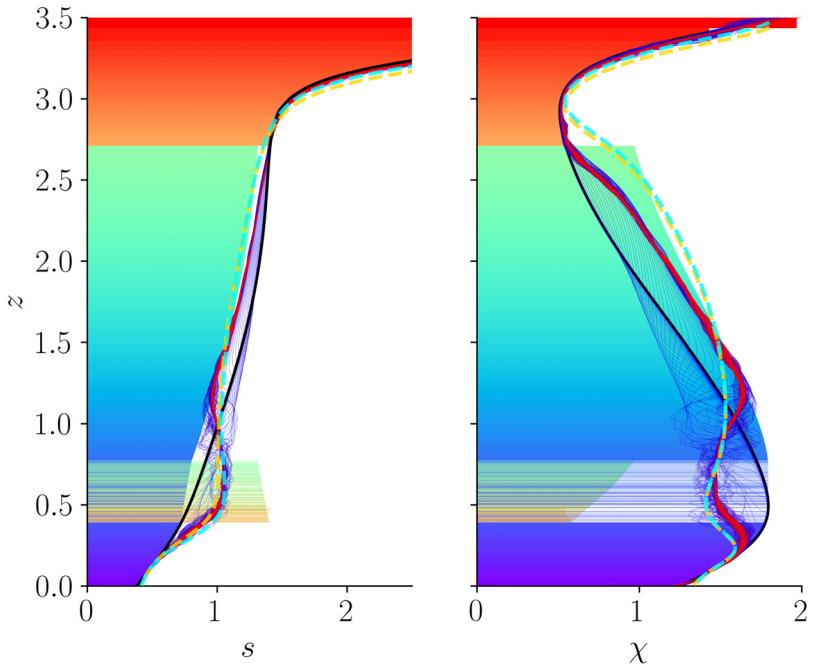

FIGURE 30. As in figure 29, but for a simulation initialised with $u_0 = 0.1$ in (4.5), which corresponds to $E_{\mathrm{kin},0} \simeq 0.1 E_{\mathrm{avail}}$.

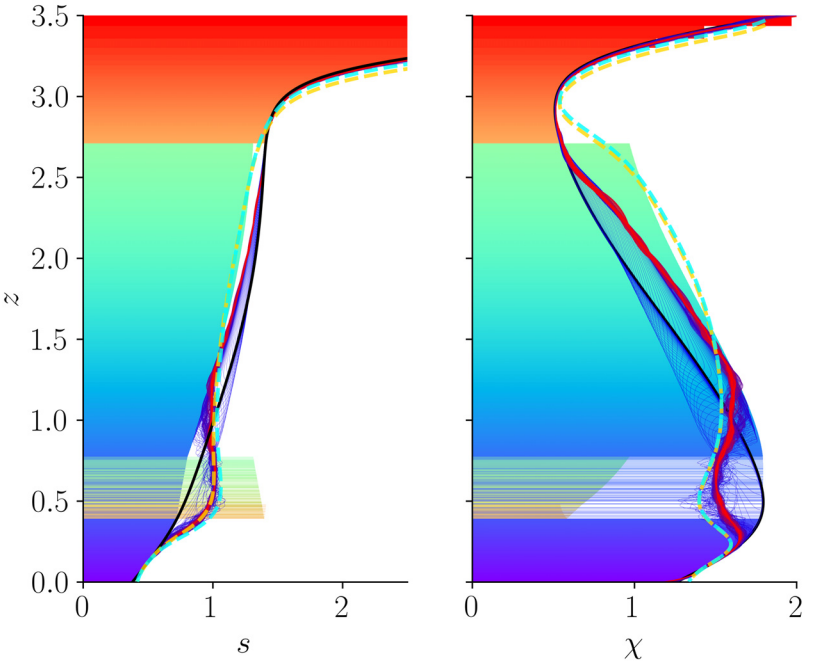

FIGURE 31. As in figure 29, but for a simulation initialised with $u_0 = 0.05$ in (4.5), which corresponds to $E_{\mathrm{kin},0} \simeq 0.03 E_{\mathrm{avail}}$.

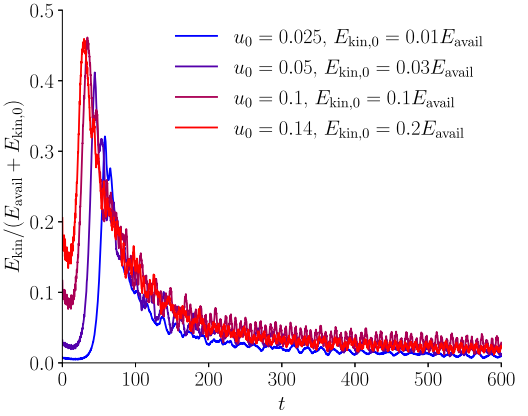

FIGURE 32. Evolution of the kinetic energy as a fraction of the total available energy, which is the kinetic energy plus the available potential energy of the initial state, for the simulations visualised in figures 29–31, as well as for one with $u_0 = 0.025$ in (4.5).

fluid can be less than that of the ambient fluid at the new location, even if this would not be the case for sufficiently local (i.e. linear) displacements.

The existence of metastability in 2D is particularly interesting because ideal (i.e. dissipationless) relaxation in 2D constitutes rearrangement of flux tubes subject to the Lagrangian invariance of entropy and magnetic flux. This enables the use of combinatorial and statistical techniques that are unavailable for the 3D problem (although see § 7.2). We

have determined the minimum-energy state consistent with 2D rearrangements by solving a LSA problem in four different illustrative cases (§ 3). We find that the available potential energy is generally small, a result that can be traced back to the fact that fluid parcels exclude each other, so that the mass of fluid that can experience a significant change in pressure is limited (§ 3.1). An interesting finding is that the minimum-energy states are two dimensional, despite the initial profiles being one dimensional (§ 3.3.1). We show that the relaxed state that develops in numerical simulations at small Reynolds number approximates this 2D minimum-energy state (§ 4), as the potential energy liberated during relaxation is dissipated by viscosity (changing the entropy of the fluid only slightly, owing to the small energy scales involved).

The two-dimensionality of the minimum-energy states may be interpreted as a consequence of the fact that the 1D states with least potential energy are, in general, Rayleigh–Taylor unstable. Provided that it is not suppressed by viscosity, this instability (and the potential energy liberated during relaxation) drives a turbulent flow that generates small-scale structure, enabling diffusion to act and violate the fluid-element-wise conservation of entropy and magnetic flux (§ 5). We have proposed that such cases can be modelled theoretically by postulating (non-rigorously, but, apparently, usefully) a separation of time scales between the one on which the flow is mixed (i.e. the ideal dynamical time scale) and the one on which the advected invariants diffuse. We suggest that the final state of the former relaxation is the equilibrium state that maximises the Boltzmann mixing entropy subject to fixed energy (§ 5.2). This is analogous to the Lynden-Bell theory of 'violent relaxation' of collisionless stellar systems and plasma, and to the RSM theory of 2D vortex turbulence. The generalisation of these theories to more than one conserved quantity (thermal entropy and magnetic flux) turns out to be uncomplicated because the theory can be recast as a maximisation of the mixing entropy associated with rearrangements of 1D slices (§ 5.1).

We take the latter, diffusive part of the relaxation to be a homogenisation of the stochastic small-scale structure present in a statistical-mechanical microstate. We obtain a prediction for the post-diffusion state by collapsing the statistical mechanical probability function onto a value determined by the local conservation of enthalpy and magnetic flux (§ 5.3). Because the enthalpy density is a nonlinear function of the ideal Lagrangian invariants of the fluid, the states that one derives in this manner are not necessarily (or, indeed, usually) stable to ideal dynamics (§ 5.4). We propose that a natural scheme for dealing with this, which is consistent with the philosophy of time scale separation between ideal and non-ideal effects, is to iterate the procedure described above for the new profiles – i.e., to seek the mixing-entropy-maximising ideal rearrangement and then allow diffusion to act upon it – and to continue iterating until a stable profile is reached. The difference between the profiles predicted at the first and final iterations is not typically very large; the qualitative outcome of the subsequent iterations is to produce plateaus in the profiles of s and χ – see § 7.2 for further discussion.

We compare the theoretical predictions described above with the results of 2D numerical simulations at large Reynolds number in § 6. Provided that the equilibrium is perturbed sufficiently strongly to generate thorough mixing, we find remarkably good agreement between the late-time state of the numerical simulations and the predictions of our statistical mechanical theory. We also observe in the simulations the formation of well-mixed plateaus whose properties are consistent with the idea that they are formed by further mixing of diffused fluid. For weak initial perturbations, the agreement between the late-time states of the numerical simulations and our theory is less good. We interpret this as a consequence of the metastability phenomenon itself – when the relaxing system

becomes 'stuck' in a new metastable state, it cannot explore the full constant-energy surface in its configuration space.

7.2. *Discussion*

Our study suggests a number of questions for future investigation. One concerns the role of metastability in driven systems. The present study has been motivated by the fact that, unlike equilibria that are very unstable linearly, equilibria that are strongly metastable (in the sense of having a large amount of available energy) are realisable in real physical systems by an evolution of the potential-energy landscape that preserves the local minimum within which the system resides. It would be illuminating to understand whether, and under what conditions, the development of a metastable state actually happens in a dynamically evolving system, such as one driven towards convective instability by external heating (and/or cooling). If metastable states do develop, do they relax periodically via sporadic eruptions? How frequent are those eruptions? And are the corresponding relaxations 'complete', in the sense of taking the system far from the linear-stability limit, or does the state of the system always remain close to this limit? Such questions are particularly pertinent in light of the observation that edge-localised modes in tokamaks – which are believed to be manifestations of metastable dynamics in driven systems (see the Introduction and Cowley & Artun 1997; Hurricane *et al.* 1997; Wilson & Cowley 2004; Cowley *et al.* 2015; Ham *et al.* 2018) – are observed to leave the plasma much below the threshold for linear instability post-eruption (see, e.g. Kirk *et al.* 2004, 2006).

A natural question is whether the methods employed within this work can be adapted to a fully 3D dynamics. In 3D, B/ρ is not an ideal invariant: the magnetic flux through any material surface is conserved, but the density of the fluid in that surface can change due to motions perpendicular to the surface as well as parallel to it. Thus, relaxation in 3D is not simply a rearrangement of fluid parcels with associated invariants. Non-ideal processes too are significantly more complex in the 3D problem than the simple diffusion of scalar quantities – with magnetic field now a vector, magnetic topology imposes constraints (some, but not all, of which can be broken by magnetic reconnection in the subsequent relaxation Taylor 1974, 1986; Zhou *et al.* 2019; Bhat, Zhou & Loureiro 2021; Hosking & Schekochihin 2021). Nonetheless, we note that the Hungarian algorithm (§ 3.2) could still be employed to calculate a rigorous lower bound on the available energy of an equilibrium with initially straight magnetic field: the smallest potential energy that the field can have under 2D interchanges of field lines evidently constitutes such a bound. Furthermore, it seems plausible that the final state of 3D relaxation would, in fact, be 2D: bent field lines would tend to reconnect and straighten out under magnetic tension. Speculatively, if the development of such a final state were to constitute an effective series of interchanges, the statistical methods developed in this paper might well be applied usefully.

Another intriguing question for future work is whether combinatorial and statistical theories of relaxation can predict the nonlinear saturation of double-diffusive instabilities, for example, the fingering instability that occurs if a quantity whose stratification is stabilising is diffusive, so that its stabilisation is not felt on sufficiently small scales (see, for example, Hughes & Brummell 2021, and references therein). Such systems saturate with staircase distributions of their compositional properties: thermohaline staircases in the ocean, which exhibit steps in their temperature and salinity profiles, are a prominent example. These staircases are apparently extremely stable – measurements of thermohaline staircases have revealed structure that persists over $\sim$ 100 km and for a time scale of years (see Merryfield 2000 for a review). Metastability to diffusive modes has been mooted as a possible explanation for the staircases (Merryfield 2000). Although diffusion is not naturally incorporated into the LSA problem of finding minimum-energy states (§ 3),

one could imagine incorporating different rates of diffusion into the iterative model for post-diffusive relaxation described in §§ 5.3 and 5.4. Whether such a scheme would reproduce staircases, or offer qualitative insights into the mechanisms by which they form, remains to be seen. Some cautious optimism can be derived from results like those shown in figure 21*c*, which constitutes a theoretical prediction of a staircase-like structure (although not as a result of double-diffusive instability).

In closing, let us consider whether the results of this work might be valuable for 'violent relaxation' in other contexts. A straightforward yet useful result is that the statistical mechanical theory can work well in predicting relaxed states, provided the system is perturbed sufficiently strongly for it to become well mixed. A second result is that the theory appears to work well in systems that possess more than one invariant. Ewart *et al.* (2023) have speculated about the possibility of predicting the relaxation of a collisionless magnetised plasma by enforcing the conservation of two invariants: the phase-space density η and the magnetic moment $\mu_b = mv_\perp^2/2B$ (here, m is the mass of a particle, and $v_\perp$ is the magnitude of the velocity of the particle in the direction perpendicular to the magnetic field B). It is interesting to note that, under such a scheme, the energy $\mathcal{E}$ associated with a volume element of phase space is a nonlinear function of conserved quantities, *viz.*, $\mathcal{E} = \eta(mv_\parallel^2/2 + \mu_b B)$. It follows that such a theory would need to address questions similar to the ones we considered in § 5.3, about how one extracts predictions from the Lynden-Bell probability distribution: the distribution function corresponding to its expectation values will, in general, not have the correct energy. We suggest that the fix might be, as in this work, to acknowledge that diffusion (in this case, particle collisions) does not conserve η and μ_b, but, rather, η and $\mathcal{E}$, and, therefore, to evaluate the distribution function according to a scheme analogous to (5.13) and (5.14). Whether or not the resulting distribution functions are unstable (as for the profiles that develop due to diffusion in our study), and, if so, the prediction of their further evolution, would, we suggest, be an interesting topic for future exploration.

Supplementary movies

Supplementary movies are available at https://doi.org/10.1017/S0022377824001521.

Acknowledgements

We thank J. Brown, R. Ewart, T. Foster, A. Kerstein, M. Kunz, H. Latter, E. McCulloch, M. Nastac, G. Ogilvie, E. Quataert, A. Schekochihin, N. Shibley and A. Spitkovsky for helpful discussions and suggestions. In particular, we are indebted to R. Ewart and A. Schekochihin for their suggestion that Lynden-Bell's violent-relaxation formalism might be applicable to our developing work on metastability in 2D MHD. We are grateful to the two anonymous referees, whose recommendations have improved this paper.

Editor Per Helander thanks the referees for their advice in evaluating this article.

Declaration of interests

The authors report no conflict of interest.

Funding

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Appendix A. Details of the numerical simulations

The numerical simulations presented in this work were conducted with the finite-difference MHD code Pencil (Pencil Code Collaboration *et al.* 2021). The code uses sixth-order finite differences and a third-order-accurate time-stepping scheme to solve the equations of 2D MHD with a constant gravitational field, i.e.

$$\frac{\partial \rho}{\partial t} + \boldsymbol{u} \cdot \boldsymbol{\nabla} \rho = -\rho \boldsymbol{\nabla} \cdot \boldsymbol{u}, \tag{A1}$$

$$\rho \left(\frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{u} \cdot \boldsymbol{\nabla} \boldsymbol{u} \right) = -\boldsymbol{\nabla} \left(p + \frac{B^2}{2} \right) - \rho g \hat{\boldsymbol{z}} + \rho \nu \left(\nabla^2 \boldsymbol{u} + \frac{1}{3} \boldsymbol{\nabla} (\boldsymbol{\nabla} \cdot \boldsymbol{u}) \right), \tag{A2}$$

$$\begin{aligned} \frac{\partial p}{\partial t} + \boldsymbol{u} \cdot \boldsymbol{\nabla} p = -\gamma p \boldsymbol{\nabla} \cdot \boldsymbol{u} + (\gamma - 1) \Bigg[\rho \nu \left(2 e_{ij} e_{ij} - \frac{2}{3} (\boldsymbol{\nabla} \cdot \boldsymbol{u})^2 \right) \\ + \eta |\boldsymbol{\nabla} \times (B \hat{\boldsymbol{y}})|^2 + \rho K \nabla^2 \left(\frac{p}{\rho} \right) \Bigg], \end{aligned} \tag{A3}$$

where

$$e_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right), \tag{A4}$$

is the rate-of-strain tensor, and

$$\frac{\partial B}{\partial t} + \boldsymbol{u} \cdot \boldsymbol{\nabla} B = -B \boldsymbol{\nabla} \cdot \boldsymbol{u} + \eta \nabla^2 B. \tag{A5}$$

In all simulations, we use reflecting boundary conditions (i.e. anti-symmetry for the component of velocity field in the direction normal to the boundary (u_z) and symmetry for the component in the other direction (u_x)) at the boundaries in z, while enforcing anti-symmetry relative to the value on the boundary (i.e. vanishing second derivative) for B, p and ρ. For the simulations reported in § 1, we simulate directly only the region with $0 \leq x \leq 1$ and $0 \leq z \leq 3.5$, with reflecting boundary conditions in the x direction (and symmetric boundary conditions for B, p and ρ) – we construct the visualisations in figures 1 and 2 by reflecting the simulation domain in the line $x = 0$. For all other simulations, we employ periodic boundary conditions in the x direction (and simulate the full range $-1 \leq x \leq 1$). The resolutions of the simulations are 1166×4080 for those reported in § 1, 1166×2040 for those reported in § 4.2, and 2332×4080 for those reported in § 6.

In all simulations, we choose the adiabatic index $\gamma = 5/3$, magnetic diffusivity $\eta = 4 \times 10^{-6}$ and thermal diffusivity $K = 6 \times 10^{-6}$. This means that $\eta/K = 2/3 = \gamma - 1$, which is the critical ratio for stability to double-diffusive instabilities at any ν (Hughes 1985). The kinematic viscosity is $\nu = 4 \times 10^{-6}$ for the simulations in §§ 1 and 6 and $\nu = 1.6 \times 10^{-3}$ for the simulations in § 4.2. The details of the equilibrium states in which we initialise the simulations are explained in § 2.6.

Appendix B. Comparison with moist hydrodynamics

In this appendix, we apply the general formalism of § 2.2 to moist hydrodynamics, thus elucidating the analogy between metastability in 2D MHD and in the terrestrial atmosphere. The appendix is mostly a review of standard results, but derived economically via the linear- and nonlinear-stability criteria (2.7) and (2.13).

B.1. *Overview*

Moist hydrodynamics is the fluid dynamics of a mixture of dry air, water vapour and liquid water in local thermodynamic equilibrium. In this appendix, we restrict attention to the case where the liquid water is suspended in the air, as in clouds and fog, and does not precipitate out. As for 2D MHD (see (2.19) and (2.20)), moist hydrodynamics has two quantities that are conserved in a Lagrangian sense in the absence of diffusion. These are the specific entropy S and the water mixing ratio w, which is the ratio of the mass of water (both liquid and vapour) to the mass of dry air. As with our use of the 'entropy function' s in the main text (see (2.20)), it is more convenient to work with a quantity derived from S – in this case, potential temperature, θ – than with S directly. The vector of conserved quantities (2.4) for moist hydrodynamics is therefore $\boldsymbol{Q} = (\theta, w)$.

In the following sections, we consider the linear and nonlinear stability of unsaturated air (composed of dry air and water vapour) and saturated air (in which water is in both vapour and liquid states) in turn. For unsaturated air, we review the definition of specific entropy S in § B.2.1, use it to derive a formula for potential temperature θ in § B.2.2 and apply the general formula (2.7) to determine the criterion for linear stability in § B.2.3. We show in § B.2.4 that the compressibility κ of unsaturated air increases with w because the specific heat capacity of water vapour is greater than that of dry air. However, because $w \ll 1$ in the atmosphere, differences in compressibility are always small, meaning that metastability does not occur with only unsaturated air in practice.

On the other hand, when a parcel of unsaturated air rises and cools sufficiently for vapour to condense (at the so-called lifting condensation level), the newly saturated parcel becomes significantly more compressible than its dry-air surroundings. This is because further decrease in pressure leads to additional cooling and condensation, which releases latent heat and re-warms the parcel somewhat, leading to additional expansion. As a result, the density of the saturated parcel decreases more in response to a change in pressure than the density of the surrounding dry air does. Furthermore, because the specific latent heat of condensation of water is much greater than the typical thermal energy per unit mass of air, differences in compressibility between saturated and unsaturated air can be significant even for $w \ll 1$. In the Earth's atmosphere, cumulonimbus clouds form as the result of nonlinear instability arising from this effect (see, e.g. Rogers & Yau 1996). We calculate the compressibility of saturated air in § B.3.3, after introducing its specific entropy and the liquid-water potential temperature in §§ B.3.1 and B.3.2, respectively. We determine the linear-stability criterion for an atmosphere containing saturated air in § B.3.4.

B.2. *Case of unsaturated air*

B.2.1. *Specific entropy of unsaturated air*

In what follows, we use the subscripts d, v, l and w to refer to dry air, water vapour, liquid water (in § B.3 only) and total water, respectively. For the cases of dry air and water vapour, we have from the first law of thermodynamics applied to an ideal gas that

$$\mathrm{d}S_i = \frac{c_i}{M_i}\frac{\mathrm{d}T}{T} - \frac{R}{M_i}\frac{\mathrm{d}p_i}{p_i} \implies S_i = S_{0i} + \frac{1}{M_i}\left(c_i \ln\frac{T}{T_0} - R\ln\frac{p_i}{P_0}\right), \tag{B1}$$

where $i \in \{d, v\}$ is the species index, S_i specific entropy, p_i partial pressure, c_i molar heat capacity at constant pressure, M_i molar mass, R the universal gas constant and S_{0i} the specific entropy in the reference state that has temperature and partial pressure equal to T_0 and P_0, respectively. With m_i the mass of species i in a mixture, the specific entropy S_{unsat} of a mixture of dry air and water vapour satisfies

$$(m_d + m_v)S_{\mathrm{unsat}} = m_d S_d + m_v S_v \implies (1+w)S_{\mathrm{unsat}} = S_d + wS_v, \tag{B2}$$

where the water content $w \equiv m_w/m_d$ is equal to m_v/m_d because all water is in the vapour state. Substituting (B1) into (B2), using $p_i V = n_i RT \implies p_v/p_d = w/\varepsilon$ where $\varepsilon \equiv M_d/M_v$ and $p_d + p_v = P$, we obtain an expression for $S_{\text{unsat}} = S_{\text{unsat}}(P, T, w)$:

$$M_d(1+w)S_{\text{unsat}} = M_d(S_{0d} + wS_{0v}) + \left(c_d + \frac{w}{\varepsilon}c_v\right)\ln\frac{T}{T_0} - R\left(1+\frac{w}{\varepsilon}\right)\ln\frac{P}{P_0} + R\left(1+\frac{w}{\varepsilon}\right)\ln\left(1+\frac{w}{\varepsilon}\right) - R\frac{w}{\varepsilon}\ln\frac{w}{\varepsilon}. \tag{B3}$$

B.2.2. *Potential temperature*

It is conventional in studies of convection to work not with the entropy directly but rather potential temperature θ, which may be defined as

$$\ln\frac{\theta}{T_0} \equiv \frac{1}{c_d + wc_v/\varepsilon}\left[M_d(1+w)S_{\text{unsat}} - M_d(S_{0d} + wS_{0v}) - R\left(1+\frac{w}{\varepsilon}\right)\ln\left(1+\frac{w}{\varepsilon}\right) + R\frac{w}{\varepsilon}\ln\frac{w}{\varepsilon}\right] \tag{B4}$$

$$= \ln\frac{T}{T_0} - \left[1 - \frac{1}{\Gamma(w)}\right]\ln\frac{P}{P_0}, \tag{B5}$$

so that

$$\theta = T\left(\frac{P}{P_0}\right)^{1/\Gamma(w)-1}, \tag{B6}$$

where the adiabatic index is

$$\Gamma(w) \equiv \frac{c_d + c_v w/\varepsilon}{c_d - R + (c_v - R)w/\varepsilon}. \tag{B7}$$

The potential temperature θ is the temperature of a fluid parcel moved at fixed entropy and water-mixing ratio to the reference pressure P_0. We see from its definition (B4) that θ is conserved under isentropic displacements that preserve the composition w.

B.2.3. *Linear stability of unsaturated air*

The linear-stability criterion (2.7) reads

$$\mathcal{L} \equiv -\frac{\mathrm{d}\boldsymbol{Q}}{\mathrm{d}z}\cdot\frac{\partial\ln\rho(P,\boldsymbol{Q})}{\partial\boldsymbol{Q}} > 0, \quad \forall\, z. \tag{B8}$$

We therefore require $\rho = \rho(P, \theta, w)$. It is straightforward to show from the ideal gas law that

$$P = \rho\frac{RT}{M_d}\frac{1+w/\varepsilon}{1+w} \implies \rho = \frac{M_d P_0}{R\theta}\left(\frac{P}{P_0}\right)^{1/\Gamma(w)}\frac{1+w}{1+w/\varepsilon}. \tag{B9}$$

The partial derivatives in (B8) are then

$$\left(\frac{\partial\ln\rho}{\partial\theta}\right)_{w,P} = -\frac{1}{\theta}, \quad \left(\frac{\partial\ln\rho}{\partial w}\right)_{\theta,P} = -\frac{\Gamma'}{\Gamma^2}\ln\left(\frac{P}{P_0}\right) + \frac{1}{1+w} - \frac{1}{\varepsilon+w}, \tag{B10a,b}$$

where $\Gamma' = \mathrm{d}\Gamma/\mathrm{d}w < 0$. The criterion for linear stability becomes [cf. (2.22)]

$$\mathcal{L} = \frac{1}{\theta}\frac{\mathrm{d}\theta}{\mathrm{d}z} + \left[\frac{\Gamma'}{\Gamma^2}\ln\left(\frac{P}{P_0}\right) + \frac{1-\varepsilon}{(1+w)(\varepsilon+w)}\right]\frac{\mathrm{d}w}{\mathrm{d}z} \tag{B11}$$

$$= \frac{\mathrm{d}\ln T}{\mathrm{d}z} + \left[\frac{1}{\Gamma(w)} - 1\right]\frac{\mathrm{d}\ln P}{\mathrm{d}z} + \frac{1-\varepsilon}{(1+w)(\varepsilon+w)}\frac{\mathrm{d}w}{\mathrm{d}z} > 0, \quad \forall z. \tag{B12}$$

The first two terms become the negative of the gradient of the potential temperature if w is constant in z; in that case, the system is stable if $\mathrm{d}\theta/\mathrm{d}z > 0$. The third term represents the contribution from moisture content: we see that the presence of water vapour is stabilising when w increases with height; this is because the molar mass of water is less than that of dry air ($\varepsilon < 1$).

B.2.4. *Metastability of unsaturated air*

By (B9), the compressibility of unsaturated air is (cf. (2.23))

$$\kappa \equiv \frac{\partial \ln \rho(P,\theta,w)}{\partial \ln P} = \frac{1}{\Gamma(w)}. \tag{B13}$$

The adiabatic index $\Gamma(w)$ is a decreasing function of w because the specific heat capacity of water vapour is greater than that of dry air. Therefore, air with greater water mixing ratio w is more compressible, and so unsaturated air can be metastable when (i) the atmosphere is sufficiently close to marginal linear stability and (ii) wetter air moves through dryer air. However, because w in the atmosphere is typically smaller than 1 %, differences in κ between different parcels of unsaturated air are small. Metastability in the atmosphere occurs in practice because of the release of latent heat during condensation, as we now describe.

B.3. *Case of saturated air*

B.3.1. *Specific entropy of saturated air*

Once the partial pressure of water vapour p_v becomes equal to the saturation vapour pressure p_{sat} (a known function of T stated explicitly in § B.3.5), the vapour condenses to form liquid water. The moist air is then composed of dry air, vapour and suspended liquid water. Its specific entropy S_{sat} satisfies

$$(m_d + m_w)S_{\mathrm{sat}} = m_d S_d + m_v S_v + m_l S_l \implies (1+w)S_{\mathrm{sat}} = S_d + (w - w_v)S_l + w_v S_v, \tag{B14}$$

where we define $w_v \equiv m_v/m_d$. In the saturated state, $p_v = p_{\mathrm{sat}}$, so $P = p_d + p_{\mathrm{sat}}$. Thus, from $p_i V = n_i R T$, we have that

$$w_v = \frac{\varepsilon p_{\mathrm{sat}}}{p_d} = \frac{\varepsilon p_{\mathrm{sat}}}{P - p_{\mathrm{sat}}}. \tag{B15}$$

Using these expressions, we can write S_d and S_v, given by (B1), in terms of P, T and w. The remaining specific entropy S_l can be determined by summing the parts that correspond to isobaric heating of vapour from T_0 to T, isothermal compression from P_0 to p_{sat}, and

finally condensation with a latent heat per mole of $L(T)$. Thus,

$$S_l = S_{0v} + \frac{1}{M_w}\left[c_v \ln\left(\frac{T}{T_0}\right) - R\ln\frac{p_{\rm sat}}{P_0} - \frac{L}{T}\right]. \tag{B16}$$

Substituting the three contributions into (B14), we obtain

$$\begin{aligned}(1+w)M_d S_{\rm sat} = M_d(S_{0d} + wS_{0v}) + \left(c_d + \frac{w}{\varepsilon}c_v\right)\ln\frac{T}{T_0} - R\left(1+\frac{w}{\varepsilon}\right)\ln\frac{P}{P_0} \\ + R\left(1+\frac{w}{\varepsilon}\right)\ln\left(1+\frac{w_v}{\varepsilon}\right) - R\frac{w}{\varepsilon}\ln\frac{w_v}{\varepsilon} - \frac{L}{\varepsilon T}(w-w_v).\end{aligned} \tag{B17}$$

This constitutes an expression for $S_{\rm sat}$ as a function of P, T and w. The functions $p_{\rm sat}(T)$ and $L(T)$ are given by (B29) and (B30), respectively.

B.3.2. *Liquid-water potential temperature*

Analogously to (B4), one can define a conserved quantity known as 'liquid-water potential temperature' by

$$\begin{aligned}\ln\frac{\theta_l}{T_0} \equiv \frac{1}{c_d + wc_v/\varepsilon}\Big[M_d(1+w)S_{\rm unsat} - M_d(S_{0d} + wS_{0v}) \\ -R\left(1+\frac{w}{\varepsilon}\right)\ln\left(1+\frac{w}{\varepsilon}\right) + R\frac{w}{\varepsilon}\ln\frac{w}{\varepsilon}\Big]\end{aligned} \tag{B18}$$

$$\begin{aligned}= \ln\frac{T}{T_0} - \left[1 - \frac{1}{\Gamma(w)}\right]\ln\frac{P}{P_0} + \frac{1}{c_d + wc_v/\varepsilon}\left[R\left(1+\frac{w}{\varepsilon}\right)\ln\left(\frac{1+w_v/\varepsilon}{1+w/\varepsilon}\right)\right. \\ \left. - R\frac{w}{\varepsilon}\ln\frac{w_v}{w} - \frac{L}{\varepsilon T}(w-w_v)\right].\end{aligned} \tag{B19}$$

Liquid-water potential temperature is the potential temperature that a parcel of air would have if all the water in it were vaporised ($w_v = w$). Because the latent heat of condensation of water is large, i.e. $L/\varepsilon RT \sim 40 \gg 1$ (see footnote 15 for characteristic sizes of these quantities), the term involving L in (B19) is typically much larger than the other terms in the second set of square brackets. Neglecting those terms, (B19) becomes

$$\theta_l = T\left(\frac{P}{P_0}\right)^{1/\Gamma(w)-1}\exp\left(-\frac{L}{\varepsilon c_d T}(w-w_v)\right). \tag{B20}$$

B.3.3. *Change in compressibility on saturation*

The compressibility κ of moist air may be determined by inverting (B19) for $T = T(P, \theta_l, w)$, substituting the result into

$$P = \rho\frac{RT}{M_d}\frac{1+w_v/\varepsilon}{1+w}, \tag{B21}$$

and then taking the derivative of ρ with respect to P at fixed θ_l and w. The result of doing so numerically is shown in figure 33 for a number of different values of w, with $L(T)$ and $p_{\rm sat}(T)$ given by (B30) and (B29) in § B.3.5, respectively.[15] We observe that there

[15]We take $R = 8.31$ J mol^{-1} K^{-1}, $M_d = 0.0290$ kg mol^{-1}, $\varepsilon = 0.622$, $c_d = 29.2$ J mol^{-1} K^{-1}, $c_v = 33.7$ J mol^{-1} K^{-1}, $c_l = 75.5$ J mol^{-1} K^{-1}, $T_c = 283$ K, $L_c = 44\,600$ J mol^{-1} and $p_{\rm sat}(T_c) = 1230$ Pa. These are equivalent to the values quoted by Stansifer *et al.* (2017).

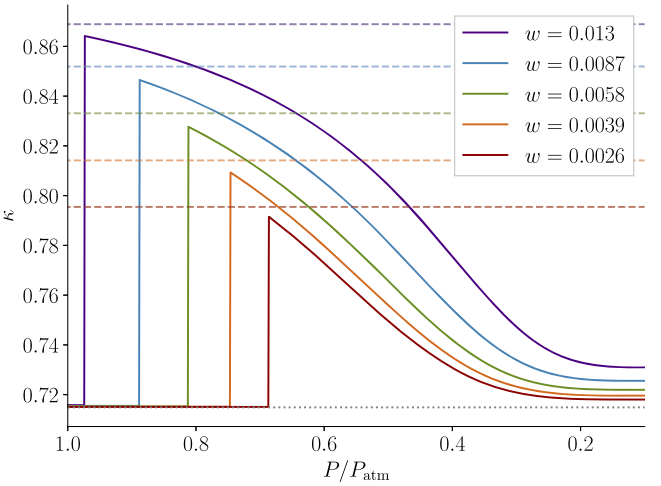

FIGURE 33. Solid lines show the compressibility κ of moist air as a function of total pressure P at fixed $\theta_l = 293$K and different mixing ratios w, calculated numerically from (B15), (B19) and (B21). Pressure is measured in units of $P_0 = P_{\rm atm} = 101.325$kPa. Dashed horizontal lines show the prediction (B24) for the maximum value of κ (small discrepancies with the maxima of the solid curves are due to approximations made in deriving (B24), specifically, the neglect of terms involving w that are not multiplied by the large ratio L/RT). The temperatures of saturation $T_{\rm sat}$ (obtained numerically) are, in order of decreasing w, 291 K, 283 K, 276K, 270 K and 263 K (temperatures smaller than that of the triple point of water, $\simeq$273 K, correspond to a supercooled liquid state – we neglect any effect of ice formation).

is a significant increase in compressibility when the pressure reaches the critical value at which condensation occurs ($p_v = p_{\rm sat}$). The height at which this happens for a rising air parcel is known as the level of lifting condensation (which can be at ground level if w is sufficiently large, as can be the case in tropical regions on Earth). The increase in compressibility that results from saturation means that as the parcel continues to rise, its density decreases relative to that of the surrounding air (provided that the latter is sufficiently close to marginal linear stability). Above the height at which they become equal, known as the level of free convection, the rising air parcel experiences an upwards force and is therefore unstable. The towering cloud formation that forms because of the resulting updraught is known as cumulonimbus (see, e.g. Rogers & Yau 1996).

The change in compressibility at saturation can be calculated analytically as follows. Due to the temperature dependencies of $L(T)$ [see (B30)] and w_v through $p_{\rm sat}(T)$ [see (B15)], we cannot invert (B20) analytically for $T = T(P, \theta_l, w)$, unlike the case for θ in § B.3.4. We therefore restrict attention to the point of saturation, which occurs at pressure $P_{\rm sat}(\theta_l, w)$ (note that $P_{\rm sat}$ is not the same as $p_{\rm sat}$: the former is the total pressure of the air at saturation while the latter is the partial pressure of the water vapour specifically). With $T_{\rm sat} \equiv T(P_{\rm sat}, \theta_l, w)$, we have, to first order in $\delta P \equiv P - P_{\rm sat}$,

$$\theta_l = \left[T_{\rm sat} + \left.\frac{\partial T(P, \theta_l, w)}{\partial P}\right|_{P=P_{\rm sat}} \delta P \right] \left(\frac{P_s}{P_0}\right)^{1/\Gamma - 1} \left[1 + \left(\frac{1}{\Gamma} - 1\right) \frac{\delta P}{P_{\rm sat}} \right]$$
$$\times \left[1 + \frac{L^2 w}{\varepsilon c_d R T_{\rm sat}^3} \left.\frac{\partial T(P, \theta_l, w)}{\partial P}\right|_{P=P_{\rm sat}} \delta P - \frac{wL}{\varepsilon c_d T} \frac{\delta P}{P_{\rm sat}} \right]. \tag{B22}$$

It follows that

$$\left.\frac{\partial T(P,\theta_l,w)}{\partial P}\right|_{P=P_{\rm sat}} = \frac{1/\Gamma - 1 - Lw/\varepsilon c_d T_{\rm sat}}{1 + L^2 w/\varepsilon c_d R T_{\rm sat}^2}. \tag{B23}$$

Substituting (B23) into (B21), we have

$$\kappa(P_{\rm sat},\theta_l,w) \equiv \left.\frac{\partial \ln\rho(P,\theta_l,w)}{\partial \ln P}\right|_{P=P_{\rm sat}} = \frac{1}{\Gamma} + \frac{w}{\varepsilon}\frac{L}{c_d T_{\rm sat}}\frac{(1-1/\Gamma)L/RT_{\rm sat} - 1}{1 + L^2 w/\varepsilon c_d R T_{\rm sat}^2}, \tag{B24}$$

where we have neglected the contribution to κ of the w_v in (B21). To leading order in w, the change in compressibility at saturation is

$$\Delta\kappa \equiv \kappa(P_{\rm sat},\theta_l,w) - 1/\Gamma(w) \simeq \frac{w}{\varepsilon}\frac{L^2}{Rc_v T_{\rm sat}^2} \sim 10^3 w. \tag{B25}$$

Thus, $\Delta\kappa \sim 0.1$ [which is the largest possible difference in κ in MHD, see (2.23)] can be achieved even with a water-mixing ratio of $\sim 10^{-4}$ (we recall from (3.9) that the fractional difference in density between a fluid parcel moved from pressure P_1 to P_2 and its surroundings at P_2 is $\sim \Delta\kappa/\kappa$ at marginal linear stability). For w much larger than this, the leading-order approximation (B25) is inaccurate and one must use (B24) instead.

B.3.4. *Linear stability of saturated air*

We note that the linear-stability criterion (B12) is modified if the air is saturated. It is straightforward to show from (2.8) that the criterion becomes

$$\mathcal{L} = \frac{{\rm d}\ln T}{{\rm d}z} + [\kappa(P(z),\theta_l(z),w(z)) - 1]\frac{{\rm d}\ln P}{{\rm d}z} > 0, \tag{B26}$$

provided that w_v and w can be neglected when compared with 1 in (B21). The compressibility κ must be evaluated numerically as explained in § B.3.3.

B.3.5. *Saturation pressure and latent heat of condensation of water vapour*

For completeness, we note that the saturation pressure of water vapour $p_{\rm sat}(T)$ can be obtained from the Clausius–Clapeyron equation

$$\frac{{\rm d}\ln p_{\rm sat}}{{\rm d}\ln T} = \frac{L}{RT}. \tag{B27}$$

The latent heat of condensation $L(T)$ can be determined by considering a reversible cycle in which vapour on the coexistence curve is heated isobarically from reference temperature T_c to T, compressed isothermally from $p_{\rm sat}(T_c)$ to $p_{\rm sat}(T)$, condensed to the liquid phase, cooled isobarically from T to T_c, allowed to expand isothermally from $p_{\rm sat}(T)$ to $p_{\rm sat}(T_c)$ and then finally evaporated, thus returning to the initial state. For the total entropy change to be zero, we must have

$$\frac{L}{T} = \frac{L_c}{T_c} - R\ln\frac{p_{\rm sat}(T)}{p_{\rm sat}(T_c)} + (c_v - c_l)\ln\frac{T}{T_c}. \tag{B28}$$

Substituting this into (B27) and integrating, we find that

$$R\ln\frac{p_{\rm sat}(T)}{p_{\rm sat}(T_c)} = \frac{L_c}{T_c}\left(1 - \frac{T_c}{T}\right) + (c_v - c_l)\left(\ln\frac{T}{T_c} - 1 + \frac{T_c}{T}\right), \tag{B29}$$

while by eliminating $p_{\rm sat}$ between (B28) and (B29), we have

$$L = L_c + (c_v - c_l)(T - T_c). \tag{B30}$$

Appendix C. Stability analysis to quadratic order

In this appendix, we present an expansion of the buoyancy force on a displaced fluid element, (2.5), for an equilibrium close to marginal linear stability, to quadratic order in the displacement. We show that the results are consistent with the conclusions of § 2.5.

We take the typical scales of variation of s, χ, P and ρ in z to be the same, denoted H, but we take the two terms that appear in the scalar product in (2.5) to have opposite signs and mostly cancel, so that the equilibrium is close to marginal linear stability, i.e.

$$\mathcal{L} \sim \epsilon \frac{\rho}{H} > 0, \tag{C1}$$

where $0 < \epsilon \ll 1$. Expanding (2.5) in $\delta z = z_2 - z_1$ yields

$$\frac{F}{gV_2} = -\mathcal{L}\delta z + \mathcal{N}\delta z^2 + \mathcal{O}(\delta z^3), \tag{C2}$$

where

$$\begin{aligned}\mathcal{N} &= -\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}z} + \frac{\mathrm{d}P}{\mathrm{d}z}\left(\frac{\mathrm{d}s}{\mathrm{d}z}\frac{\partial^2\rho}{\partial P\partial s} + \frac{\mathrm{d}\chi}{\mathrm{d}z}\frac{\partial^2\rho}{\partial P\partial \chi}\right)\\ &= \frac{\mathrm{d}P}{\mathrm{d}z}\left(\frac{\mathrm{d}s}{\mathrm{d}z}\frac{\partial^2\rho}{\partial P\partial s} + \frac{\mathrm{d}\chi}{\mathrm{d}z}\frac{\partial^2\rho}{\partial P\partial \chi}\right) + \mathcal{O}(\epsilon),\end{aligned} \tag{C3}$$

where we have used that $\mathrm{d}\mathcal{L}/\mathrm{d}z \sim \epsilon\rho/H^2$ by (C1).

Unless both s and χ increase with height [in which case (C1) demands that their gradients each be small], $\mathcal{N} \sim \rho/H^2$ and does not generally vanish at marginal stability. The equilibrium is therefore stable to linear perturbations with $\delta z \ll \mathcal{L}/\mathcal{N} \sim \epsilon H$ but unstable to nonlinear ones with $\epsilon H \ll \delta z \ll H$ [the latter condition ensuring that the $\mathcal{O}(\delta z^3)$ terms in (C2) are negligible]. Replacing partial derivatives by their expressions in 2D MHD, (C3) becomes

$$\mathcal{N} = -\frac{\rho g c_s^2}{c^4}(2-\gamma)\frac{\mathrm{d}\ln s}{\mathrm{d}z} + \mathcal{O}(\epsilon), \tag{C4}$$

which implies that $\mathcal{N} < 0$ for a stabilising entropy gradient $\mathrm{d}\ln s/\mathrm{d}z > 0$ (and $\mathrm{d}\ln\chi/\mathrm{d}z < 0$), and therefore the atmosphere is metastable to downwards displacements (provided $\gamma < 2$). Conversely, the atmosphere is metastable to upwards displacements if it has a destabilising entropy gradient but stabilising gradient of magnetic flux. These are the same conclusions as would be obtained by determining the direction of metastable displacements as the ones in the direction that the ratio s/χ decreases, as should be the case according to the argument in § 2.5.

Appendix D. Available energy of a two-phase atmosphere

In this appendix, we consider the simple case of a 'two-phase' atmosphere in which a mass m_s of fluid with $\chi = 0$ ($\beta = \infty$) is situated initially below a mass m_χ of fluid with

$s = 0$ ($\beta = 0$), for which the available energy can be determined analytically. This case is of pedagogical value as it illustrates why the available energy is always a small fraction of the total.

According to (2.13), the large-β fluid experiences a destabilising force when displaced upwards into the small-β fluid (indeed, the destabilising force is the greatest possible, as the difference in compressibility is maximal). Intuitively, therefore, we expect the minimum-energy state to be obtainable by re-stacking the atmosphere such that the large-β fluid sits above the small-β fluid. We consider the most optimistic case from the perspective of available energy – that of an atmosphere initially at marginal linear stability, i.e. $\mathcal{L} = 0$ – which means that, by (2.7), s and χ are piecewise-constant functions of m

$$s = \begin{cases} 0, & m < m_\chi, \\ s_0, & m_\chi < m < m_t, \end{cases} \qquad \chi = \begin{cases} \chi_0, & m < m_\chi, \\ 0, & m_\chi < m < m_t, \end{cases} \tag{D1a,b}$$

where $m_t = m_s + m_\chi$ is the total mass of the atmosphere. The interface of the two phases at $m = m_\chi$ is stable provided the density of the fluid above it is smaller than the density of the fluid below – the case of marginal linear stability is the one where the fluid density is continuous at the interface. By (2.26), this implies that

$$\frac{s_0}{\chi_0} = \frac{1}{\sqrt{2}}(m_\chi g)^{1/\gamma - 1/2}. \tag{D2}$$

The change in energy ΔE when the entire mass m_s of large-β fluid is moved from the bottom to the top may then be shown straightforwardly to satisfy

$$\frac{\Delta E}{E_\chi} = C(r^{2-1/\gamma} - (1+r)^{2-1/\gamma} + 1) + (1+r)^{3/2} - 1 - r^{3/2}, \tag{D3}$$

where $r = m_s/m_\chi$, $C = 3\gamma^2/4(\gamma - 1)(2\gamma - 1)$ and

$$E_\chi = \frac{2}{3}(2m_\chi^3 g)^{1/2}\chi_c, \tag{D4}$$

is the total energy of the small-β fluid at the top of the atmosphere, which is related to the total energy of the atmosphere E_0 via

$$\frac{E_0}{E_\chi} = 1 + C[(1+r)^{2-1/\gamma} - 1]. \tag{D5}$$

In the limit of $r \to 0$, (D3) and (D5) give

$$\frac{\Delta E}{E_\chi} = \frac{\Delta E}{E_0} = -\frac{3r}{8} = -\frac{3m_s}{8m_t}. \tag{D6}$$

Thus, the fractional energy change of a small-β atmosphere at marginal linear stability when a small parcel of large-β fluid rises through it is proportional to the mass fraction of the fluid moved, with a proportionality factor of order unity. However, returns in the fractional energy change diminish rapidly as the amount of mass moved increases. In figure 34(*a*), we visualise ΔE normalised by E_0 (black line) and E_χ (blue line) as a function of the mass fraction m_s/m_t of large-β fluid in the atmosphere. We observe that, despite its order-unity initial gradient of $-3/8$, $\Delta E/E_0$ shallows rapidly, reaching a minimum of

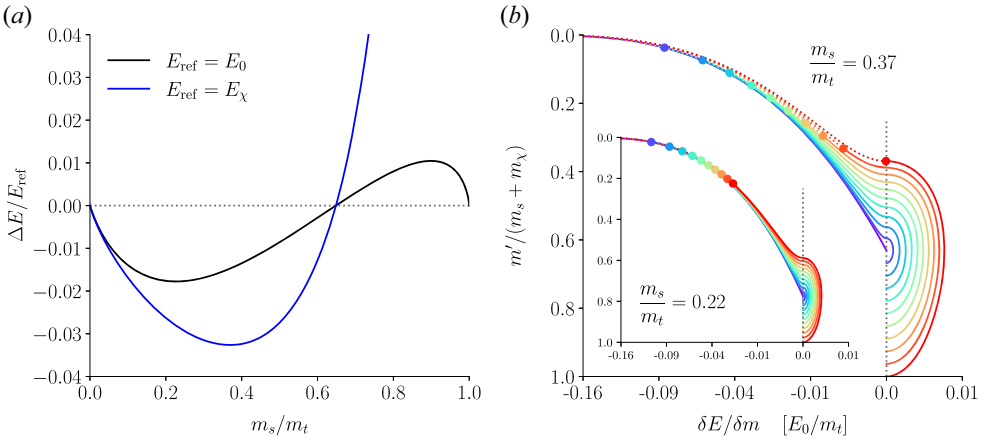

FIGURE 34. (*a*) The available energy of the 'two-phase' atmosphere considered in Appendix D, as a function of the mass m_s of fluid with $\beta = \infty$. The black line normalises the available energy by the total initial potential energy, the blue line by the initial potential energy of the fluid with $\beta = 0$ only. (*b*) The energy liberated per unit mass of moving fluid when a slice moves from its initial position to a new location where the supported mass is m'. The main plot shows the case of $m_s/m_t = 0.37$, which corresponds to the largest possible available energy (minimum of the blue line in (*a*)); the inset shows the case of $m_s/m_t = 0.37$, which corresponds to the largest possible fractional available energy [minimum of the black line in (*a*)].

$-1.8\,\%$ where the large-β fluid has a mass fraction of $22\,\%$. If the mass fraction of the large-β is greater than this, the fractional energy liberated in moving the large-β fluid is smaller, and, for a mass fraction of greater than $65\,\%$, moving the large-β fluid incurs an energetic cost.

To elucidate the reason for the diminishing return, we consider the rearrangement as a sequence of displacements of parcels of the large-β fluid from just below the interface to new positions at the top of the atmosphere. Figure 34*b* shows, for a sequence of parcel displacements represented by lines coloured from purple (the first parcel to move) to red (the last parcel to move), the energy δE that is liberated when a mass δm of the large-β fluid is moved from the (evolving) position of the interface to a new, smaller, supported mass m' (end points of the particle motions are shown as filled circles). The decrease in the magnitude of $\delta E/\delta m$ for subsequent parcels is seen to occur because fluid parcels obey an exclusion principle – two slices cannot have the same supported mass. This means that a parcel of fluid that has already risen from the bottom to the top of the atmosphere will prevent the next parcel from rising to the same height, thus limiting the difference in total pressure that this parcel experiences over its motion, and so reducing the work done on the parcel by the buoyancy force (recall that for an atmosphere at marginal linear stability, the latter is proportional to the ratio of pressures at the initial and final positions).

A second effect visible in figure 34*b* is that the motion of the first fluid parcel stabilises the atmosphere somewhat, because it moves the interface between the large- and small-β fluid downwards, where the total pressure is greater (the constants s_0 and χ_0 were chosen such that the large- and small-β fluids had the same density at $P = m_\chi g$; at larger pressure, the more compressible, large-β fluid underneath the interface is denser). The buoyancy force on a rising parcel is downwards (restoring) until it passes the point of neutral buoyancy at the original position of the interface. If the work required to overcome the restoring force is greater than the energy liberated when the parcel moves to the greatest

height permitted by the exclusion principle, then the reassignment incurs a net energetic cost. This situation occurs for $m_s \gtrsim 0.37m_t$: for m_s larger than this, $\Delta E/E_\chi$ rises with m_s (figure 34*a*), becoming positive for $m_s \gtrsim 0.65m_t$. $\Delta E/E_0$ reaches a minimum at smaller m_s/m_t because the small increase in available energy is outweighed by the energetic cost incurred by increasing the total mass of the atmosphere.

Appendix E. A necessary condition for 'one-to-many' optimal assignment

We can deduce a necessary condition for the one-to-many optimal assignment described in § 3.3.2 in a similar manner to the one in which we deduced the equal-density condition (3.14) for 'many-to-one' assignments in § 3.3.1. Let us suppose that we are given the optimal assignment of all slices apart from those with initial supported mass $m_i = m_a + \delta m_i$. Their contribution to the total energy of the atmosphere is

$$\begin{aligned}\delta E &= \Delta m \sum_i \mathcal{E}(m_{\sigma(i)}, m_a + \delta m_i) \\ &= \Delta m \sum_i \left[\mathcal{E}(m_{\sigma(i)}, m_a) + \delta m_i \frac{\partial \mathcal{E}}{\partial \mu}(m_{\sigma(i)}, m_a) + \mathcal{O}(\delta m_i^2) \right], \end{aligned} \tag{E1}$$

where

$$\frac{\partial \mathcal{E}}{\partial \mu} = \frac{1}{\gamma - 1} \rho^{\gamma - 1} \frac{\mathrm{d}s^\gamma}{\mathrm{d}\mu} + \frac{1}{2} \rho \frac{\mathrm{d}\chi^2}{\mathrm{d}\mu}. \tag{E2}$$

Thus, a necessary (but not sufficient) condition for the optimal assignment to be one-to-two is that $\partial \mathcal{E}/\partial \mu$ must be equal (up to a difference proportional to Δm) when evaluated (with $s = s(\mu)$ and $\chi = \chi(\mu)$) at each of the two different supported masses $m_{\sigma(i)}$ to which slices in the vicinity of m_a are to be assigned. If this is not the case for any δm_i in the given range, then (E1) is minimised to leading order in δm by assigning the slice with largest initial supported mass to the new supported mass for which $\partial \mathcal{E}/\partial \mu$ is smallest, and so on – this will always be a one-to-one assignment, as $\partial \mathcal{E}/\partial \mu$ is continuous in $m_{\sigma(i)}$ for fixed m_a.

Denoting the density of slice i at the two values of $m_{\sigma(i)}$ by ρ and ρ', the condition for a one-to-two mapping is (cf. equation (3.14))

$$\frac{1}{\gamma - 1} (\rho'^{\gamma - 1} - \rho^{\gamma - 1}) \frac{\mathrm{d}s^\gamma}{\mathrm{d}m} + \frac{1}{2} (\rho' - \rho) \frac{\mathrm{d}\chi^2}{\mathrm{d}m} = \mathcal{O}(\Delta m). \tag{E3}$$

From this expression, we see that, for any given ρ, there can be at most one solution for ρ' in addition to the trivial $\rho' = \rho$. The existence of a second solution requires that gradients of s and χ with m have opposite signs in the initial profile. Thus, while an optimal assignment may in general be one-to-two, as in figure 11, it cannot be one-to-X with $X > 2$.

Appendix F. Viscous relaxation in the case of downwards metastability, (2.31)

Figures 35 and 36 are analogues of figures 17 and 18 for the equilibrium defined by (2.31). The dimensionless numbers Re, Pr_m and Pr_t are the same as in § 4.2, but $z_0 = 2.25$ in (4.5). In figure 36, We again compare $u_0 = 0.1$, $u_0 = 0.05$ and $u_0 = 0.025$; in this case, this corresponds to $E_{\mathrm{kin},0} \simeq 0.1E_{\mathrm{avail}}$, $E_{\mathrm{kin},0} \simeq 0.03E_{\mathrm{avail}}$ and $E_{\mathrm{kin},0} \simeq 0.01E_{\mathrm{avail}}$, respectively. Again, we observe the formation of a long-lived 2D state (subject to slow diffusion) that is consistent with the minimum-energy state obtained in § 3.3.2 (see lower panels of figures 35 and 36), with the quality of agreement between theory and simulation being better at $E_{\mathrm{kin},0} \simeq 0.1E_{\mathrm{avail}}$ than $E_{\mathrm{kin},0} \simeq 0.01E_{\mathrm{avail}}$.

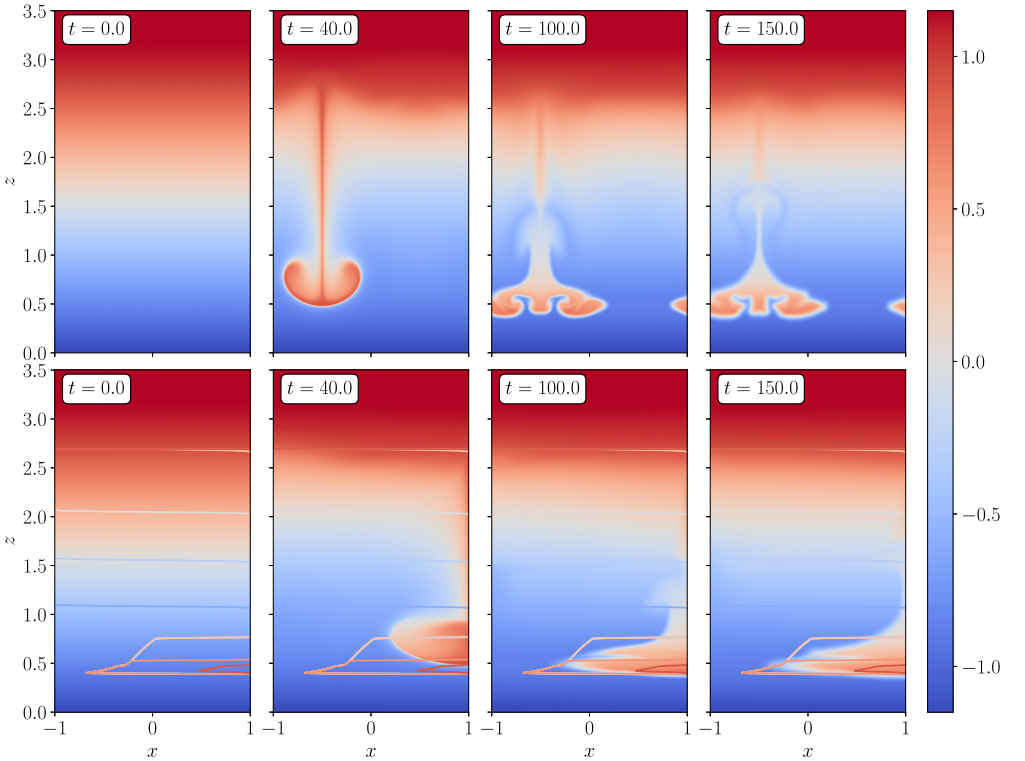

FIGURE 35. As for figure 17, but for the equilibrium defined by (2.31).

Appendix G. Alternative formulation of the Lynden-Bell statistical mechanics

For completeness, we note in this appendix how Lynden-Bell statistical mechanics may be formulated for the joint probability-distribution function for s and χ, rather than for μ as in § 5.2. In that case, we define $\mathcal{P}(m, s, \chi)\,\mathrm{d}s\,\mathrm{d}\chi$ to be the probability that a fluid element at supported mass m has specific entropy and flux in the ranges s to $s + \mathrm{d}s$ and χ to $\chi + \mathrm{d}\chi$, respectively. We obtain $\mathcal{P}(m, s, \chi)$ by maximising the number of microstates with which it is consistent, after coarse graining. This corresponds to maximising the mixing entropy (Robert & Sommeria 1991)

$$S = -\int \mathrm{d}s \int \mathrm{d}\chi \int \mathrm{d}m \mathcal{P}(m, s, \chi) \ln \mathcal{P}(m, s, \chi). \tag{G1}$$

Maximisation of S is subject to the constraints of fixed total probability (i.e. the normalisation of P)

$$\int \mathrm{d}s \int \mathrm{d}\chi \mathcal{P}(m, s, \chi) = 1, \tag{G2}$$

fixed potential energy E_{pot}

$$\int \mathrm{d}m \int \mathrm{d}s \int \mathrm{d}\chi \mathcal{E}(mg, s, \chi)\mathcal{P}(m, s, \chi) = E_{\mathrm{pot}}, \tag{G3}$$

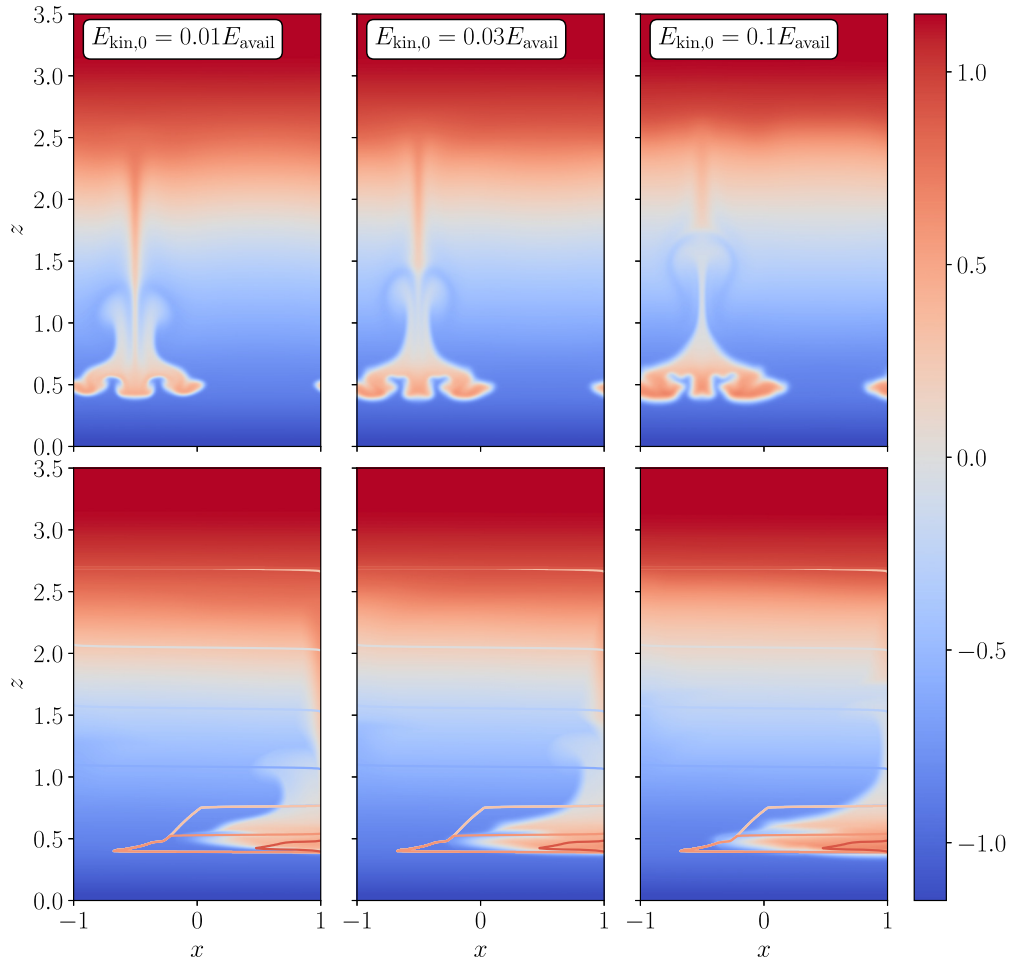

FIGURE 36. As for figure 18, but for the equilibrium defined by (2.31).

and, for all s and χ, a fixed mass of fluid $M(s,\chi)\mathrm{d}s\mathrm{d}\chi$ having both a value of s in the range s to $s+\mathrm{d}s$ and a value of χ in the range χ to $\chi+\mathrm{d}\chi$

$$\int \mathrm{d}m\mathcal{P}(m,s,\chi) = M(s,\chi). \tag{G4}$$

The equivalence of the above constraints to the corresponding ones in § 5.2 [(5.3), (5.4) and (5.5)] is readily demonstrated by substituting

$$\mathcal{P}(m,s,\chi) = \int \mathrm{d}\mu\mathcal{P}(m,\mu)\delta(s-s(\mu))\delta(\chi-\chi(\mu)), \tag{G5}$$

and evaluating integrals over s and χ. Substitution of (G5) into the expression (G1) for the thermodynamic entropy S yields

$$S = -\int \mathrm{d}\mu \int \mathrm{d}m\mathcal{P}(m,\mu)\ln\int \mathrm{d}\mu'\mathcal{P}(m,\mu')\delta(s(\mu)-s(\mu'))\delta(\chi(\mu)-\chi(\mu')). \tag{G6}$$

This reduces to (5.2) provided that each value of μ has a different pair of values of s and χ: in that case, both delta functions on the second line are simultaneously non-zero only

when $\mu' = \mu$. $\mathcal{P}(m, \mu)$ can then be brought outside of the μ' integral, leaving (up to an additive constant that, after application of (5.5), does not depend on $\mathcal{P}(m, \mu)$)

$$S = -\int \mathrm{d}\mu \int \mathrm{d}m \mathcal{P}(m, \mu) \ln \mathcal{P}(m, \mu), \tag{G7}$$

which is (5.2).

The solution of the constrained maximisation is (cf. (5.7))

$$\mathcal{P}(m, s, \chi) = \frac{\mathrm{e}^{-\beta_T[\mathcal{E}(mg,s,\chi)-\psi(s,\chi)]}}{\displaystyle\int \mathrm{d}s' \int \mathrm{d}\chi' \, \mathrm{e}^{-\beta_T[\mathcal{E}(m,s',\chi')-\psi(s',\chi')]}}, \tag{G8}$$

where the Lagrange multipliers β_T and $\psi(s, \chi)$ are determined by (5.4) and (5.5), respectively, for given E and $M(s, \chi)$.

Appendix H. The minimum-energy state of § 3 as the $\beta_T \to \infty$ limit of Lynden-Bell statistical mechanics

On physical grounds, we expect that the $\beta_T \to \infty$ limit of (5.7), i.e. the limit of zero statistical mechanical temperature, corresponds to the smallest energy permitted by the system, i.e. the result of solving the LSA problem (§ 3). In this appendix, we verify that the solution to the LSA problem is indeed recoverable from (5.7) in the $\beta_T \to \infty$ limit.

H.1. *Discrete problem: non-degenerate case*

Because the LSA problem is discrete, it is strictly the $\beta_T \to \infty$ limit of (5.3), (5.4), (5.5) and (5.7) only after they are discretised over a small but finite scale Δm (we consider reversing the order of the $\Delta m \to 0$ and $\beta_T \to \infty$ limits in § H.3). We adopt the economical notation $\mathcal{P}_{ij} \equiv \mathcal{P}(m_j, m_i)$, $\mathcal{E}_{ij} \equiv \mathcal{E}(m_j, s_i, \chi_i) = \mathcal{E}(m_j, m_i)$ and $\psi_i \equiv \psi(m_i)$. Converting integrals to sums, (5.3), (5.5) (5.4) and (5.8) become

$$\Delta m \sum_i \mathcal{P}_{ij} = 1, \tag{H1}$$

$$\Delta m \sum_j \mathcal{P}_{ij} = 1, \tag{H2}$$

$$\Delta m^2 \sum_{ij} \mathcal{E}_{ij} \mathcal{P}_{ij} = E_{\mathrm{pot}}, \tag{H3}$$

and

$$\mathcal{P}_{ij} = \frac{1}{\Delta m} \frac{\mathrm{e}^{-\beta_T(\mathcal{E}_{ij}-\psi_i)}}{\displaystyle\sum_k \mathrm{e}^{-\beta_T(\mathcal{E}_{kj}-\psi_k)}}, \tag{H4}$$

respectively. We define $\tilde{\mathcal{E}}_{ij} = \mathcal{E}_{ij} - a_j - b_i$ and $\tilde{\psi}_i = \psi_i - b_i$, whence

$$\mathcal{P}_{ij} = \frac{1}{\Delta m} \frac{\mathrm{e}^{-\beta_T(\tilde{\mathcal{E}}_{ij}-\tilde{\psi}_i)}}{\displaystyle\sum_k \mathrm{e}^{-\beta_T(\tilde{\mathcal{E}}_{kj}-\tilde{\psi}_k)}}. \tag{H5}$$

The advantage to introducing the 'tilded' variables in (H5) is that we may always choose the functions b_i and a_j to be such that $\tilde{\mathcal{E}}_{ij}$ is the normal form of the cost matrix $\mathcal{E}_{ij}$

(see (3.11)). As explained in § 3.2, this means that $\tilde{\mathcal{E}}_{ij} \geq 0$ and has at least one zero in each row and column, or, equivalently, there exists at least one bijection σ such that $\tilde{\mathcal{E}}_{i\sigma(i)} = 0$ for all i ($j = \sigma(i)$ constitutes an optimal assignment for the LSA problem).

In the simplest case, $\tilde{\mathcal{E}}_{ij}$ has exactly one zero in each row and column. We expect this to be the generic case, and, indeed, this is true for all the profiles examined in § 3.3. These zeros define the unique optimal assignment $j = \sigma(i)$ – taking $\tilde{\psi}_i = 0$ [the natural choice, as with only one zero in each row and column, the 'exclusion principle' is obeyed automatically and the constraint (H2) becomes superfluous] then implies that

$$\mathcal{P}_{ij} \to \begin{cases} 1/\Delta m & \text{if } \tilde{\mathcal{E}}_{ij} = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{H6}$$

Thus, in the $\beta_T \to \infty$ limit, the system is certain to found in the state given by the optimal assignment σ for the LSA problem. Equation (H6) straightforwardly satisfies (H2) – the sum picking out the single value of j for which $\mathcal{P}_{ij} \neq 0$ for any given i – which justifies formally our setting $\tilde{\psi}_i = 0$. Substituting (H6) into (H3) yields

$$E = \Delta m \sum_i \mathcal{E}_{i\sigma(i)} = E_{\min}. \tag{H7}$$

H.2. *Discrete problem: degenerate case*

In principle, $\tilde{\mathcal{E}}_{ij}$ may have more than one zero in each row and column: these may either be part of other optimal assignments or not be part of any optimal assignment. In such cases, (H6) becomes

$$\mathcal{P}_{ij} \to \frac{1}{\Delta m} \frac{a_{ij} F_i}{\sum_k a_{kj} F_k} \text{as } \beta_T \to \infty, \tag{H8}$$

where $a_{ij} = 1$ if $\tilde{\mathcal{E}}_{ij} = 0$ and $a_{ij} = 0$ otherwise, and $F_i = \mathrm{e}^{\beta_T \tilde{\psi}_i}$ is the quantity sometimes known as 'fugacity' [in writing (H8), we assumed that $\tilde{\psi}_i < 0$; we were free to do so because taking $\tilde{\psi}_i \to \tilde{\psi}_i + C$ for all i with C a constant leaves (H5) unchanged]. The values of F_i are determined by (H2), which requires that

$$\sum_j \frac{a_{ij} F_i}{\sum_k a_{kj} F_k} = 1, \quad \forall\, i, \tag{H9}$$

an equation that may be solved iteratively for F_i, and hence $\mathcal{P}_{ij}$. The possible outcomes of such a procedure are constrained by the Birkhoff–von Neumann theorem, which states that any doubly stochastic matrix – one whose rows and columns all sum to unity – can be expressed as a sum of permutation matrices with positive weights that also sum to unity. From (H1) and (H2), $\mathcal{P}_{ij}$ is a doubly stochastic matrix, so that

$$\mathcal{P}_{ij} = \frac{1}{\Delta m} \sum_k \theta_k S^{(k)}_{ij}, \quad \theta_k > 0, \quad \sum_k \theta_k = 1, \tag{H10}$$

where $S^{(k)}_{ij}$ are permutation matrices. In the $\beta_T \to \infty$ limit, $\mathcal{P}_{ij} \propto a_{ij}$ (H8), so $\{S^{(k)}_{ij}\}$ is the maximal set of permutations that vanish wherever $\tilde{\mathcal{E}}_{ij} > 0$, which we recognise as the

optimal assignments of the LSA problem: with $\sigma^{(k)}$ the kth optimal assignment, $S_{ij}^{(k)}$ is the matrix representation of $\sigma^{(k)}$, i.e. $S_{ij}^{(k)} = \delta_{j\sigma^{(k)}(i)}$. It follows that $\mathcal{P}_{ij} = 0$ for any ij that is not part of an optimal assignment, even if $\tilde{\mathcal{E}}_{ij} = 0$, and that

$$E_{\mathrm{pot}} = \Delta m \sum_{ijk} \mathcal{E}_{ij}\theta_k S_{ij}^{(k)} = \Delta m \sum_k \theta_k \sum_i \mathcal{E}_{i\sigma^{(k)}(i)} = E_{\mathrm{min}}. \tag{H11}$$

H.3. *Continuous problem*

We have so far determined that the discretised probability-density function $\mathcal{P}_{ij}$ reproduces the solution of the LSA problem (or some weighted combination of its solutions if they are not unique) as $\beta_T \to \infty$ at fixed Δm. This is not, however, the large-β_T limit of the continuous system (5.4), (5.5) and (5.7): discretising the integrals is valid only if Δm is small compared with the scale of variation of the continuous probability density $\mathcal{P}(m, \mu)$, which shrinks to zero as $\beta_T \to \infty$. Nonetheless, the $\beta_T \to \infty$ limit of the continuous Lynden-Bell equations is equal to the $\Delta m \to 0$ limit of the coarse-grained solution of the LSA problem, provided it is unique, as follows.

Let $\tilde{\mathcal{E}}(m, \mu)$ be the continuous limit of $\tilde{\mathcal{E}}_{ij}$ obtained by taking $\Delta m \to 0$. As $\beta_T \to \infty$, it is clear that

$$\mathcal{P}(m, \mu) \propto \mathrm{e}^{-\beta_T \tilde{\mathcal{E}}(m,\mu)}, \tag{H12}$$

becomes increasingly localised to the curve in (m, μ) space on which $\tilde{\mathcal{E}}(m, \mu) = 0$. The same is true for the coarse-grained discrete solution $\mathcal{P}_{ij}$ of the LSA problem, which we define by

$$\langle \mathcal{P} \rangle \equiv \frac{1}{2n} \int_{m-n\Delta m}^{m+n\Delta m} \mathrm{d}m' \int_{\mu-n\Delta m}^{\mu+n\Delta m} \mathrm{d}\mu' \sum_{ij} \mathcal{P}_{ij}\delta(\mu - m_i)\delta(m - m_j), \tag{H13}$$

with n a constant satisfying $1 \ll n \ll N$. As $\beta_T \to \infty$, both $\mathcal{P}(m, \mu)$ and $\mathcal{P}$ tend to a delta distribution centred on the optimal-solution curve. The most general form such a distribution can take is

$$A(m, \mu)\delta(F(m, \mu)), \tag{H14}$$

where F vanishes along the optimal-solution curve and, without loss of generality, can be taken to have $|\nabla F| = 1$ there (F contributes no further degrees of freedom in the expression (H14)). However, as we show in Appendix I, the function $A(m, \mu)$ is uniquely determined by the conditions (H1) and (H2), which must be satisfied by both $\langle P \rangle$ and P. It follows that their large-β_T limits are the same. A practical consequence of this is that one may determine the horizontal composition of 2D minimum-energy ground states from the $\beta_T \to \infty$ limit of $\mathcal{P}(m, \mu)$ (as in figure 7).

Appendix I. Uniqueness of doubly stochastic delta distributions

In this appendix, we show that the delta distribution

$$\mathcal{P}(x, y) = A(x, y)\delta(F(x, y)), \tag{I1}$$

where $A > 0$, and $F(x, y) = 0$ on a piecewise-once-differentiable curve $\mathcal{C}$, is unique under the conditions that

(i) $\mathcal{C}$ intersects every horizontal and vertical line in the unit square at least once;

(ii) $\mathcal{C}$ does not admit any cycles formed by traversing horizontal and vertical lines (in the context of (H13), this corresponds to uniqueness of the optimal assignment);

(iii) $\mathcal{P}(x, y)$ satisfies the integral constraints

$$\int_0^1 \mathrm{d}x \mathcal{P}(x, y) = 1, \tag{I2}$$

and

$$\int_0^1 \mathrm{d}y \mathcal{P}(x, y) = 1. \tag{I3}$$

Without loss of generality, we can take $|\nabla F| = 1$; any other choice amounts to a redefinition of A. Then, (I2) and (I3) become

$$\sum_i \frac{A(x_i(y), x)}{|t_x(x_i(y), y)|} = 1, \quad \sum_i \frac{A(x, y_i(x))}{|t_y(x, y_i(x))|} = 1, \tag{I4}$$

respectively, where $y_i(x)$ is the y-coordinate of the ith intersection of $\mathcal{C}$ with the vertical line through $(x,0)$, $x_i(y)$ is the x-coordinate of the ith intersection of $\mathcal{C}$ with the horizontal line through $(0, y)$, and

$$(t_x, t_y) = \left(\frac{\partial F}{\partial y}, -\frac{\partial F}{\partial x}\right), \tag{I5}$$

is the unit tangent vector to $\mathcal{C}$.

For each point on $\mathcal{C}$, (I4) each provide an equation relating A at that point to the value of A at each other point on $\mathcal{C}$ with the same x or y coordinate. In turn, (I4) provide one additional equation for each such point of intersection, relating A evaluated there to A evaluated at other points on $\mathcal{C}$ with the same x or y-coordinate, and so on for the points with which those points share an x or y coordinate. There are three possible outcomes for this proliferation of coupled equations: (*a*) a stage of the process is reached when all new equations only involve a single point on $\mathcal{C}$, and so the proliferation is arrested when $n + 1$ independent (prior to the specification of the tangent vectors) linear equations couple the values of A at n different points; (*b*) proliferation of equations ceases, but the n points on $\mathcal{C}$ form one or more cycles of the sort envisioned in (ii) and are therefore coupled by fewer than $n + 1$ equations; (*c*) proliferation of equations never ceases – in this case, infinitely many points are coupled, some of them arbitrarily close to each other (infinite cycle). By condition (ii), we restrict attention to case (*a*). Then, the first n of the $n + 1$ independent linear equations can be solved for the values of A at each of the n points, so the function $A(x, y)$ is indeed unique given conditions (i)–(iii), as claimed. The final equation gives a constraint on the tangent vectors, which must be satisfied in order for a delta distribution (I1) satisfying conditions (i)–(iii) to exist.

For clarity, we illustrate the procedure described above for the example curve $\mathcal{C}$ shown in figure 37.

I.1. *Case where vertical and horizontal lines intersect $\mathcal{C}$ once only – red point in figure 37*

In this case, (I4) become

$$\frac{A}{|t_y|} = 1, \quad \frac{A}{|t_x|} = 1 \implies A = |t_x| = |t_y| = \frac{1}{\sqrt{2}}, \tag{I6}$$

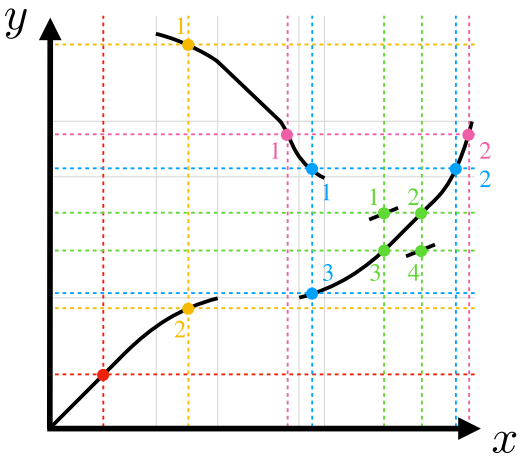

FIGURE 37. Visualisation of the construction of the delta distribution (I1) for the curve $\mathcal{C}$ by solving (I4). Coloured dashed lines indicate the relevant vertical and horizontal lines used in the evaluation of (I4). For ease of comparison between different parts of $\mathcal{C}$, the grey lines mark certain points at which $\mathcal{C}$ has discontinuities.

from which it follows that

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \pm 1 \tag{I7}$$

must be satisfied at the point in question. As claimed, we find the unique value of A and a condition on the tangent vector.

I.2. *Case where one vertical line intersects $\mathcal{C}$ twice – yellow points in figure 37*

With subscripts indicating evaluation at points according to the labelling scheme in figure 37, (I4) yield

$$\frac{A_1}{|t_{y1}|} = 1, \quad \frac{A_2}{|t_{y2}|} = 1, \quad \frac{A_1}{|t_{x1}|} + \frac{A_2}{|t_{x2}|} = 1, \tag{I8a–c}$$

from which it follows that

$$A_1 = |t_{y2}|, \quad A_2 = |t_{y2}|, \quad \left|\frac{\mathrm{d}y}{\mathrm{d}x}\right|_1 + \left|\frac{\mathrm{d}y}{\mathrm{d}x}\right|_2 = 1. \tag{I9a–c}$$

As claimed, we find the unique values of A_1 and A_2 and a condition on the tangent vector at the relevant points.

I.3. *Case where one horizontal line intersects $\mathcal{C}$ twice – pink points in figure 37*

In this case, (I4) yield

$$\frac{A_1}{|t_{x1}|} = 1, \quad \frac{A_2}{|t_{x2}|} = 1, \quad \frac{A_1}{|t_{y1}|} + \frac{A_2}{|t_{y2}|} = 1, \tag{I10a–c}$$

from which it follows that

$$A_1 = |t_{x2}|, \quad A_2 = |t_{x2}|, \quad \left|\frac{\mathrm{d}y}{\mathrm{d}x}\right|_1^{-1} + \left|\frac{\mathrm{d}y}{\mathrm{d}x}\right|_2^{-1} = 1. \tag{I11a–c}$$

Again, we find the unique values of A_1 and A_2 and a condition on the tangent vector at the relevant points.

I.4. *Case where both one vertical and one horizontal line intersects $\mathcal{C}$ – blue points in figure 37*

In this case, three points are coupled by (I4), which yield

$$\frac{A_2}{|t_{x2}|} = 1, \quad \frac{A_3}{|t_{y3}|} = 1, \quad \frac{A_1}{|t_{x1}|} + \frac{A_3}{|t_{x3}|} = 1, \quad \frac{A_1}{|t_{y1}|} + \frac{A_2}{|t_{y2}|} = 1, \tag{I12a–d}$$

from which it follows that the values of A_i are

$$A_1 = \left(1 - \frac{|t_{x2}|}{|t_{y2}|}\right)|t_{y1}|, \quad A_2 = |t_{x2}|, \quad A_3 = |t_{y3}|, \tag{I13a–c}$$

and the condition on $\mathcal{C}$ is

$$\left|\frac{\mathrm{d}y}{\mathrm{d}x}\right|_3 = \left|\frac{\mathrm{d}y}{\mathrm{d}x}\right|_1 \left[\left|\frac{\mathrm{d}y}{\mathrm{d}x}\right|_1^{-1} + \left|\frac{\mathrm{d}y}{\mathrm{d}x}\right|_2^{-1} - 1\right]. \tag{I14}$$

I.5. *Case in which $\mathcal{C}$ has a simple cycle – green points in figure 37*

To illustrate how uniqueness fails if condition (ii) is not satisfied, we consider as a final example the case where four points form a cycle along horizontal and vertical lines. Equations (I4) yield

$$\frac{A_1}{|t_{y1}|} + \frac{A_2}{|t_{y2}|} = 1, \quad \frac{A_1}{|t_{x1}|} + \frac{A_3}{|t_{x3}|} = 1, \quad \frac{A_3}{|t_{y3}|} + \frac{A_4}{|t_{y4}|} = 1, \quad \frac{A_2}{|t_{x2}|} + \frac{A_4}{|t_{x4}|} = 1. \tag{I15a–d}$$

These equations can be reduced to the matrix equation

$$\begin{pmatrix} \dfrac{1}{|t_{y1}|} & -\dfrac{|t_{x2}|}{|t_{y2}||t_{x4}|} \\ -\dfrac{|t_{x3}|}{|t_{y3}||t_{x1}|} & \dfrac{1}{|t_{y4}|} \end{pmatrix} \begin{pmatrix} A_1 \\ A_4 \end{pmatrix} = \begin{pmatrix} 1 - \dfrac{|t_{x2}|}{|t_{y2}|} \\ 1 - \dfrac{|t_{x3}|}{|t_{y3}|} \end{pmatrix}, \tag{I16}$$

which has unique solutions for A_1 and A_4 unless the determinant of the matrix is zero, which requires

$$\left|\frac{\mathrm{d}y}{\mathrm{d}x}\right|_1 \left|\frac{\mathrm{d}y}{\mathrm{d}x}\right|_4 = \left|\frac{\mathrm{d}y}{\mathrm{d}x}\right|_2 \left|\frac{\mathrm{d}y}{\mathrm{d}x}\right|_3. \tag{I17}$$

In this case, the A_i are not determined uniquely. It is straightforwardly verified that (I17) is precisely the condition for the four points separated by an infinitesimal distance along $\mathcal{C}$ from the green ones to also be on a cycle of horizontal and vertical lines. Thus, it appears that condition (ii) is actually too strong a condition to guarantee uniqueness of (I1): cycles are permissible if they only exist in a set of measure zero in x and y. It follows that cases where the optimal solution to the assignment problem is non-unique on a set of measure zero only do not violate the conclusion of § H.3.

REFERENCES

BALBUS, S.A. 2000 Stability, instability, and 'backward' transport in stratified fluids. *Astrophys. J.* **534**, 420.

BHAT, P., ZHOU, M. & LOUREIRO, N.F. 2021 Inverse energy transfer in decaying, three-dimensional, non-helical magnetic turbulence due to magnetic reconnection. *Mon. Not. R. Astron. Soc.* **501**, 3074.

BURKARD, R., DELL'AMICO, M. & MARTELLO, S. 2012 *Assignment Problems*. Society for Industrial and Applied Mathematics.

CHAVANIS, P.H. 2003 Gravitational instability of isothermal and polytropic spheres. *Astron. Astrophys.* **401**, 15.

CHEN, P.F. 2011 Coronal mass ejections: models and their observational basis. *Living Rev. Sol. Phys.* **8**, 1.

CONNOR, J.W., HASTIE, R.J. & TAYLOR, J.B. 1979 High mode number stability of an axisymmetric toroidal plasma. *Proc. R. Soc. Lond.* **365**, 1.

COWLEY, S.C. & ARTUN, M. 1997 Explosive instabilities and detonation in magnetohydrodynamics. *Phys. Rep.* **283**, 185.

COWLEY, S.C., COWLEY, B., HENNEBERG, S.A. & WILSON, H.R. 2015 Explosive instability and erupting flux tubes in a magnetized plasma. *Proc. R. Soc. Lond.* **471**, 20140913.

CRANMER, S.R. & WINEBARGER, A.R. 2019 The properties of the solar corona and its connection to the solar wind. *Annu. Rev. Astron. Astrophys.* **57**, 157.

EWART, R.J., NASTAC, M.L. & SCHEKOCHIHIN, A.A. 2023 Non-thermal particle acceleration and power-law tails via relaxation to universal Lynden-Bell equilibria. *J. Plasma Phys.* **89**, 905890516.

FRIEDRICHS, K.O. 1960 Pinch buckling. *Rev. Mod. Phys.* **32**, 889.

GARAUD, P. 2018 Double-diffusive convection at low Prandtl number. *Annu. Rev. Fluid Mech.* **50**, 275.

HAM, C.J., COWLEY, S.C., BROCHARD, G. & WILSON, H.R. 2018 Nonlinear ballooning modes in tokamaks: stability and saturation. *Plasma Phys. Control. Fusion* **60**, 075017.

HENDER, T.C., WESLEY, J.C., BIALEK, J., BONDESON, A., BOOZER, A.H., BUTTERY, R.J., GAROFALO, A., GOODMAN, T.P., GRANETZ, R.S., GRIBOV, Y., *et al.* 2007 Progress in the ITER physics basis, chapter 3: MHD stability, operational limits and disruptions. *Nucl. Fusion* **47**, S128.

HIERONYMUS, M. & NYCANDER, J. 2015 Finding the minimum potential energy state by adiabatic parcel rearrangements with a nonlinear equation of state: an exact solution in polynomial time. *J. Phys. Oceanogr.* **45**, 1843.

HOSKING, D.N. & SCHEKOCHIHIN, A.A. 2021 Reconnection-controlled decay of magnetohydrodynamic turbulence and the role of invariants. *Phys. Rev.* X **11**, 041005.

HUGHES, D.W. 1985 Magnetic buoyancy instabilities for a static plane layer. *Geophys. Astrophys. Fluid Dyn.* **32**, 273.

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

HURRICANE, O.A., FONG, B.H. & COWLEY, S.C. 1997 Nonlinear magnetohydrodynamic detonation: part I. *Phys. Plasmas* **4**, 3565.

ITER PHYSICS EXPERT GROUP ON CONFINEMENT AND TRANSPORT, ITER PHYSICS EXPERT GROUP ON CONFINEMENT MODELLING & ITER PHYSICS BASIS (Eds.) 1999 ITER physics basis, chapter 2: plasma confinement and transport. *Nucl. Fusion* **39**, 2175.

KIRK, A., KOCH, B., SCANNELL, R., WILSON, H.R., COUNSELL, G., DOWLING, J., HERRMANN, A., MARTIN, R. & WALSH, M. 2006 Evolution of filament structures during edge-localized modes in the MAST tokamak. *Phys. Rev. Lett.* **96**, 185001.

KIRK, A., WILSON, H.R., COUNSELL, G.F., AKERS, R., ARENDS, E., COWLEY, S.C., DOWLING, J., LLOYD, B., PRICE, M. & WALSH, M. 2004 Spatial and temporal structure of edge-localized modes. *Phys. Rev. Lett.* **92**, 245002.

KUHN, H.W. 1955 The hungarian method for the assignment problem. *Nav. Res. Logist. Quart.* **2**, 83.

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

LORENZINI, R., MARTINES, E., PIOVESAN, P., TERRANOVA, D., ZANCA, P., ZUIN, M., ALFIER, A., BONFIGLIO, D., BONOMO, F., CANTON, A., *et al.* 2009 Self-organized helical equilibria as a new paradigm for ohmically heated fusion plasmas. *Nat. Phys.* **5**, 570.

LYNDEN-BELL, D. 1967 Statistical mechanics of violent relaxation in stellar systems. *Mon. Not. R. Astron. Soc*. **136**, 101.

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

MILLER, J. 1990 Statistical mechanics of Euler equations in two dimensions. *Phys. Rev. Lett*. **65**, 2137.

MILLER, J., WEICHMAN, P.B. & CROSS, M.C. 1992 Statistical mechanics, Euler's equation, and Jupiter's Red Spot. *Phys. Rev.* A **45**, 2328.

MUNKRES, J. 1957 Algorithms for the assignment and transportation problems. *J. Soc. Ind. Appl. Maths* **5**, 32.

PENCIL CODE COLLABORATION, *et al.* 2021 The Pencil Code, a modular MPI code for partial differential equations and particles: multipurpose and multiuser-maintained. *J. Open Source Softw*. **6**, 2807.

QUATAERT, E. 2008 Buoyancy instabilities in weakly magnetized low-collisionality plasmas. *Astrophys. J*. **673**, 758.

ROBERT, R. 1991 A maximum-entropy principle for two-dimensional perfect fluid dynamics. *J. Stat. Phys*. **65**, 531.

ROBERT, R. & SOMMERIA, J. 1991 Statistical equililbrium states for two-dimensional flows. *J. Fluid Mech*. **229**, 291.

ROGERS, R.R. & YAU, M.K. 1996 *A Short Course in Cloud Physics*. Elsevier Science.

RUTHERFORD, P.H., FURTH, H.P. & ROSENBLUTH, M.N. 1971Non-linear kink and tearing mode effects in Tokamaks. In *Plasma Physics and Controlled Nuclear Fusion Research, Volume II*, p. 553. https://www.osti.gov/biblio/4668895.

SCHWARZSCHILD, K. 1906 On the equilibrium of the Sun's atmosphere. *Gottinger Nachr*. **195**, 41.

SINGH, M.S. & O'NEILL, M.E. 2022 The climate system and the second law of thermodynamics. *Rev. Mod. Phys*. **94**, 015001.

STANSIFER, E.M., O'GORMAN, P.A. & HOLT, J.I. 2017 Accurate computation of moist available potential energy with the Munkres algorithm. *Q. J. R. Meteorol. Soc*. **143**, 288.

STEVENS, B. 2005 Atmospheric moist convection. *Annu. Rev. Earth Planet. Sci*. **33**, 605.

SU, Z. & INGERSOLL, A.P. 2016 On the minimum potential energy state and the eddy size-constrained APE density. *J. Phys. Oceanogr*. **46**, 2663.

TAYLOR, J.B. 1974 Relaxation of toroidal plasma and generation of reverse magnetic fields. *Phys. Rev. Lett*. **33**, 1139.

TAYLOR, J.B. 1986 Relaxation and magnetic reconnection in plasmas. *Rev. Mod. Phys*. **58**, 741.

WHITE, R.B., MONTICELLO, D.A., ROSENBLUTH, M.N. & WADDELL, B.V. 1977 Saturation of the tearing mode. *Phys. Fluids* **20**, 800.

WILSON, H.R. & COWLEY, S.C. 2004 Theory for explosive ideal magnetohydrodynamic instabilities in plasmas. *Phys. Rev. Lett*. **92**, 175006.

ZHOU, M., BHAT, P., LOUREIRO, N.F. & UZDENSKY, D.A. 2019 Magnetic island merger as a mechanism for inverse magnetic energy transfer. *Phys. Rev. Res*. **1**, 012004.