

Linear and non-linear evolution of the vertical shear instability in accretion discs

Richard P. Nelson,¹★ Oliver Gressel^{1,2}★ and Orkan M. Umurhan^{1,3}★

¹*Astronomy Unit, Queen Mary University of London, Mile End Road, London E1 4NS, UK*

²*NORDITA, KTH Royal Institute of Technology and Stockholm University, Roslagstullsbacken 23, SE-106 91 Stockholm, Sweden*

³*School of Natural Sciences, University of California, Merced, 5200 North Lake Rd, Merced, CA 95343, USA*

Accepted 2013 August 6. Received 2013 August 6; in original form 2012 September 12

ABSTRACT

We analyse the stability and non-linear dynamics of power-law accretion disc models. These have mid-plane densities that follow radial power laws and have either temperature or entropy distributions that are strict power-law functions of cylindrical radius, R . We employ two different hydrodynamic codes to perform high-resolution 2D axisymmetric and 3D simulations that examine the long-term evolution of the disc models as a function of the power-law indices of the temperature or entropy, the disc scaleheight, the thermal relaxation time of the fluid and the disc viscosity. We present an accompanying stability analysis of the problem, based on asymptotic methods, that we use to guide our interpretation of the simulation results. We find that axisymmetric disc models whose temperature or entropy profiles cause the equilibrium angular velocity to vary with height are unstable to the growth of perturbations whose most obvious character is modes with horizontal and vertical wavenumbers that satisfy $|k_R/k_Z| \gg 1$. Instability occurs only when the thermodynamic response of the fluid is isothermal, or the thermal evolution time is comparable to or shorter than the local dynamical time-scale. These discs appear to exhibit the Goldreich–Schubert–Fricke or ‘vertical shear’ linear instability. Closer inspection of the simulation results uncovers the growth of two distinct modes. The first are characterized by very short radial wavelength perturbations that grow rapidly at high latitudes in the disc, and descend down towards the mid-plane on longer time-scales. We refer to these as ‘finger modes’ because they display $k_R/k_Z \gg 1$. The second appear at slightly later times in the main body of the disc, including near the mid-plane. These ‘body modes’ have somewhat longer radial wavelengths. Early on they manifest themselves as fundamental breathing modes, but quickly become corrugation modes as symmetry about the mid-plane is broken. The corrugation modes are a prominent feature of the non-linear saturated state, leading to strong vertical oscillation of the disc mid-plane. In a viscous disc with aspect ratio $H/r = 0.05$, instability is found to operate when the viscosity parameter $\alpha < 4 \times 10^{-4}$. In three dimensions the instability generates a quasi-turbulent flow, and the associated Reynolds stress produces a fluctuating effective viscosity coefficient whose mean value reaches $\alpha \sim 10^{-3}$ by the end of the simulation. The evolution and saturation of the vertical shear instability in astrophysical disc models which include realistic treatments of the thermal physics has yet to be examined. Should it occur on either global or local scales, however, our results suggest that it will have significant consequences for their internal dynamics, transport properties and observational appearance.

Key words: accretion, accretion discs – hydrodynamics – instabilities – methods: analytical – methods: numerical – protoplanetary discs.

1 INTRODUCTION

Accretion discs play important roles in a broad range of astrophysical phenomena. Protostellar discs orbiting young stars provide conduits through which most of the mass accretes during the star formation process, and they are the nascent environments for planetary

★ E-mail: r.p.nelson@qmul.ac.uk (RPN); o.gressel@qmul.ac.uk (OG);
oumurhan@ucmerced.edu (OMU)

system formation. Mass transfer through Roche lobe overflow in close binary systems leads to highly energetic and time-variable phenomena because of accretion through a disc on to compact objects such as white dwarfs in cataclysmic variables and neutron stars or black holes in low-mass X-ray binaries. Quasars and active galactic nuclei are powered by disc accretion on to supermassive black holes. Understanding the dynamics and evolution of astrophysical discs is key to understanding these and related phenomena.

Since the early work of Shakura & Sunyaev (1973) and Lynden-Bell & Pringle (1974), significant efforts have been made to understand the internal dynamics of discs, including mechanism(s) that allow them to transport angular momentum and accrete at their observed rates. Several ideas based on dynamical instabilities in non-magnetized flows have been explored, including the Papaloizou–Pringle instability for thick accretion tori driven by unstable non-axisymmetric wave modes (Papaloizou & Pringle 1984), convective instability (Cameron & Pine 1973; Lin & Papaloizou 1980; Ruden, Papaloizou & Lin 1988; Ryu & Goodman 1992), gravitational instability (Toomre 1964; Lin & Pringle 1987; Papaloizou & Savonije 1991), the global and subcritical baroclinic instabilities (SBI; Klahr & Bodenheimer 2003; Johnson & Gammie 2006; Petersen, Stewart & Julien 2007; Lesur & Papaloizou 2010) and the vertical shear instability (Urpin & Brandenburg 1998; Urpin 2003; Arlt & Urpin 2004) which is closely related to the Goldreich–Schubert–Fricke (GSF) instability (Goldreich & Schubert 1967; Fricke 1968) developed in the context of differentially rotating stars. The presence of a weak magnetic field in the disc, however, leads to the development of magnetohydrodynamic (MHD) turbulence driven by the magnetorotational instability (MRI; Balbus & Hawley 1991; Hawley & Balbus 1991), and it is now generally accepted that this is likely to be the source of anomalous viscosity in most accretion discs in which the magnetic field is well coupled to the gas. In addition to transporting angular momentum, disc instabilities may have other important dynamical consequences. In the context of protoplanetary discs, for example, the formation of vortices is believed to have potential importance for planet formation (e.g. Barge & Sommeria 1995; Klahr & Bodenheimer 2003). Instabilities may also lead to general disturbance of the disc, regulating the growth and settling of dust grains (Dullemond & Dominik 2005) and the formation and growth of planetesimals (Nelson & Gressel 2010). Considering a broader context, the growth and maintenance of unstable normal modes has been invoked to explain time-dependent behaviour in X-ray binary discs (e.g. Okazaki, Kato & Fukue 1987). Clearly, having a full understanding of dynamical processes that can be manifest in discs is important for explaining a broad range of phenomena.

In this paper, we present an extensive analysis of the hydrodynamic stability and non-linear dynamics of disc models with power-law mid-plane density distributions, and either temperature or entropy profiles that are power-law functions of R only, where R is the cylindrical radius. Our investigation employs high-resolution 2D axisymmetric and 3D hydrodynamic simulations and a linear stability analysis based on asymptotic methods. Models with power-law temperature profiles, adopting locally isothermal equations of state, have been used extensively in the study of protostellar disc dynamics and disc–planet interactions. These studies normally assume that the disc is viscous (e.g. Kley, D’Angelo & Henning 2001; Cresswell et al. 2007; D’Angelo & Lubow 2010; Fagner & Nelson 2010) or magnetized (e.g. Fromang & Nelson 2006; Beckwith, Armitage & Simon 2011), but in this study we include neither viscosity nor magnetic fields. Adoption of radial variations in temperature or entropy in the models causes them to have angular velocity profiles that are a function of both radius and height, $\Omega(R, Z)$. The height

variation of Ω is often referred to as the thermal wind in studies of atmospheric dynamics, and arises because of the baroclinic nature of the flow.

We find that disc models for which $d\Omega/dZ \neq 0$, and which experience thermal relaxation on \sim dynamical time-scales or shorter, are unstable to the growth of modes with $|k_R/k_Z| \gg 1$, where (k_R, k_Z) are the radial and vertical wavenumbers. This instability appears to be closely related to the GSF instability, as studied by Urpin & Brandenburg (1998), Urpin (2003) and Arlt & Urpin (2004) in the context of accretion discs, an interpretation that is further supported by our own stability analysis. Growth of the instability is favoured when the thermodynamic response of the gas is isothermal, or near-isothermal, although the strength of this dependence varies with model parameters. Steeper thermal gradients display a greater tendency towards instability. The one 3D simulation we present suggests that non-linear development of the instability leads to a turbulent flow whose associated Reynolds stress leads to an effective alpha parameter $\alpha \sim 10^{-3}$, causing non-negligible outward angular momentum transport.

This paper is organized as follows. In Section 2, we present the basic equations of the problem and the disc models we examine. In Section 3, we discuss the hydrodynamic stability of rotating shear flows and previous work in the literature relevant to the present study. In Section 4, we describe the numerical methods employed, and in Section 5 we present the results of the non-linear simulations. A stability analysis is presented in Section 6 that includes a reworking of the Goldreich & Schubert (1967) calculation, which is then extended using asymptotic methods (the mathematical details of which are given in Appendix A). We discuss the application of our study to protoplanetary discs in Section 7 and draw our conclusions.

2 BASIC EQUATIONS

In this paper, we make use of both cylindrical (R, ϕ, Z) and spherical polar coordinates (r, θ, ϕ) . We solve the continuity, momentum and internal energy equations of hydrodynamics

$$\begin{aligned} \partial_t \rho + \nabla \cdot (\rho \mathbf{v}) &= 0, \\ \partial_t (\rho \mathbf{v}) + \nabla \cdot [\rho \mathbf{v} \mathbf{v}] &= -\nabla P - \nabla \Phi, \\ \partial_t (e) + \nabla \cdot (e \mathbf{v}) &= -P \nabla \cdot \mathbf{v} + S - Q, \end{aligned} \quad (1)$$

where ρ is the density, \mathbf{v} is the velocity, P is the pressure, e is the internal energy per unit volume, S and Q are energy source and sink terms, and $\Phi = -GM/r$ is the gravitational potential due to the central star. Here G is the gravitational constant and M is the mass of the star.

2.1 Disc models

We are concerned primarily with two basic equilibrium disc models in this paper. In the first, the temperature, T , and mid-plane density, ρ_{mid} , are simple power-law functions of cylindrical radius:

$$T(R) = T_0 \left(\frac{R}{R_0} \right)^q \quad (2)$$

$$\rho_{\text{mid}}(R) = \rho_0 \left(\frac{R}{R_0} \right)^p, \quad (3)$$

where T_0 is the temperature at the fiducial radius R_0 and ρ_0 is the mid-plane gas density there. Adopting an ideal gas equation of state

$$P = \frac{\mathcal{R}}{\mu} T \rho, \quad (4)$$

where \mathcal{R} is the gas constant and μ is the mean molecular weight; we note that the isothermal sound speed is related to the temperature through the expression $c_s^2 = \mathcal{R}T/\mu$, such that q also represents the radial power-law dependence of $c_s^2(R)$:

$$c_s^2(R) = c_0^2 \left(\frac{R}{R_0} \right)^q. \quad (5)$$

In the second disc model, we adopt a power-law function for the mid-plane density, as in equation (3), and specify the entropy function, K_s , as a strict power-law function of R in the initial model:

$$K_s(R) = K_0 \left(\frac{R}{R_0} \right)^s, \quad (6)$$

where the entropy function is defined through the expression

$$P = K_s \rho^\gamma, \quad (7)$$

and γ is assumed to be constant. The entropy per unit mass is given by

$$S = C_v \log(T \rho^{1-\gamma}), \quad (8)$$

where C_v is the specific heat at constant volume, and the entropy function is given in terms of the entropy by the expression

$$K_s = C_v(\gamma - 1) \exp\left(\frac{S}{C_v}\right). \quad (9)$$

2.1.1 Equilibrium solutions

In order to construct initial conditions for our simulations, we need to obtain equilibrium disc models. The equations of force balance in the radial and vertical directions are given by

$$R\Omega^2 - \frac{GM R}{(R^2 + Z^2)^{3/2}} - \frac{1}{\rho} \frac{\partial P}{\partial R} = 0 \quad (10)$$

$$- \frac{GM Z}{(R^2 + Z^2)^{3/2}} - \frac{1}{\rho} \frac{\partial P}{\partial Z} = 0. \quad (11)$$

Combining equations (10), (11), (2) and (3) leads to expressions for the equilibrium density and angular velocity, Ω , as functions of (R, Z) for the disc with a power-law temperature profile:

$$\rho(R, Z) = \rho_0 \left(\frac{R}{R_0} \right)^p \exp\left(\frac{GM}{c_s^2} \left[\frac{1}{\sqrt{R^2 + Z^2}} - \frac{1}{R} \right]\right), \quad (12)$$

$$\Omega(R, Z) = \Omega_K \left[(p + q) \left(\frac{H}{R} \right)^2 + (1 + q) - \frac{qR}{\sqrt{R^2 + Z^2}} \right]^{1/2}, \quad (13)$$

where $\Omega_K = \sqrt{GM/R^3}$ is the Keplerian angular velocity at radius R and $H = c_s/\Omega_K$ is the local disc scaleheight (also see Takeuchi & Lin 2002; Fromang, Lyra & Masset 2011). The definition of c_s given in equation (5) implies that

$$H = H_0 \left(\frac{R}{R_0} \right)^{(q+3)/2}, \quad (14)$$

where $H_0 = c_0/\sqrt{GM/R_0^3}$ is the disc scaleheight at radius R_0 .

Similarly, the equilibrium density and angular velocity for the disc model with a power-law entropy function profile are

given by

$$\rho(R, Z) = \left(\rho_{\text{mid}}^{(\gamma-1)} + \frac{(\gamma-1)GM}{\gamma K_s} \left[\frac{1}{\sqrt{R^2 + Z^2}} - \frac{1}{R} \right] \right)^{1/(\gamma-1)} \quad (15)$$

$$\Omega(R, Z) = \Omega_K \left[\frac{p}{\mathcal{M}_{\text{mid}}^2} + \frac{s}{\gamma \mathcal{M}^2} + (1 + s) - \frac{sR}{\sqrt{R^2 + Z^2}} \right]^{1/2}, \quad (16)$$

where $\mathcal{M}_{\text{mid}} = v_K/a_{\text{mid}}$ is the Mach number at the disc mid-plane and $\mathcal{M} = v_K/a_s$ is the Mach number at each disc location. $v_K = \sqrt{GM/R}$ is the Keplerian velocity at radius R , and the adiabatic sound speed $a_s = \sqrt{\gamma P/\rho}$, which takes the value a_{mid} at the disc mid-plane. We note that combining equations (4), (6), (7) and (15) demonstrates that the temperature in this model is a function of both R and Z in general:

$$T(R, Z) = K_s \rho^{\gamma-1} \frac{\mu}{\mathcal{R}}. \quad (17)$$

As such, this model provides a useful contrast to the one with temperature constant on cylinders, and is convenient to implement numerically because of the existence of analytic solutions for the equilibrium $\rho(R, Z)$ and $\Omega(R, Z)$ profiles.

Equations (2), (3), (12) and (13) fully specify the initial disc models with power-law temperature profiles that we examine in this paper, subject to appropriate choices for p and q . Expressions (3) and (6), along with (15) and (16), specify the initial models for which the entropy function is a power-law function of the cylindrical radius, subject again to appropriate choices for the power-law exponents p and s . We note that for all models in which the initial temperature, T , or entropy function, K_s , is strict power-law functions of R , the equilibrium angular velocities are explicit functions of both R and Z , a fact that appears to play a key role in the disc instability that we examine in this paper.

2.2 Thermodynamic evolution

The thermodynamic evolution of both disc models described above in Section 2.1 is assumed to be one of three types in the simulations presented here: *locally isothermal*, for which the local temperature at each (R, Z) position in the disc is kept strictly equal to its original value; *adiabatic*, where the entropy of the fluid is kept constant [equivalent to there being no source/sink term in the energy equation (1)]; *thermally relaxing*, where we relax the temperature at each location in the disc towards its initial value on some time-scale, τ_{relax} . The thermal relaxation model we adopt is

$$\frac{dT}{dt} = -\frac{(T - T_0)}{\tau_{\text{relax}}}, \quad (18)$$

where T_0 is the initial temperature. For simplicity, we assume that τ_{relax} is a function of R , being a fixed multiple or fraction of the local Keplerian orbital period. Equation (18) has a simple analytic solution of the form

$$T(t + \Delta t) = T_0 + (T(t) - T_0) \exp\left(-\frac{\Delta t}{\tau_{\text{relax}}}\right), \quad (19)$$

where $T(t)$ is the temperature at time t and $T(t + \Delta t)$ is the temperature at some later time $t + \Delta t$.

In the locally isothermal models, we use an isothermal equation of state $P = c_s^2 \rho$ and do not evolve the energy equation in equation (1). In the adiabatic models, we use the equation of state

$P = (\gamma - 1)e$, solve the energy equation in equation (1), and neglect the source and sink terms. The energy equation is also solved in the thermally relaxing models, along with equation (18) which plays the role of the source and sink terms in the energy equation (1).

3 HYDRODYNAMIC STABILITY OF DISC MODELS

3.1 The Rayleigh and Solberg–Høiland criteria

The Rayleigh criterion indicates that accretion discs with strictly Keplerian angular velocity profiles are hydrodynamically stable since

$$\frac{dj^2}{dr} > 0,$$

where $j = R^2\Omega$ and $\Omega = \sqrt{GM/R^3}$. More generally, a differentially rotating, compressible fluid with angular velocity varying with height and radius, $\Omega(R, Z)$, subject to axisymmetric adiabatic perturbations (i.e. $DS/Dt = 0$, where D/Dt is the total time derivative for fluid elements), is stable if both of the Solberg–Høiland criteria are satisfied (e.g. Tassoul 1978):

$$\frac{1}{R^3} \frac{\partial j^2}{\partial R} + \frac{1}{\rho C_p} (-\nabla P) \cdot \nabla S > 0 \quad (20)$$

$$-\frac{\partial P}{\partial Z} \left(\frac{\partial j^2}{\partial R} \frac{\partial S}{\partial Z} - \frac{\partial j^2}{\partial Z} \frac{\partial S}{\partial R} \right) > 0. \quad (21)$$

Accretion disc models generally possess negative radial pressure gradients, and a negative vertical pressure gradient in the disc hemisphere above the mid-plane, leading to stability criteria (20) and (21) of the form

$$\frac{1}{R^3} \frac{\partial j^2}{\partial R} + \frac{1}{\rho C_p} \left(\left| \frac{\partial P}{\partial R} \right| \frac{\partial S}{\partial R} + \left| \frac{\partial P}{\partial Z} \right| \frac{\partial S}{\partial Z} \right) > 0 \quad (22)$$

$$\frac{\partial j^2}{\partial R} \frac{\partial S}{\partial Z} - \frac{\partial j^2}{\partial Z} \frac{\partial S}{\partial R} > 0. \quad (23)$$

Considering a disc with a strictly Keplerian j profile, we see that stability according to equation (23) requires $\partial S/\partial Z > 0$, in agreement with the Schwarzschild condition for convective stability. Condition (22) shows that a large-amplitude negative radial entropy gradient $\partial S/\partial R < 0$ can also drive instability in principle, provided the gradient is strong enough to overcome the positive angular momentum gradient. This is distinct from the global baroclinic and/or SBI discussed in the introduction that does not require violation of the Solberg–Høiland criteria to operate, but does require thermal evolution of the fluid on time-scales similar to the dynamical time to re-establish the initial radial entropy gradient, in addition to the presence of finite-amplitude perturbations (Lesur & Papaloizou 2010).

Considering a thin ($H/R \ll 1$) quasi-Keplerian disc whose angular velocity depends on height and radius, inspection of equation (13) shows that $\partial j/\partial Z \simeq q(H/R)\partial j/\partial R$. Inserting this expression into equation (23) gives

$$\frac{\partial S}{\partial Z} - q \left(\frac{H}{R} \right) \frac{\partial S}{\partial R} > 0, \quad (24)$$

so that a thin disc with $\partial S/\partial Z > 0$ and $q < 0$ can be destabilized only if $\partial S/\partial R < 0$ and satisfies

$$q \left(\frac{H}{R} \right) \frac{\partial S}{\partial R} > \frac{\partial S}{\partial Z}. \quad (25)$$

We consider two basic disc models in this paper that are the subject of non-linear simulations in which the fluid evolution is adiabatic and for which the Solberg–Høiland criteria determine hydrodynamic stability. One model assumes that the temperature profile is a strict power-law function of radius, R , such that the density and angular velocity are given by equations (12) and (13). With values $p = -1.5$ and $q = -1, -0.5, -0.25$ and 0 in equations (2) and (3), respectively, these discs are stable according to criterion (22). The term involving $\partial S/\partial R$ provides the only destabilizing contribution (because $\partial S/\partial Z > 0$), but is always far too small to overcome the positive radial gradient in j^2 and the strongly stabilizing positive vertical entropy gradient. This disc is also stable according to criterion (25) because the negative radial entropy gradient is insufficient to overcome the positive vertical entropy gradient.

Our second set of disc models assume that entropy is a strict power-law function of radius R . The density and angular velocity are given by expressions (15) and (16), and the values ($p = 0, s = -1$) and ($p = -1.5, s = 0$) are adopted in equations (3) and (6). Both of these models are stable according to criterion (22) as the destabilizing term proportional to $\partial S/\partial R$ is either zero or too small ($\partial S/\partial Z = 0$ in these models). Criterion (25), however, predicts instability for the ($p = 0, s = -1$) model because the term on the left-hand side is positive and the right-hand side is zero. The adiabatic simulation that we present later in this paper will provide an example of non-linear evolution of a model that is unstable according to one of the Solberg–Høiland criteria. Finally, the model with ($p = -1.5, s = 0$) is stable according to criterion (25) as both the right- and left-hand sides are zero.

3.2 The GSF instability

When thermal and viscous diffusion play a role, the stability properties of rotating flows are no longer controlled by the Solberg–Høiland criteria, but instead are determined by stability criteria obtained originally by Goldreich & Schubert (1967) and Fricke (1968) in application to the radiative zones of differentially rotating stars. Axisymmetric rotating flows are susceptible to the GSF instability when viscous diffusion is much less efficient than thermal diffusion, such that a fluid element retains its initial angular momentum but quickly attains the entropy of the surrounding fluid when perturbed from its equilibrium location. Under these circumstances, the stabilizing influence of entropy gradients provided by the Solberg–Høiland criteria diminishes and instability ensues for wave modes satisfying the instability criterion

$$\frac{\partial j^2}{\partial R} - \frac{k_R}{k_Z} \frac{\partial j^2}{\partial Z} < 0. \quad (26)$$

For a rotating flow where the angular momentum is a function of both R and Z , and the appropriate conditions on thermal and viscous diffusion are satisfied, unstable modes are guaranteed to exist since wavevectors with ratios k_R/k_Z that satisfy equation (26) can always be found. Adopting the scalings discussed in Section 3.1 for a thin ($H/R \ll 1$) quasi-Keplerian disc, according to equation (26) instability occurs for modes satisfying $k_R/k_Z > R/H$ (i.e. unstable modes will have radial wavelengths that are much shorter than vertical ones).

Application of the GSF instability to accretion discs has not received a great deal of attention in the literature (but see the discussion below). Indeed, the study presented in this paper has arisen from unrelated attempts to generate 3D models of protoplanetary discs within which turbulence is generated by the MRI in active layers near the disc surface, with extensive dead zones that are

magnetically inactive (or at least stable against the MRI) near the mid-plane. The adoption of a locally isothermal equation of state in these models, combined with the absence of a physical viscosity, rendered them unstable to the growth of vertical corrugation oscillations that in the non-linear regime became quite violent. This behaviour appears to be a non-linear manifestation of the GSF instability, which we study here in more detail. The two classes of models we consider (temperature constant on cylinders and entropy constant on cylinders) both have angular velocity profiles that depend on both R and Z , and we expect that models in which perturbations evolve quasi-isothermally will display the GSF instability.

Previous analysis of the GSF instability in the context of accretion discs was initiated by Urpin & Brandenburg (1998) in which they presented a local linear analysis utilizing the short-wavelength approximation. A more extensive analysis was presented in Urpin (2003), where it was suggested that the GSF may provide a source of hydrodynamic turbulence in accretion discs, and non-linear simulations were presented by Arlt & Urpin (2004). These non-linear simulations adopted a basic disc model with a strictly isothermal equation of state. As such, the underlying equilibrium disc model has an angular velocity profile that is independent of height, Z (see equation 13), and should not be susceptible to the GSF instability. Arlt & Urpin (2004), however, adopted initial conditions that allowed relaxation around the equilibrium state leading to variations of Ω with height, and when initial perturbations with $|k_R/k_Z| \gg 1$ were applied, growth of the GSF was observed. It is also worth noting the related linear and non-linear study of adiabatic accretion disc models presented by Rüdiger, Arlt & Shalybkov (2002) in this context. This examined the stability of discs whose angular momentum and entropy profiles rendered them stable according to the previously discussed Solberg–Høiland criteria. As expected, applied perturbations were found to always decay. We present a linear analysis of the instability we study in Section 6 for a fully compressible fluid under the assumption that the equation of state is locally isothermal.

4 NUMERICAL METHODS

The simulations presented in this paper were performed using two different codes that utilize very different numerical schemes. We use an older version of *NIRVANA*, which utilizes an algorithm very similar to the *ZEUS* code to solve the equations of ideal MHD (Stone & Norman 1992; Ziegler & Yorke 1997). This scheme uses operator splitting, dividing the governing equations into source and transport terms. Advection is performed using the second-order monotonic transport scheme (van Leer 1977). We also use the more modern *NIRVANA-III* code, which is a second-order Godunov-type MHD code (Ziegler 2004), which has recently been extended to orthogonal-curvilinear coordinate systems (Ziegler 2011). All presented simulations were performed using a standard spherical coordinate system (r, θ, ϕ) .

4.1 Initial and boundary conditions

The numerical study presented here applies to the two general classes of disc model discussed in Section 2.1, and as such we adopt a numerical set-up that is not specific to any particular physical system (although, as already mentioned, our motivation for undertaking this study arose from earlier attempts to establish stationary equilibrium solutions for protostellar disc models). Based on numerous test calculations performed with a wide range of resolutions during an early stage of this project, we know that the insta-

bility we study here is characterized by having a radial wavelength much shorter than the vertical wavelength (i.e. $k_R \gg k_Z$) during its early growth phase. Consequently, we consider disc models with narrow radial domains to facilitate high-resolution simulations. The spherical polar grid we adopt for our standard numerical set-up has inner radius $r_{\text{in}} = R_0 = 1$ and outer radius $r_{\text{out}} = 2$, but for a small number of simulations that examine detailed behaviour as a function of H/R we use smaller radial domains that are equal to five disc scaleheights, running between $r_{\text{in}} = 1$ and $r_{\text{out}} = 5 \times H_0$. Most simulations we present are axisymmetric. Our one non-axisymmetric run covers a restricted azimuthal domain of $\pi/4$, again to facilitate a high-resolution study.

Disc models in which the initial temperature is a strict function of the cylindrical radius, R , have meridional domains $\pi/2 - 5H_0/R_0 \leq \theta \leq \pi/2 + 5H_0/R_0$. For a disc with radial temperature profile $q = -1$, corresponding to a disc with constant H/R , the meridional domain covers ± 5 scaleheights above and below the mid-plane. For larger (less negative) values of q , the disc covers ± 5 scaleheights at the inner radius, but a reduced number of scaleheights as one moves out in radius. Prior to initiating these simulations, the disc models are specified using equations (3), (2), (12) and (13). In most models the initial velocity field was seeded with random noise distributed uniformly in each component with a peak amplitude equal to 1 per cent of the local sound speed. For the set of runs that examines evolution as a function of H/R , we employ seed noise with amplitude equal to $10^{-6}c_s$ as we wish to examine the earliest behaviour during the linear phase of disc instability.

Discs for which the initial entropy function, K_s , depends only on R have physical surfaces where the density goes to zero. In these models, the meridional boundaries are placed at a location that is 5 per cent smaller than the angular distance from the mid-plane where the density vanishes. The initial models are specified using equations (3), (6), (15) and (16). In order to determine the value of K_0 in equation (6), we specify the mid-plane Mach number, \mathcal{M}_{mid} , at radius R_0 . We adopt $\mathcal{M}_{\text{mid}} = 20$ to be consistent with the models for which $H/R = 0.05$. Seed noise with amplitude equal to 1 per cent of the sound speed was added to all velocity components.

For most simulations, we adopt standard outflow or reflecting boundary conditions at the inner and outer radial boundaries. Periodic boundaries are applied in the azimuthal direction, and either standard outflow or reflecting conditions are applied at the meridional boundaries (all simulations performed using *NIRVANA-III* adopted outflow boundary conditions at the radial and meridional boundaries). The density is obtained in ghost zones by means of linear extrapolation. A variety of different boundary conditions were used in test simulations at an early stage of this project, and the results were found to be insensitive to the choice adopted. These tests included adoption of damping boundary conditions that absorb incoming waves, indicating that reflecting boundaries are not required to drive the instability discussed in this paper.

Owing to the unsplit character of finite-volume schemes, such as used in *NIRVANA-III*, it is difficult numerically to preserve static equilibria (Zingale et al. 2002). In particular, problems arise with the constant extrapolation of the density profile in the vertical direction. This is because the weight of the gas is not balanced by the now vanishing pressure gradient in the case of an isothermal equation of state, leading to a standing accretion shock in the first grid cell of the meridional domain. In the case of an adiabatic evolution equation, the same problem arises in the boundary condition for the thermal energy (which enters the boundary condition for the total energy). To alleviate this situation, we obtain the density (and in the case of solving an energy equation, the pressure) by integrating

the hydrostatic equilibrium in each cell adjacent to the domain boundary. This is done with a second-order Runge–Kutta shooting method.

Unless indicated otherwise, we use a fixed number of grid cells $N_r \times N_\theta$ of 1328×1000 for all our NIRVANA runs. The non-axisymmetric run had $N_\phi = 300$ cells in the azimuthal direction (covering an azimuthal domain $0 \leq \phi \leq \pi/4$). The NIRVANA-III runs used a resolution of 1344×1024 .

We adopt a system of units in which $M = 1$, $G = 1$ and $R_0 = 1$. When presenting our results, the unit of time is the orbital period at the disc inner edge, $P_{\text{in}} = 2\pi$.

5 RESULTS

The main aims of the following simulations are to delineate conditions under which the disc models outlined in Section 2.1 are unstable to the growth of the vertical shear instability, and to examine the effect that different physical and numerical set-ups have on its growth and evolution. We also aim to characterize the final saturated state of these unstable discs, although our adoption of mainly axisymmetric simulations provides some restriction in achieving this final aim.

We define the volume-integrated meridional and radial kinetic energies through the expressions

$$e_\theta = \frac{1}{2} \int_V \rho v_\theta^2 dV, \quad e_r = \frac{1}{2} \int_V \rho v_r^2 dV. \quad (27)$$

When presenting our results, we normalize these energies by the total kinetic energy contained in Keplerian motion in the disc initially.

We begin discussion of our results below by describing simulations of locally isothermal discs for which $T(R) \sim R^{-q}$ and $\rho(R) \sim R^{-p}$. We describe one fiducial model in detail, before discussing briefly the influence of the temperature profile in controlling the instability. We present a comparison between results obtained using the two codes described in Section 4, and also demonstrate how the instability evolves as a function of disc viscosity and H/R .

The next set of results we present are for disc models in which the initial temperature is a strict function of R , but we set $\gamma = 1.4$, solve the energy in equation (1) and allow the temperature to relax towards its initial value using equation (18). In this section, we examine how the thermal relaxation rate controls the instability, covering the full range of thermodynamic behaviour from locally isothermal through to adiabatic.

In the penultimate part of our numerical study, we consider disc models for which the entropy function, K_s , is a strict function of R , and again employ thermal relaxation to examine the conditions under which accretion discs display the vertical shear instability. The final numerical experiment we present examines the instability in a full 3D model and provides an estimate of the Reynolds stress induced by the instability.

In the discussion section, we present a modified stability analysis that follows that presented by Goldreich & Schubert (1967), except that we consider a fully compressible gas and adopted a locally isothermal equation of state. We then extend this analysis by relaxing the purely local assumption and, using simplifications motivated by the simulation results, examine the stability of wave modes that are predicted to exist by the equations thus obtained. These analyses are used to interpret the results of the simulations.

Table 1. Simulation parameters and results. Labels beginning with a ‘T’ denote runs with T being a function of R . Those with a ‘K’ denote runs where $K_s = K_s(R)$. The letter R denotes reflecting boundary conditions, and a letter O denotes outflow boundary conditions. Digits after the hyphens denote thermal relaxation times. The final column indicates whether the disc was stable (S) or unstable (U).

Run	p	q or s	H_0/R_0	τ_{Relax}	N_r, N_θ, N_ϕ	U/S
T1R/O-0	−1.5	−1	0.05	0.00	1328, 1000, 1	U
T2R/O-0	−1.5	−0.5	0.05	0.00	1328, 1000, 1	U
T3R/O-0	−1.5	−0.25	0.05	0.00	1328, 1000, 1	U
T4R/O-0	−1.5	0	0.05	0.00	1328, 1000, 1	U
T5R-0	−1.5	−1	0.01	0.00	1328, 1000, 1	U
T6R-0	−1.5	−1	0.02	0.00	1328, 1000, 1	U
T7R-0	−1.5	−1	0.03	0.00	1328, 1000, 1	U
T8R-0	−1.5	−1	0.05	0.00	1328, 1000, 1	U
T9R-0	−1.5	−1	0.1	0.00	1328, 1000, 1	U
T10R-0.01	−1.5	−1	0.05	0.01	1328, 1000, 1	U
T11R-0.1	−1.5	−1	0.05	0.10	1328, 1000, 1	S
T12R-1.0	−1.5	−1	0.05	1.00	1328, 1000, 1	S
T13R-10.0	−1.5	−1	0.05	10.0	1328, 1000, 1	S
T14R-∞	−1.5	−1	0.05	∞	1328, 1000, 1	S
K1R-0	0	−1	0.05	0.00	1328, 1000, 1	U
K2R-0.01	0	−1	0.05	0.01	1328, 1000, 1	U
K3R-0.1	0	−1	0.05	0.10	1328, 1000, 1	U
K4R-1.0	0	−1	0.05	1.00	1328, 1000, 1	U
K5R-10.0	0	−1	0.05	10.0	1328, 1000, 1	U
K6R-∞	0	−1	0.05	∞	1328, 1000, 1	S
K7R-0.01	−1.5	0	0.05	0.01	1328, 1000, 1	S
T1R-0-3D	−1.5	−1	0.05	0.00	1328, 1000, 300	U

5.1 A fiducial model

We begin presentation of the simulation results by discussing one particular model in detail to illustrate the nature of the instability that is the focus of this paper. The fiducial model is T1R-0 listed in Table 1, with temperature, T , defined as a function of R only, and $p = -1.5$ and $q = -1$ in equations (3) and (2). A locally isothermal equation of state is adopted. As such, this disc model has parameters very similar to those used in numerous previous studies of disc-related phenomena (e.g. Kley et al. 2001; Cresswell & Nelson 2006; Fromang et al. 2011), although we focus primarily on the inviscid non-magnetized evolution here.

The time evolution of the normalized meridional and radial kinetic energies defined in equation (27) is shown in Fig. 1 using both

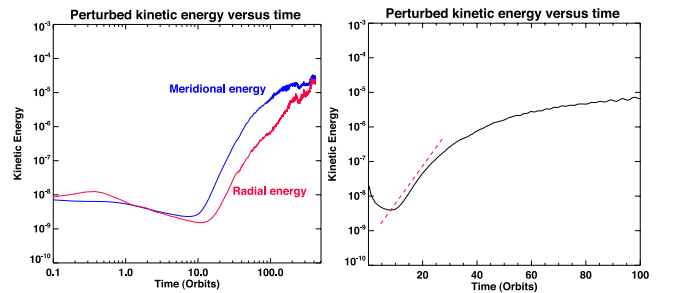


Figure 1. Left-hand panel: time evolution of the normalized perturbed kinetic energy in the meridional and radial coordinate directions for model T1R-0 with $p = -1.5$, $q = -1$ and reflecting boundary conditions at the meridional boundaries. Right-hand panel: total perturbed kinetic energy versus time plotted using log–linear axes. The red dashed line shows exponential growth with a growth rate equal to 0.25 orbit^{-1} .

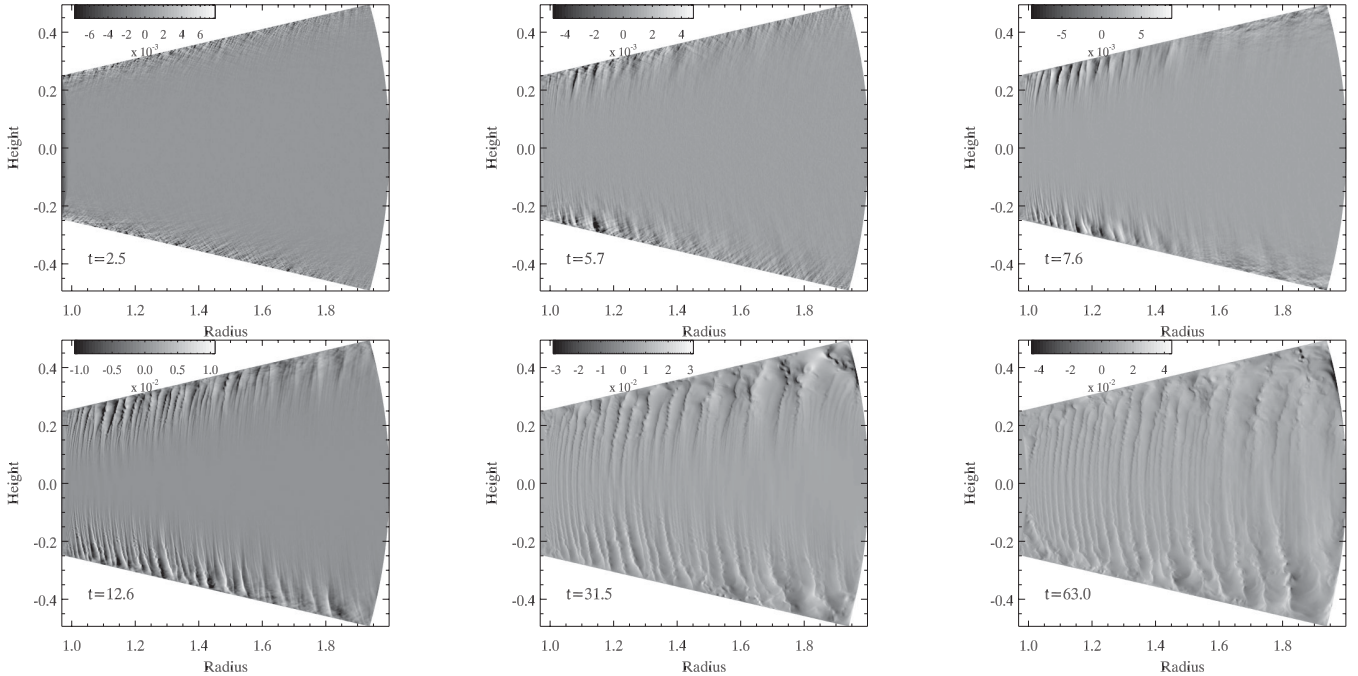


Figure 2. Edge-on contours of the perturbed vertical velocity as a function of R , Z and time for model T1R-0.

a log–log and a log–linear plot. The initial values at $t = 0$ originate from the seed noise, and we observe that after ~ 10 orbits, during which the perturbed kinetic energies damp slightly, rapid growth of the perturbation energies arises. The normalized energies reach non-linear saturation after ~ 400 orbits having reached values of a few $\times 10^{-5}$. We note here that the restricted size of the computational domain, combined with the adoption of reflecting boundary conditions, prevents us from running simulations for much longer than presented in Fig. 1, as the results become unreliable with the perturbation energies increasing unphysically (see Fig. 7, for example). Lower resolution simulations with larger simulation domains, which are not presented here, have been run for extended time periods of more than 1000 orbits, and the perturbation energies in those runs were observed to maintain an approximately constant saturated level. Inspection of the evolution of the sum of the meridional and radial kinetic in the log–linear plot indicates that the linear growth rate of the perturbed energy is $\simeq 0.25 \text{ orbit}^{-1}$ during the early linear growth phase.

Contour plots of vertical velocity perturbations, v_z , that arise at different stages of the disc evolution are shown in Figs 2 and 3. These two figures show the perturbed velocity field at identical times, but whereas Fig. 2 maps linearly between the velocity values and the grey-scale, Fig. 3 plots the values $\text{sign}(v_z) \times |v_z|^{1/4}$ so that the grey-scale is stretched to enable the morphology of the perturbations to be more clearly discerned. Both figures demonstrate that perturbations start to grow near the upper and lower disc surfaces, where $|d\Omega/dZ|$ is largest, and towards the inner edge of the disc. The perturbations are characterized by having short radial and long vertical wavelengths, as expected for the vertical shear instability described in Sections 3.2 and 6. The short radial wavelength gives rise to significant radial shear in the vertical velocity dv_z/dR , and this apparently causes small-scale eddies to form at the shearing interfaces. As time proceeds, the instability extends towards the disc mid-plane and out to larger radii, until the entire disc participates in the instability (although it should be noted that the mid-plane where $d\Omega/dZ = 0$ is formally stable to local growth of the vertical

shear instability). We note that during early stages of evolution, the emerging velocity perturbations located at high latitudes have wavelengths of the order of $\lambda_R \sim 0.01$, a length-scale that is resolved by ≈ 13 grid cells.

Close inspection of the lower panels in Fig. 3 shows that the velocity perturbations have odd symmetry about the mid-plane initially (this is particularly apparent in the fourth panel between radii $1.1 \leq r \leq 1.4$). In other words, the disc exhibits a breathing motion about the mid-plane as the instability first becomes apparent at lower disc latitudes. As time proceeds, however, the velocity perturbations become symmetric about the mid-plane as demonstrated by the fifth and sixth panels. These perturbations correspond to corrugation of the disc characterized by coherent oscillations of the vertical centre of mass position whose phase depends on radius in a time-dependent manner. The development of this disc corrugation is illustrated by Fig. 4, which shows the vertical centre of mass position of the disc at each radius for five different times (note that each plot is off-set in the vertical direction to aid clarity, and each curve has been multiplied by a unique factor so that the corrugation may be observed). The vertical centre of mass has been normalized by the local disc scale-height at each radius. Moving from the lower to the upper curve, we note that the vertical centre of mass position has a very small variation with radius after one time step, but this becomes progressively larger in amplitude and more spatially coherent as time progresses (times corresponding to each curve are given in the figure caption). The final curve, corresponding to an evolution time of ~ 92 orbits, has a maximum vertical displacement approximately equal to $0.006H$. It is interesting to note that the initial disc instability begins with $|k_R/k_Z| \gg 1$, but as the disc approaches the non-linear state, the development of coherent corrugation waves causes the radial wavelengths of the most apparent perturbations to approach or modestly exceed the local scaleheight. At the end of the simulation (~ 420 orbits), the maximum vertical displacement of the disc centre of mass reaches $\sim 0.01H$, but we are cautious about interpreting the results at this late stage of evolution as the reflecting boundary conditions may play a role.

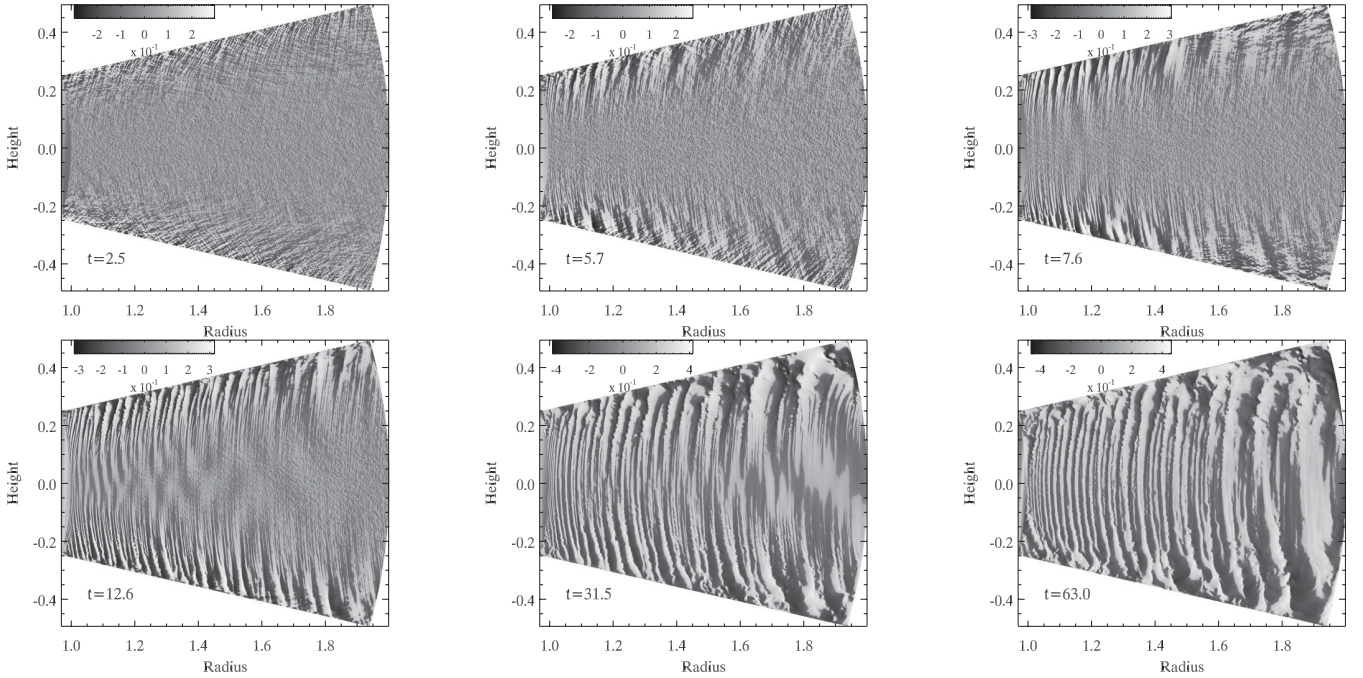


Figure 3. Edge-on contours of the perturbed vertical velocity as a function of R , Z and time for model T1R-0. Note that for clarity, the grey-scale of the image has been stretched by plotting the quantity $\text{sign}(v_z) \times |v_z|^{1/4}$. Note that the spectrum bar shows values of $v_z^{1/4}$.

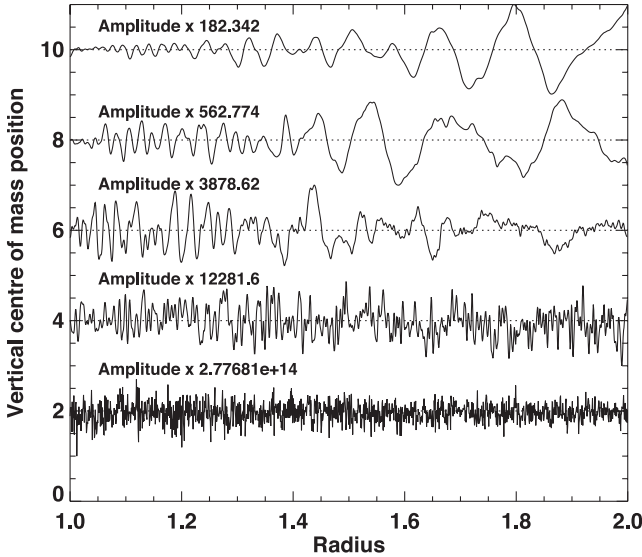


Figure 4. Time evolution of the vertical centre of mass (c.o.m.) position for each radial location in the disc for the fiducial model T1R-0 with $H/R = 0.05$. Note that the vertical c.o.m. position has been normalized by the local scale-height at each radius. Starting from bottom to top, the plots correspond to times (in orbits): 9×10^{-4} , 9.18, 27.86, 65.00, 92.66. The multiplicative factor indicated in each legend causes the maximum amplitude of the normalized c.o.m. position in each graph to equal unity.

Contour plots of the density perturbations $\delta\rho/\rho_0$ corresponding to the previously discussed velocity contours are displayed in Fig. 5. As with the velocity contours, we see perturbations first arise near the disc surface. Coherent density structures first become visible in the third panel, but are substantially more apparent in the lower panels.

The above discussion about the early disc evolution is influenced by the fact that the initial perturbations applied to the velocity

field were actually quite large (amplitude $0.01 \times c_s$). Anticipating the discussion later in the paper (Section 5.5) regarding models that examine behaviour as a function of H/R (for which much smaller amplitude seed noise was applied), we note that a more accurate description of the early evolution is one in which two distinct mode types appear and grow. The first are those that appear at the highest latitudes in the disc and have the appearance of thin, radially confined vertical velocity perturbations with $k_R \gg k_Z$. We refer to these as ‘finger modes’, and they are the ones that are first seen to emerge in the simulations, and as time proceeds they extend down towards the mid-plane. The second type of modes have longer radial wavelengths and appear in the main body of the disc near the mid-plane long before the finger modes have extended down to that region. We refer to these modes as ‘body modes’. Although just about discernible in Fig. 3, the sequence of events in which the two types of modes grow independently before interacting is most clearly exemplified in the contours of vertical velocity shown in Fig. 11, where the left-hand and middle panels show the early development and coexistence of the finger and body modes, and the right-hand panel shows that the rather turbulent appearance of the velocity field in the upper disc layers is caused by their non-linear interaction. It is obvious from the anti-symmetry about the mid-plane that the body modes are associated initially with the breathing modes mentioned above, and evolve to become the corrugation modes that dominate the non-linear state.

The final saturated state consists of locally unstable disc annuli that oscillate vertically at close to the local frequency, $\Omega_k(R)$, superposed on which are a spectrum of oscillations with different frequencies caused by travelling waves excited by vertical oscillations at neighbouring disc radii. A region of the disc lying at intermediate radii will thus experience a locally generated corrugation, in addition to inward travelling corrugation waves that propagate as inertial modes (or r modes) launched from exterior disc locations, and outward propagating corrugation waves propagating as acoustic or

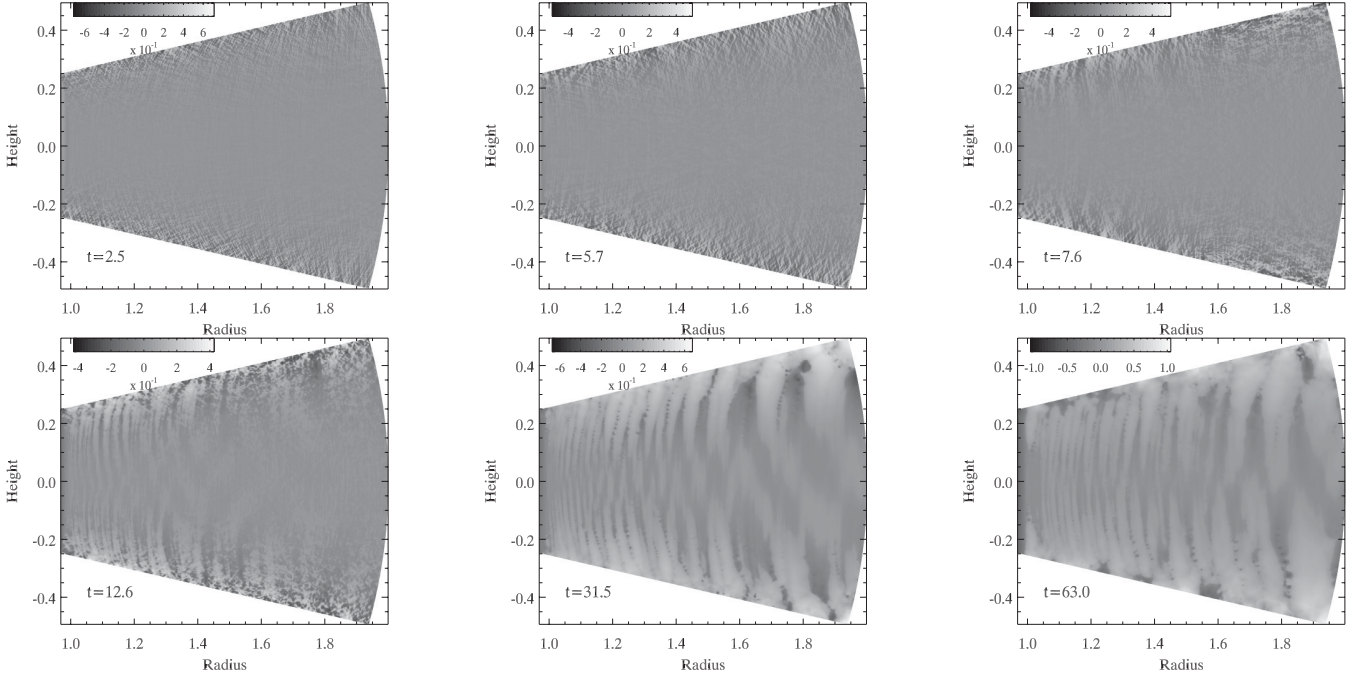


Figure 5. Edge-on contours of the disc relative density perturbations $\delta\rho/\rho_0$ as a function of R, Z and time. Note that we have effectively stretched the grey-scale by plotting the quantity $\text{sign}(\delta\rho) \times \sqrt{\delta\rho/\rho_0}$.

fundamental modes launched from interior disc locations (Lubow & Pringle 1993; Lubow & Ogilvie 1998).

Although we only present simulations with finite-amplitude initial perturbations to the velocity fields in this paper, we have conducted numerous experiments in which the peak amplitude of the imposed perturbations varies, including the cases where perturbations just grow from numerical round-off errors. Although this requires a larger time interval for the instability to become apparent, we nonetheless observe instability for all perturbation amplitudes, demonstrating that the instability is linear. This is further supported by the linear stability analysis presented later in the paper.

5.2 Evolution as a function of the radial temperature profile

In this section, we discuss simulations T1R-0 to T4R-0 (which utilize reflecting boundary conditions in the meridional direction) and T1O-0 to T4O-0 (which use open boundary conditions). These simulations have $p = -1.5$ and a range of values for q running from $q = -1$ (a constant H/R disc) up to $q = 0$ (a purely isothermal disc in which $H/R \propto R^{1/2}$).

The left-hand panel of Fig. 6 shows the time evolution of the normalized perturbed kinetic energy summed over the radial and meridional directions in the disc models T1R-0 to T4R-0. We see that as the value of q increases from $q = -1$ through to $q = -0.25$, the growth rate of the instability decreases, and for $q = 0$ it switches off altogether. Inspection of the evolution of the sum of the meridional and radial kinetic energies on a log-linear plot indicates that the linear growth rate for the $q = -0.5$ case is $\approx 0.12 \text{ orbit}^{-1}$, and for the $q = -0.25$ run is $\approx 0.052 \text{ orbit}^{-1}$ (to be contrasted with the growth rate $\approx 0.25 \text{ orbit}^{-1}$ obtained for the $q = -1$ run). The azimuthal velocity in the $q = 0$ model is independent of Z , as indicated by equation (13), so the observed stability of this disc model is in agreement with the expectations discussed in Section 3. It is also noteworthy that the saturated values of the perturbation energies in

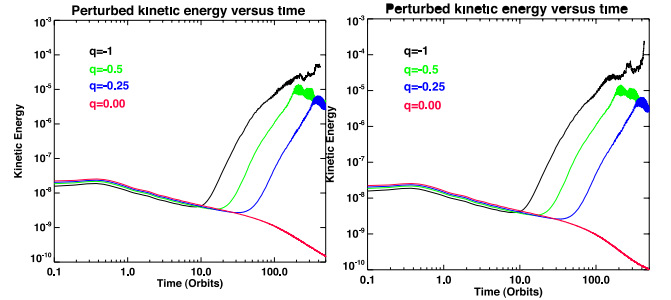


Figure 6. Time evolution of the perturbed meridional plus radial kinetic energies (normalized) as a function of the radial temperature profile. The left/right-hand panel shows results for simulations which adopted reflecting/outflow boundary conditions at the upper and lower disc surfaces.

each unstable disc, normalized to the total energy in Keplerian motion, differ substantially from one another in accord with the trend in the growth rates.

The right-hand panel of Fig. 6 shows the corrugation energies for runs T1O-0 to T4O-0, which are identical to runs T1R-0 to T4R-0 except that the boundary conditions applied at the meridional boundaries are outflow rather than reflecting. During the growth phase of the instability, the results of the T1O-0 to T4O-0 runs are essentially identical to the corresponding T1R-0 to T4R-0 calculations, and the saturated energies are very similar. We also observe the important result that the transition between stable and unstable disc models is independent of the boundary conditions, with both $q = 0$ models showing decay of the perturbed meridional and radial kinetic energies during the simulations. It is clear that the existence of a radial temperature profile plays a fundamental role in determining whether or not a disc becomes unstable. This appears to apply to both the finger and body modes described earlier in Section 5.1.

5.3 Code comparison

Numerical modelling of the Navier–Stokes equations can pose arcane pitfalls, particularly if hydrodynamic instabilities are involved. In the absence of rigorous analytical reference solutions, it has become customary to substantiate the physical reliability of the solutions obtained by means of code comparisons. While we have already shown the generality of a GSF-like instability occurring under various physical settings, we demonstrate here its comparative development using two numerical schemes. Although very similar in their names, the two codes we used are fundamentally different in the numerical schemes they apply. *NIRVANA* utilizes the same finite-difference scheme as the *ZEUS* code, whereas *NIRVANA-III* applies a finite-volume Godunov scheme very similar to the ones used in, e.g., *RAMSES* (Teyssier 2002) or *ATHENA* (Gardiner & Stone 2008).

We have run the simulations labelled T1O-0 to T4O-0 in Table 1 using both codes. The codes generally agree well in the development of the instability and its saturation level as demonstrated by Fig. 7, although it should be noted that the realization of the initial seed noise in the two runs is slightly different. The decay of these initial seed perturbations occurs slightly faster in the *NIRVANA-III* simulations, and the saturation amplitude is slightly smaller. A detailed look at the vertical velocity perturbations at time $t \simeq 12.8$ orbits is shown in Fig. 8, and both codes show the characteristic $|k_R/k_Z| \gg 1$ perturbations associated with the early development of the vertical shear instability. The codes are in decent agreement about the magnitude of the velocity perturbations and also in the dominant wavelength of instability. The codes, however, also show some differences in their solution at this time. The continued presence of the initial seed noise is more apparent in the *NIRVANA* run than in the *NIRVANA-III* run, in agreement with the evolution of the kinetic energies in Fig. 7, and the *NIRVANA* run shows a greater degree of structure in the velocity perturbations, perhaps indicative of higher order modes being present at this time. Overall the comparison is very satisfactory and demonstrates that the conditions for the vertical shear instability to occur are predicted accurately by both numerical schemes, which also show reasonable agreement for the growth rates under different radial temperature profiles.

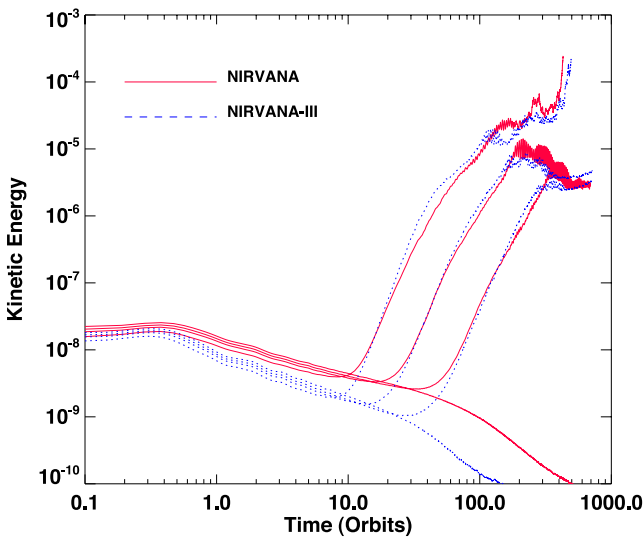


Figure 7. Comparison of the time evolution of the normalized (meridional + radial) kinetic energies for models T1O-0 to T4O-0 that were run with the two codes.

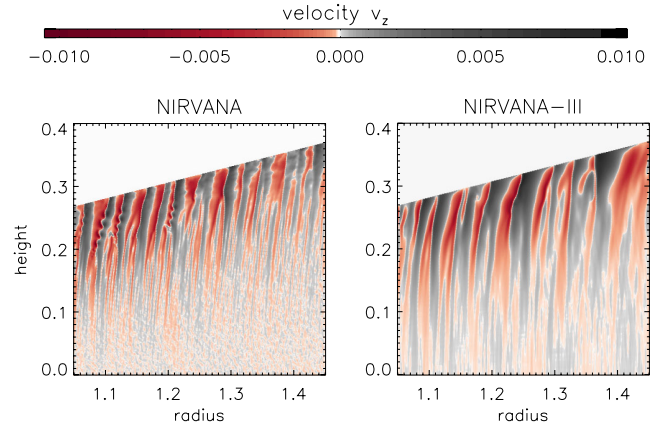


Figure 8. Comparison between the vertical velocity perturbations at time $t = 12.8$ orbits for the *NIRVANA* and *NIRVANA-III* runs. The results are for simulation T1O-0.

5.4 Evolution as a function of viscosity

In this section, we present results from simulations that examine the amplitude of the saturated state as a function of imposed viscosity. We apply a constant kinematic viscosity, ν , to the disc model T1R-0 and vary its value in the range $10^{-8} \leq \nu \leq 10^{-5}$ (a value of $\nu = 10^{-6}$ corresponds to the Shakura–Sunyaev viscous stress parameter $\alpha = 4 \times 10^{-4}$ at $R = 1$ (Shakura & Sunyaev 1973), and to a Reynolds number $\text{Re} = Hc_s/\nu = 2500$). The results are shown in Fig. 9, which displays the time evolution of the perturbed meridional plus radial kinetic energies. As expected, the results have a strong dependence on viscosity. For $\nu = 10^{-5}$ the instability is damped completely, which explains why previous 3D simulations of locally isothermal discs have not been reported seeing the vertical shear instability (e.g. Kley et al. 2001; Cresswell & Nelson 2006; Fromang et al. 2011). For decreasing values of ν , the amplitude of instability increases, until at a value of $\nu = 10^{-8}$ there is little

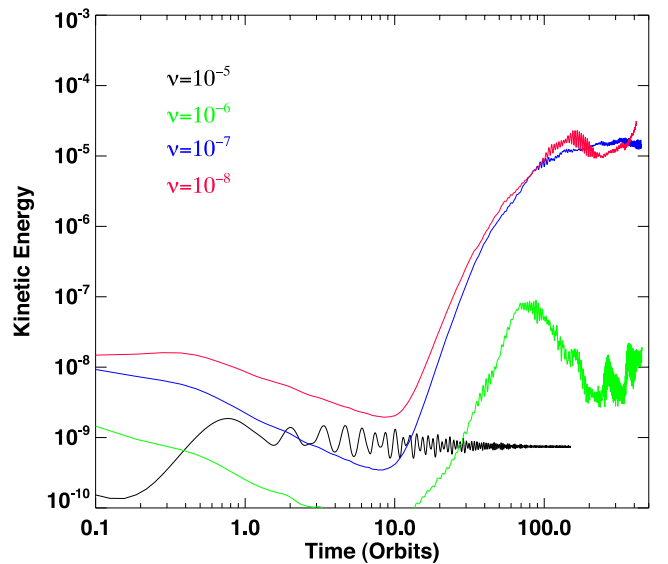


Figure 9. Time evolution of the normalized perturbed kinetic energy in the meridional and radial coordinate directions for model T1R-0 with $p = -1.5$, $q = -1$ and reflecting boundary conditions at the meridional boundaries. Each curve corresponds to a different value of the imposed kinematic viscosity, ν , as indicated.

difference between the result in Fig. 9 and the inviscid result shown in the left-hand panel of Fig. 6.

Interestingly, it is found that in fully turbulent models where the MRI is active throughout the disc and $\alpha \simeq 0.01$, the corrugation instability is not observed (Fromang & Nelson 2006). As part of this project, we have computed models similar to those presented in Fromang & Nelson (2006) and find that corrugation of the disc does not develop. Although these MHD simulations adopt a significantly lower resolution than the pure hydrodynamic runs we have presented here, we note that hydrodynamic runs performed at low resolution still show the development of the instability even when the short radial wavelength perturbations of the initial growth phase are not resolved. Instead, the disc displays growth of the longer wavelength breathing and corrugation modes that eventually cause the disc to undergo violent vertical oscillations. This suggests that the short-wavelength finger modes discussed earlier in the paper are not actually required for the disc to achieve a highly disturbed state. In magnetized global disc models with dead zones whose vertical height covers $\simeq 2.5$ scaleheights, which support Reynolds stresses in the dead zone with an effective value of $\alpha \simeq 10^{-4}$, the development of these corrugation oscillations is observed in models that adopt a locally isothermal equation of state with $q = -1$.

5.5 Evolution as a function of H/R

As discussed later in Section 6.1, a linear analysis of the vertical shear instability predicts that the ratio v_z/v_R scales as $\sim (H/R)^{-1}$, such that we expect $e_\theta/e_r \sim (H/R)^{-2}$. Indeed, we show that this scaling is implicit within the original analysis of Goldreich & Schubert (1967). Here, we present a suite of high-resolution simulations that examine this scaling. We consider disc models with $H/R = 0.1, 0.05, 0.03, 0.02$ and 0.01 . These models are labelled T5R-0 to T9R-0 in Table 1. The radial domain for each run covers just 5 scaleheights, and the full vertical domain covers 10 scaleheights ($\pm 5H$ about the mid-plane). Given that we wish to examine the evolution of e_θ/e_r from the earliest stages of linear growth, we have reduced the amplitude of the initial seed noise from $0.01c_s$ down to $10^{-6}c_s$.

The growth of $(e_r + e_\theta)$, normalized to the kinetic energy in Keplerian motion, is shown in the left-hand panel of Fig. 10. We clearly see that the growth times increase with decreasing H/R as expected from linear theory (see Section 6.1). The right-hand panel shows the growth of the perturbed kinetic energy for the $H/R = 0.05$ case in a log-linear plot. The dashed line shows the gradient expected for a growth time equal to two orbital periods, and we can see that during the earliest phase of linear growth, the growth time is slightly smaller than this (approximately 1.7 orbital periods) and increases as the disc evolves. If we compare this with the $H/R = 0.05$ fiducial model, T1R-0, discussed in Section 5.1, for which the initial velocity perturbations were stronger ($\sim 0.01c_s$, see Fig. 1), we notice that the linear growth phase for the model with smaller perturbations lasts longer. Furthermore, the growth time for this model during the early linear phase is found to be shorter than reported for model T1R-0, where a growth time of ~ 4 orbits was reported. One possible explanation for this is that the smaller amplitude initial perturbations allow for a different set of modes to grow in the disc at early times that have higher growth rates than the fundamental breathing and corrugation modes that characterize the later stage evolution. Indeed, the linear perturbation analysis that we present in Section 6.2 predicts that fundamental corrugation modes will have growth times of ~ 4 orbits, and fundamental breathing modes will have growth times of ~ 2.5 orbits. The fastest growing body modes, however, are predicted to be the first overtone breathing

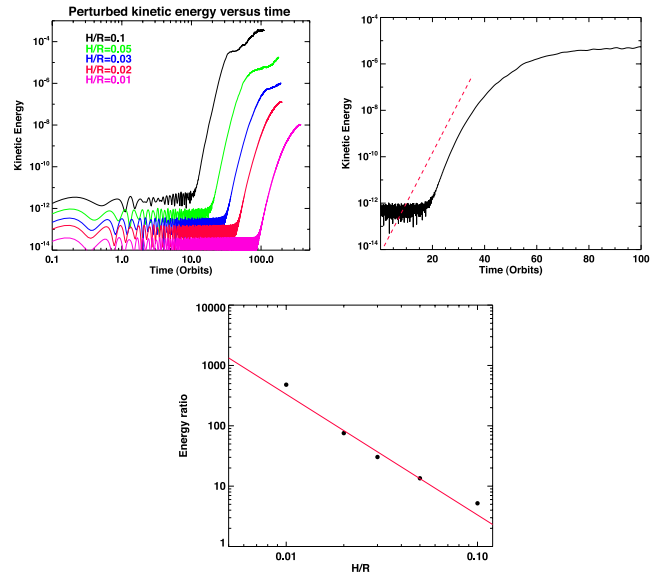


Figure 10. Left-hand panel: time evolution of the perturbed meridional plus radial kinetic energies (normalized) as a function of the disc thickness H/R . Right-hand panel: time evolution of the perturbed meridional plus radial kinetic energy (normalized) for the $H/R = 0.05$ run T8R-0. The dashed line shows the gradient expected for a linear growth time equal to 2 orbits. Lowest panel: maximum ratio of the volume-integrated meridional to radial kinetic energies versus H/R during the runs described in the text.

mode and the second overtone corrugation mode, with growth times of ~ 1.7 orbits. We note that this latter growth time is very similar to that measured during the earliest linear growth phase in the right-hand panel of Fig. 10.

Returning to the question of how e_θ/e_r scales with H/r , the lower panel in Fig. 10 shows the maximum value of e_θ/e_r obtained during each simulation, represented by the black dots. The red line represents a by-eye fit that scales with $(H/R)^{-2}$, demonstrating that the vertical-to-radial energy ratio approximately follows this scaling in the simulations, in agreement with the analysis presented in Section 6.1.

Contour plots of v_z for the run with $H/R = 0.05$ are shown in Fig. 11 at three different times. This figure has already been discussed briefly in Section 5.1, where the appearance of both high-latitude finger modes and lower latitude body modes was brought to attention. This behaviour is seen in all models with different H/R , although the distinct appearance of the two types of modes is most clearly seen in thicker discs. The model with $H/R = 0.01$, for example, shows evidence for the coexistence of the two mode types, but their distinct character is not so easy to discriminate in that run. We further comment that the first panel in Fig. 11 shows that in the outer regions of the disc, there is evidence for the first overtone breathing mode being present at early evolution times. As discussed above, according to the linear theory presented in Section 6.2, this mode is expected to grow before the other body modes, so its presence in the disc is expected. At earlier evolution times than shown in Fig. 11, we find that its presence can also be discerned in the inner disc regions, supporting our conjecture that the rapid linear growth of the perturbation energy at early times observed in the right-hand panel of Fig. 10 may arise because of the early growth of this mode. The evolution of velocity perturbations appears to follow the predicted pattern of first overtone breathing modes arising first, followed by fundamental breathing modes, with the late linear phase and non-linear phase being dominated by fundamental

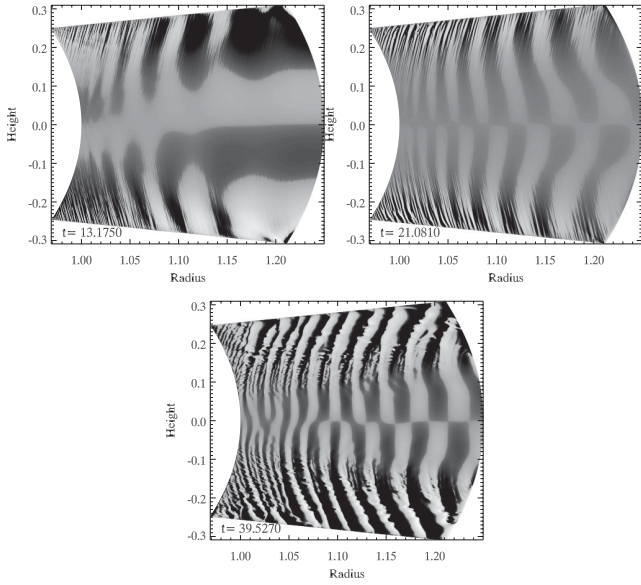


Figure 11. Edge-on contours of the perturbed vertical velocity as a function of R , Z and time for the model with $H/R = 0.05$ initiated with random perturbations of amplitude $10^{-6}c_s$. Note that for clarity, the grey-scale of the image has been stretched by plotting the quantity $\text{sign}(v_z) \times |v_z|^{0.35}$.

corrugation modes. A more quantitative analysis of the growth rates associated with the individual modes and their contributions to the overall perturbed kinetic energy budget will be presented in future work.

5.6 Thermal relaxation in models with $T(R)$

We now consider the evolution of models where we relax the locally isothermal assumption associated with the fluid response to perturbations. We evolve the energy equation in (1) and introduce thermal relaxation by integrating equation (18). We adopt the equation of state $P = (\gamma - 1)e$ and set $\gamma = 1.4$. The gas is assumed to be inviscid. Power-law profiles for the initial temperatures, $T(R)$, and mid-plane density, $\rho_{\text{mid}}(R)$, are adopted with $q = -1$ and $p = -1.5$ in equations (2) and (3). The aim of these models is to examine the robustness of the vertical shear instability as a function of the thermal relaxation time, τ_{relax} , defined in equation (18) and expressed as a fixed multiple or fraction of the local orbital period. These runs are labelled T10R-0.01 to T14R- ∞ in Table 1.

The evolution of the normalized perturbed kinetic energies for models with relaxation times in the range $0 \leq \tau_{\text{relax}} \leq \infty$ is plotted in Fig. 12. It is immediately obvious that instability only occurs in either the locally isothermal case ($\tau_{\text{relax}} = 0$) or when $\tau_{\text{relax}} = 0.01$ orbits. All other simulations result in the perturbed kinetic energy contained in the initial seed noise decaying with time. We note that the case of $\tau_{\text{relax}} = \infty$ is directly comparable to a previous study on the adiabatic evolution of a stratified disc by Rüdiger et al. (2002). The authors considered the hydrodynamic stability under the Solberg–Høiland criterion and also find stability in this case.

Our results indicate that the vertical shear instability requires that the initial temperature profile of the fluid is re-established rather rapidly during dynamical evolution, at least for the equilibrium temperature and density profiles adopted in these particular models. We discuss the implications of this for applicability of the instability to protoplanetary discs in Section 7.1.

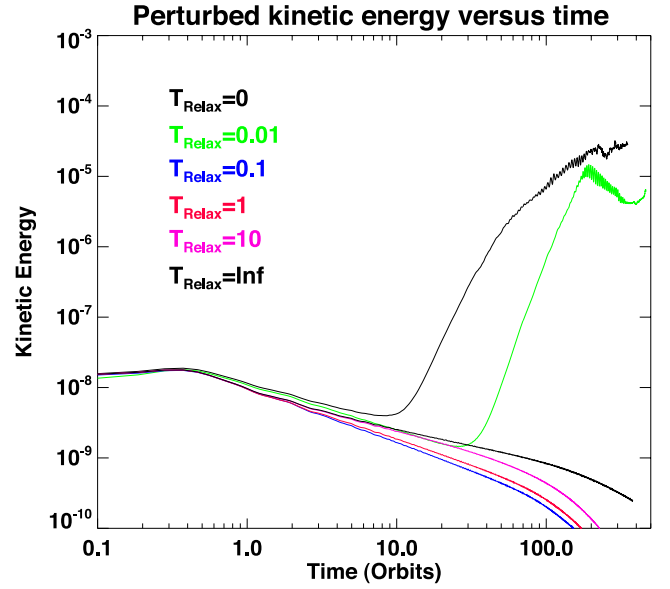


Figure 12. Time evolution of the sum of the (normalized) perturbed radial and meridional kinetic energies in discs where the temperature was initially constant on cylinders, as a function of the thermal relaxation time. Note that only the $\tau_{\text{relax}} = 0$ and 0.01 cases show growth.

5.7 Thermal relaxation in models with $K_s(R)$

We now consider models in which the initial entropy function, $K_s(R)$, follows a strict power-law function of radius given by equation (6), and the mid-plane density, $\rho_{\text{mid}}(R)$, follows a radial power-law given by equation (3). For all models except one, we adopt the values $s = -1$ and $p = 0$, leading to the mid-plane Mach number, \mathcal{M}_{mid} , being constant at all radii. Our normalization of $K_s(R_0)$ sets $\mathcal{M}_{\text{mid}} = 20$. We also consider a single model with $p = -1.5$ and $s = 0$, so that there is no initial radial entropy gradient. The entropy in this case is normalized so that $\mathcal{M}_{\text{mid}} = 20$ at $R = R_0 = 1$. These runs are listed in Table 1 as K1R-0 to K7R-0.01.

We impose reflecting conditions at the meridional boundaries and consider inviscid evolution. As described in Section 2.1.1, these models are convenient to implement numerically because analytic solutions can be obtained for the equilibrium density and velocity fields through equations (3) and (16). As such, these models allow us to explore the vertical shear instability as a function of the thermal relaxation time, τ_{relax} , in discs where the initial distribution of temperature no longer follows a power-law function of the cylindrical radius, but instead varies with both R and Z . We note that equation (16) also demonstrates that a radial power law in $K_s(R)$ with $s = -1$ gives rise to an equilibrium v_ϕ that varies with height Z at each radius R , but adopting $s = 0$ gives v_ϕ independent of Z . The normalized sum of the radial and meridional kinetic energies is plotted in Fig. 13 for thermal relaxation times in the range $0 \leq \tau_{\text{relax}} \leq \infty$ local orbits. The model with $s = 0$ and $p = -1.5$ employed a relaxation time $\tau_{\text{relax}} = 0.01$ orbits and is labelled ‘ $T_{\text{relax}} = 0.01$, $s = 0$ ’.

Interestingly, for the model with $s = -1$ and $p = 0$, all values of τ_{relax} result in growth of the perturbed meridional and radial energies, and only the strictly adiabatic simulation with $\tau_{\text{relax}} = \infty$ shows eventual decay over long time-scales of ~ 100 orbits. It is noteworthy that this adiabatic model is unstable according to the second Solberg–Høiland criterion discussed in Section 3, and this is probably the cause of the initial growth in perturbation energy. The subsequent decay may arise because the instability causes the

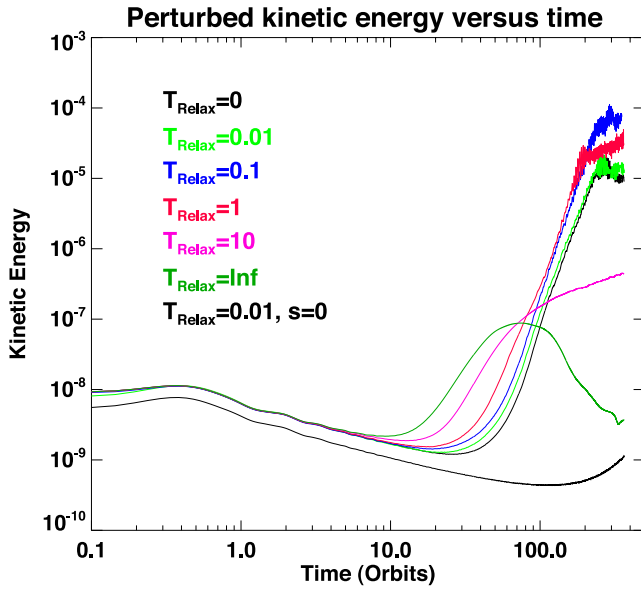


Figure 13. Time evolution of the sum of the normalized radial and meridional kinetic energies in discs where the entropy function, $K_s(R)$, was initially constant on cylinders.

equilibrium disc to move to a stable state. Inspection of perturbed velocity profiles in contour plots similar to Figs 2 and 3 (not shown here) for all of the runs with $p = 0$ and $s = -1$ shows that the previously discussed characteristic high-latitude finger perturbations with $|k_R/k_Z| \gg 1$ grow initially and descend towards the mid-plane, even for the adiabatic disc model with no thermal relaxation. Later development of the body modes that lead to breathing followed by corrugation of the disc is seen clearly in the discs with thermal relaxation time $\tau_{\text{relax}} \leq 1$ orbit, and in the disc with $\tau_{\text{relax}} = 10$ orbits velocity contours indicate that this model is also transitioning to this state on long time-scales. In agreement with these observations, all models with finite thermal relaxation time show long-term growth in perturbation energy, although in the case of $\tau_{\text{relax}} = 10$ orbits the growth time is very long indeed.

The instability displayed by the models with $\tau_{\text{relax}} \leq 10$ orbits shows that the requirement of very rapid thermal relaxation observed in the models presented in Section 5.6 depends on the detailed temperature and density structure of the disc, with the entropy gradients and their stabilizing/destabilizing effects probably playing an important role. Disc models clearly exist for which thermal relaxation times in the range $0 \leq \tau_{\text{relax}} \leq 10$ orbits lead to the growth of the instability, and as such its range of applicability to the dynamics of astrophysical discs may be broader than suggested by the results presented in Section 5.6.

A full exploration and understanding of the range of disc models that display GSF-like instabilities, and the thermal time-scales required as a function of the density, temperature and entropy profiles goes beyond the more limited remit of this paper. Contrasting the very small thermal relaxation times required for the discs with $T(R)$, however, with the less stringent requirements for the discs with $K_s(R)$, we note that the discs which are isothermal in Z have vertical entropy profiles that are highly stabilizing because $\partial S/\partial Z > 0$, as we have discussed already in the context of the Solberg–Høiland criteria (see Section 3.1). It is therefore not surprising that instability for these discs only occurs in the presence of rapid thermal relaxation, as this is required to overcome the stabilizing entropy gradient. The discs for which entropy varies only with R , however,

are neutrally stable to vertical convection, and therefore there is no entropy gradient to provide a stabilizing influence. Indeed, the disc with $p = 0$ and $s = -1$ is unstable according to the Solberg–Høiland criterion, but the non-linear evolution does not lead to sustained unstable behaviour in the absence of thermal evolution, as the disc appears able to find a neighbouring stable equilibrium. The role of thermal evolution, therefore, must be to enhance the instability of this model by increasing the influence of the $\partial S/\partial R$ and $\partial S/\partial Z$ terms in equation (23), preventing the disc from accessing a stable equilibrium. Given that this disc model is unstable according to the Solberg–Høiland criterion, we should not be surprised to find that a moderate rate of thermal forcing is required to drive sustained unstable behaviour compared with models for which the vertical entropy gradient is stabilizing. In fact, it is likely that models with null vertical entropy gradients and non-zero vertical angular momentum gradients are optimal for the development of the GSF instability for long thermal relaxation times.

Turning to the run with $p = -1.5$ and $s = 0$, we see that the initial perturbation energy decreases for the first ~ 200 orbits before increasing again. The expectation is that this model will not display the vertical shear instability, and inspection of velocity contour plots (not shown here) confirms that the characteristic perturbations with $|k_R/k_Z| \gg 1$ do not appear in this case. These velocity plots, however, indicate that over secular time-scales sound waves are generated close to the meridional boundaries, and this appears to be the reason for the up-turn in the perturbed kinetic energies seen in Fig. 13 after 200 orbits.

5.8 Non-axisymmetric model

We now consider briefly the evolution of a non-axisymmetric model T1R-0-3D, in which the azimuthal domain covers $\pi/4$ radians. This model is the 3D equivalent of model T1R-0. The simulation was performed using the NIRVANA code, and the model was set up using equations (3), (2) and (13), with values $p = -1.5$ and $q = -1$. The velocity field was seeded with noise (amplitude $0.01c_s$). Details are given in Table 1.

The total normalized kinetic energy (meridional + radial) versus time is displayed in Fig. 14. Comparing with the equivalent

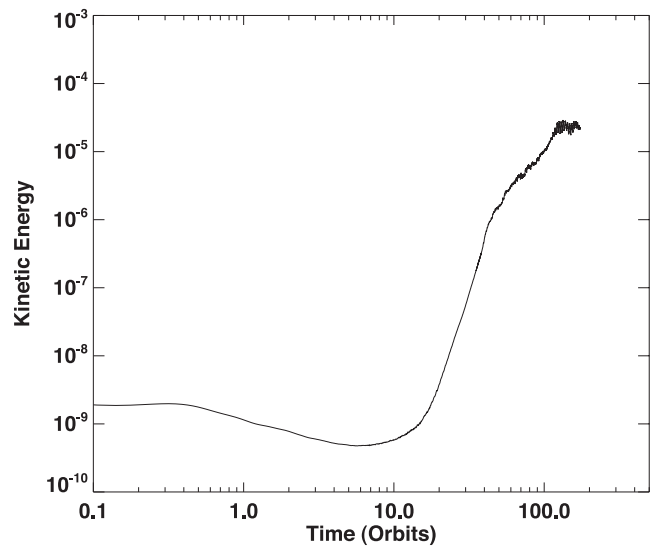


Figure 14. Time evolution of the perturbed meridional + radial kinetic energies (normalized by the total energy in Keplerian motion) for the full 3D simulation T1R-0-3D.

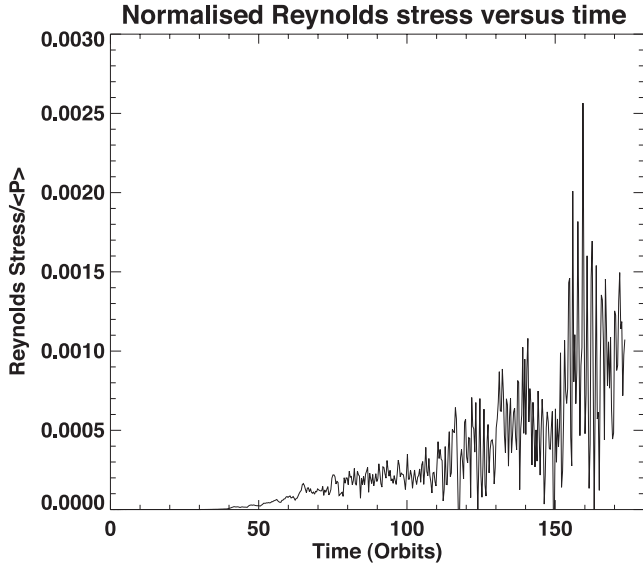


Figure 15. Time evolution of the volume-averaged Reynolds stress (normalized by the mean pressure) for the full 3D simulation T1R-0-3D.

axisymmetric plot (Fig. 6), we see that the evolution is similar, with growth in the perturbed energy occurring after ~ 10 orbits and saturation at a value of a few $\times 10^{-5}$ beginning to occur after ~ 100 orbits.

In Fig. 15, we plot the time evolution of the volume-averaged Reynolds stress normalized by the mean pressure in the disc. This quantity is computed as follows. We define an azimuthally averaged Reynolds stress $T_R(r, \theta)$ obtained by averaging the quantity $\rho \delta v_R \delta v_\phi$ over azimuth. Here δv_R and δv_ϕ are the local radial and azimuthal velocity fluctuations. We also define a density-weighted mean pressure as a function of r , $\bar{P}(r)$, obtained by averaging over θ and ϕ . We define a local value of the Shakura–Sunyaev stress parameter $\alpha(r, \theta) = T_R(r, \theta) / \bar{P}(r)$. The simple arithmetic average of $\alpha(r, \theta)$ over r and θ is the quantity plotted in Fig. 15.

Although rather noisy, we see that the normalized stress starts to level off at a value $\alpha \sim 10^{-3}$ by the end of the simulation. The spatial distribution of $\alpha(r, \theta)$, time averaged between $t = 130$ and 140 orbits, is shown in Fig. 16. Here we see that local values of the stress reach $\sim 2 \times 10^{-3}$, indicating that the vertical shear instability generates a quasi-turbulent flow capable of supporting significant outward angular momentum transport in astrophysical discs, given favourable conditions for its development.

The upper panels of Fig. 17 show contours of the perturbed density, $\delta \rho / \rho_0$, in a slice parallel to the meridional plane at three different times during the simulation, showing similar features to those presented for the 2D axisymmetric simulation in Fig. 5. Perhaps more interesting are the lower panels of Fig. 17 which show the actual density ρ in the (R, ϕ) plane located at the disc mid-plane. Here the development of spiral density waves may be observed, similar in morphology to those that arise in discs where turbulence is driven by the MRI (e.g. Papaloizou & Nelson 2003; Heinemann & Papaloizou 2009). The 3D simulation presented here suggests that if the appropriate conditions prevail in astrophysical discs, the vertical shear instability may lead to a turbulent flow capable of supporting significant angular momentum transport.

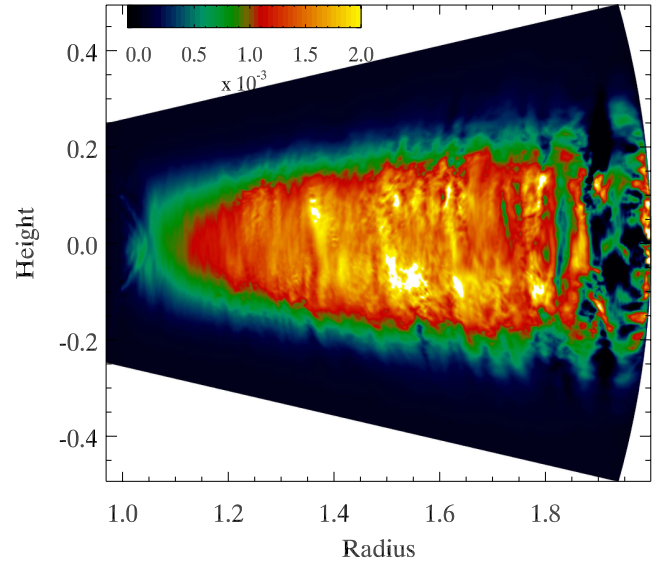


Figure 16. Spatial distribution of the time and horizontally averaged Reynolds stress (normalized by the mean pressure at each radius) for model T1R-0-3D.

6 THEORETICAL CONSIDERATIONS

The GSF instability appears by rendering the inertial modes of a rotating atmosphere unstable. The original analysis in Goldreich & Schubert (1967, GS67 hereafter) demonstrated the possibility of this instability by performing a point analysis at a given location in a stellar radiative zone away from the equator, equivalent to considering a location away from the mid-plane in a disc. In this section, we examine the instability by extending previous analyses, including those of Urpin (2003) and Arlt & Urpin (2004). We begin by reworking the point analysis of GS67 for a locally isothermal equation of state, without the Boussinesq assumption, in order to highlight some of the main results that are inherent to that analysis. We then extend this analysis by relaxing the point assumption. The mathematical details associated with this extended analysis are presented in Appendix A. To maintain the narrative flow in the main body of the paper, we present a shortened discussion of this extended analysis in Section 6.2, highlighting the assumptions employed and the main results obtained regarding the emerging modes and their growth rates.

6.1 Reworking the GS67 analysis with compressible, local isothermal assumption

Owing to the importance of the original study presented by GS67, we repeat here the calculation contained in that work wherein we assume that the equation of state of the gas is locally isothermal with a temperature profile varying in the nominal radial direction. We also extend that analysis to include the effects of compressibility and we do *not* assume outright that the hydrodynamic flow is incompressible.

The original analysis was performed in a local rotating reference frame (the original intention being to examine the possibility of such instabilities in the interiors of rotating stars). The equations of motion considered in that work are akin to the local shearing sheet approximation familiar to accretion disc theory (Goldreich & Lynden-Bell 1965). Since this is our concern here, we shall use that

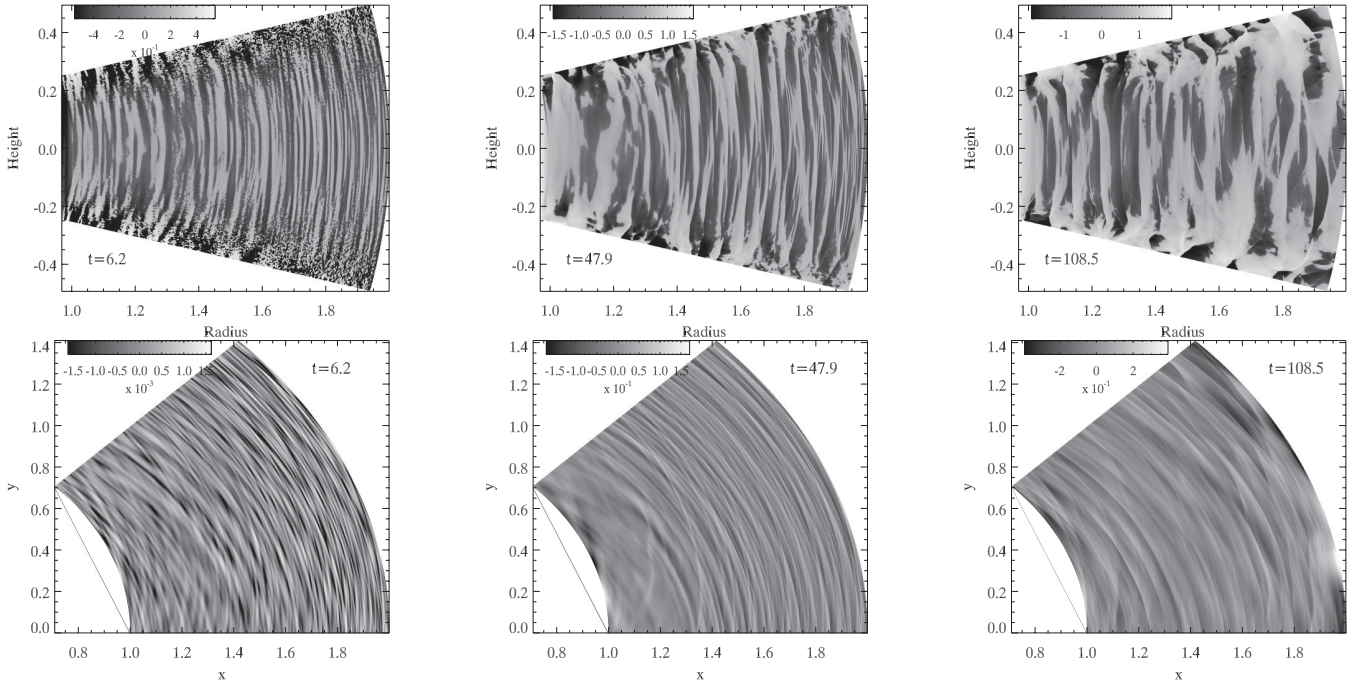


Figure 17. Perturbed density, $\delta\rho/\bar{\rho}$, in the meridional plane at $\phi = \pi/8$ (upper three panels) for the 3D simulation T1R-0-3D. Note that we have effectively stretched the grey-scale by plotting the quantity $\text{sign}(\delta\rho) \times |\delta\rho/\bar{\rho}|^{1/4}$ in the upper panels. The lower panels show the relative density perturbations $\delta\rho/\bar{\rho}$ at the disc mid-plane. No grey-scale stretching has been applied to these lower panels.

as our starting model set. The equations for axisymmetric dynamics in that setting are

$$\begin{aligned} \frac{du}{dt} - 2\Omega_0 v &= -c_s^2 \frac{\partial \Pi}{\partial x} - \frac{\partial c_s^2}{\partial x} + 3\Omega_0^2 x, \\ \frac{dv}{dt} + 2\Omega_0 u &= 0, \\ \frac{dw}{dt} &= -c_s^2 \frac{\partial \Pi}{\partial z} - g, \\ \frac{d\Pi}{dt} + \frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} &= 0, \end{aligned} \quad (28)$$

where $\Pi = \ln \rho$ and $\Omega_0 = \text{constant}$. The remaining expressions are defined for convenience below. Lower case letters are used for the radial, azimuthal and vertical velocities (i.e. u , v , w) in the Cartesian shearing box in order to distinguish them from the ones used to describe dynamics in a cylindrical geometry examined in Appendix A. In general, the vertical component of gravity is vertically varying and is given by $g = \Omega_0^2 z$, but for this analysis it is treated as a constant.

As done in GS67, one can do a point expansion (see the discussion of GS67 immediately before equation 17 of that work) around any nominal level $z = z_0$. We start by considering the mean states which we represent with overbars. Radial and vertical equilibria give

$$\begin{aligned} -2\Omega_0 \bar{v} &= -c_s^2 \frac{\partial \bar{\Pi}}{\partial x} - \frac{\partial c_s^2}{\partial x}, \\ 0 &= -c_s^2 \frac{\partial \bar{\Pi}}{\partial z} - g. \end{aligned}$$

Note that the mean azimuthal flow state \bar{v} has been decomposed into a Keplerian component (the term $-(3/2)\Omega_0 x$) plus a deviation

about that state \bar{V} , i.e. $\bar{v} = -q\Omega_0 x + \bar{V}$, and c_s^2 is the sound speed (implicitly a function of x because of the radial dependence of the vertically isothermal temperature profile). In steady state, we find that

$$\frac{\partial \bar{\Pi}}{\partial z} = -\frac{g}{c_s^2},$$

and, most importantly, the mean gradient of the azimuthal flow is

$$\bar{V}_z \equiv \frac{\partial \bar{V}}{\partial z} = \frac{g}{2\Omega_0} \frac{\partial \ln c_s^2}{\partial x}. \quad (29)$$

Perturbations of Π and all other variables around their reference mean states are introduced with prime notation, e.g. $\Pi \rightarrow \bar{\Pi} + \Pi'$, etc. We assume an isothermal equation of state for the perturbations as well. Linearized perturbations of the equations of motion (28) lead to the following expressions:

$$\begin{aligned} \frac{\partial u'}{\partial t} - 2\Omega_0 v' &= -c_s^2 \frac{\partial \Pi'}{\partial x}, \\ \frac{\partial v'}{\partial t} + \Omega_0 (1/2 + \bar{V}_x) u' + \bar{V}_z w' &= 0 \\ \frac{\partial w'}{\partial t} &= -c_s^2 \frac{\partial \Pi'}{\partial z}, \\ \frac{\partial \Pi'}{\partial t} + u' \frac{\partial \bar{\Pi}}{\partial x} + w' \frac{\partial \bar{\Pi}}{\partial z} + \frac{\partial u'}{\partial x} + \frac{\partial w'}{\partial z} &= 0. \end{aligned} \quad (30)$$

Note we have utilized the shortened notation $\bar{V}_x \equiv \partial \bar{V} / \partial x$. To emphasize, the analysis we carry out here departs from that done in GS67 in two respects: (i) we assume an isothermal equation of state for the disturbances and (ii) we allow for compressibility (cf. equation 29 of GS67). Since g is constant, the above equations are

easily combined into a single one for Π' yielding

$$\begin{aligned} \frac{\partial^4 \Pi'}{\partial t^4} - \left[\left(\frac{\partial \bar{\Pi}}{\partial z} + \frac{\partial}{\partial z} \right) c_s^2 \frac{\partial}{\partial z} + \left(\frac{\partial \bar{\Pi}}{\partial x} + \frac{\partial}{\partial x} \right) c_s^2 \frac{\partial}{\partial z} - \kappa_0^2 \right] \frac{\partial^2 \Pi'}{\partial t^2} \\ + \left[2\Omega_0 \bar{V}_z \left(\frac{\partial \bar{\Pi}}{\partial x} + \frac{\partial}{\partial x} \right) c^2 - \kappa_0^2 \left(\frac{\partial \bar{\Pi}}{\partial z} + \frac{\partial}{\partial z} \right) c_s^2 \right] \frac{\partial \Pi'}{\partial z} = 0, \end{aligned} \quad (31)$$

where the epicyclic frequency is represented by κ_0 and is related to the steady-state quantities by

$$\kappa_0^2 = 2\Omega_0^2(1/2 + \bar{V}_x).$$

Paraphrasing directly from GS67 (prior to equation 17 of that work), the next step is to expand the unperturbed variables and their derivatives in Taylor series about some point $x = x_0$ and $z = z_0$. Discarding terms of order $(x - x_0)/x_0$ and $(z - z_0)/z_0$, the perturbation variables may be expanded in plane waves of the form $\sim e^{i(\omega t + k_x x + k_z z)}$, revealing the dispersion relationship

$$\begin{aligned} \omega^4 - [c_0^2(k_x^2 + k_z^2) + i g k_z + \kappa_0^2] \omega^2 \\ - 2\Omega_0 \bar{V}_z c_0^2 k_x k_z + \kappa_0^2 (c_0^2 k_z^2 + i k_z g) = 0, \end{aligned}$$

in which c_0^2 is the sound speed at the point in question and where $\kappa_0 \rightarrow \Omega_0$. As is standard for problems of atmospheres, we can recast the vertical wave dependence to have a complex character (indicating a basic wavy pattern) by defining $k_z \rightarrow k_z - i g/(2c_0^2)$ which renders the dispersion relation into the form

$$\begin{aligned} \omega^4 - [c_0^2(k_x^2 + k_z^2) + \kappa_0^2 + N_0^2] \omega^2 \\ - 2\Omega_0 \bar{V}_z c_0^2 k_x k_z + \kappa_0^2 (c_0^2 k_z^2 + N_0^2) = 0, \end{aligned} \quad (32)$$

where we have defined the Brunt–Vaisaila frequency

$$N_0^2 = \frac{g^2}{4c_0^2}.$$

Solutions for ω are easy to obtain and write down. It is more instructive, however, to assess the stability characteristics straight from an analysis of equation (32) itself.

The classical GSF instability is one in which the modes pass through zero frequency before becoming unstable and they describe the influence that the vertical shear gradient has upon *inertial modes*. In our dispersion relation, this amounts to the instability condition

$$-2\Omega_0 \bar{V}_z c_0^2 k_x k_z + \kappa_0^2 (c_0^2 k_z^2 + N_0^2) < 0. \quad (33)$$

When $c_0^2 k_z^2 \gg N_0^2$ the condition reduces to

$$-2\Omega_0 \bar{V}_z + \kappa_0^2 \frac{k_z}{k_x} < 0,$$

which is essentially the first term appearing in equation 33 of GS67. Also, this condition is identical to that found in equation 20 of Urpin (2003). These latter correspondences follow from realizing that

$$\bar{V}_z \leftrightarrow \frac{\partial \Omega}{\partial z}, \quad \frac{\kappa_0^2}{2\Omega_0} \leftrightarrow \frac{1}{R_0^2} \left(\frac{\partial R^2 \Omega}{\partial R} \right)_{R=R_0}$$

and, thus, recovering the classical GS67 condition. The dispersion relation (32) has the general form

$$\omega^4 - B\omega^2 + C = 0,$$

where B and C are obviously identified with the terms in (32). This equation has imaginary solutions for ω if (i) $C < 0$ with $B > 0$ or (ii) $B^2 - 4C < 0$. The GSF criterion (33) is essentially the condition that

$C < 0$. Condition (ii) is for acoustic-inertial modes to be unstable. This amounts to

$$\begin{aligned} [c_0^2(k_x^2 + k_z^2) + \kappa_0^2 + N_0^2]^2 \\ - 4[-2\Omega_0 \bar{V}_z c_0^2 k_x k_z + \kappa_0^2 (c_0^2 k_z^2 + N_0^2)] < 0. \end{aligned} \quad (34)$$

A detailed examination of the behaviour of this condition shows that it does not occur for reasonable values of the parameters leading us to conclude that the acoustic modes (per se) do not become unstable as well.

Denoting σ as the growth rate, we find that the inertial mode response is roughly given by

$$\sigma^2 = \frac{-\kappa_0^2 (c_0^2 k_z^2 + N_0^2) + 2\Omega_0 c_0^2 k_x k_z \frac{\partial \bar{V}}{\partial z}}{c_0^2 (k_x^2 + k_z^2) + \kappa_0^2 + N_0^2}, \quad (35)$$

k_z and k_r are the corresponding vertical and radial disturbance wavenumbers, respectively (note that we have changed notation so that $k_x \rightarrow k_r$ and have reinstated the vertical velocity gradient so that it is now explicit). The quantity $\partial \bar{V}/\partial z$ scales of the order of magnitude of $(q/2)\Omega_0(H_0/R_0)$, where q is the same exponent of the radially varying isothermal sound speed discussed in Section 2. Supposing for this discussion that N_0 is negligible, it follows from this expression that if H_0/R_0 is small, then instability can only happen if the radial wavenumber conspires to be correspondingly large. In that limit, the above expression implies

$$\sigma^2 \sim 2\Omega_0 \frac{k_z}{k_r} \frac{\partial \bar{V}}{\partial z} - \kappa_0^2 \frac{k_z^2}{k_r^2}, \quad (36)$$

indicating that instability is possible if $k_z/k_r \sim \mathcal{O}(qH_0/R_0)$. The analysis of Arlt & Urpin (2004), for example, also similarly indicates that for the same rough conditions the growth rate ought to scale as $\mathcal{O}(q\Omega_0 H_0/R_0)$. The simulations we have performed are consistent with this tendency where the radial length-scales of the emerging structures are significantly shorter than the vertical ones with growth rates of the instability $\lesssim 4$ orbit times for $H_0/R_0 \sim 1/20$ and $q = -1$.

The goal of our extended analysis discussed below is to develop a better physical understanding of the processes responsible for this instability beyond invoking Solberg–Høiland criteria. In this respect, we notice from Fig. 1 that the integrated radial energies are smaller in magnitude than the corresponding integrated meridional energies during the growth phase of the linear instability. Furthermore, in Fig. 10 we showed that the ratio of meridional to radial energies observed in the simulations scales as $e_\theta/e_r \sim (H/R)^{-2}$. We now examine whether or not this conforms to what is expected of the GSF instability.

If we combine the time-scales ($\sim q\Omega_0 H_0/R_0$) and length-scales ($k_z \sim k_r H_0/R_0$) implied by the unstable mode with equations (30) describing the point analysis found in GS67, then we find that the relative orders of magnitude of the various terms in the perturbed continuity equations (30) are

$$-w' \frac{g}{c_s^2} + \frac{\partial w'}{\partial z} + \frac{\partial u'}{\partial x} = \mathcal{O} \left(w' \frac{H_0}{R_0} \right)^2, \quad (37)$$

where u' , w' are the perturbed (cylindrical) radial and vertical velocities, respectively. Now since $c_s^2/g \sim \mathcal{O}(H_0)$ and if we assume that the vertical perturbation scales k_z^{-1} are similarly $\sim H_0$, then we find the following: because in order for the instability to operate we need $k_r \sim k_z(H_0/R_0)^{-1}$, it necessarily follows that the only way to

non-trivially get the sum of the order 1 terms on the LHS of (37) to be approximately zero is for

$$u' = \mathcal{O}\left(w' \frac{H_0}{R_0}\right).$$

In other words, the radial perturbation velocities must be a factor of H_0/R_0 smaller than the perturbed vertical velocities while the instability is growing.¹ This immediately points out that those growing disturbances are largely *anelastic* rather than just simply incompressible – a fact we shall examine further in Appendix A. More importantly, if we define $r_e = e_r/e_z$ to be the ratio of the corresponding domain-integrated radial to vertical (perturbation) energies while the linear instability is active, then we expect that it should scale according to

$$r_e = \frac{e_r \approx \int_{\text{Vol}} \bar{\rho} u'^2 d^3x}{e_z \approx \int_{\text{Vol}} \bar{\rho} w'^2 d^3x} \sim \mathcal{O}\left(\frac{H_0}{R_0}\right)^2,$$

all other quantities being equal.

6.2 An extended linear analysis: approximate solutions and double instability

With these considerations in mind, in Appendix A we re-examine the linear instability without adopting the point-analysis strategy of GS67. Instead, we seek to express in a relatively simplified way the dynamics occurring when the radial length-scales are much smaller than the vertical scales assumed to be of the order of H_0 . Through careful scaling analysis we show how, subject to the adopted assumptions that are outlined below, the processes involved in bringing about the instability are largely *anelastic* and *radially geostrophic* – by the latter expression we mean to indicate dynamics which are in constant radial force balance between Coriolis effects and pressure gradients (note that these conditions are also implicit within the GS67 analysis). Related to this, we show how the preceding considerations based on the GS67 theory carry over: namely that while the instability is in its linear phase, the vertical velocities exceed the corresponding perturbed radial velocities by a factor of H_0/R_0 .

In Appendix A, we detail and motivate an asymptotic scaling analysis of the inviscid equations of motion by making use of the above observations. A short summary of what is assumed is the following: we examine radially propagating wave solutions about the fiducial radius R_0 which have radial length-scales of the order of H_0^2/R_0 and represented by the non-dimensional variable x , while the vertical length-scales are of the order of H_0 and denoted similarly by z . Times are scaled by $(2\pi/\Omega_0)(R_0/H_0)$, i.e. a factor of R_0/H_0 longer than the local orbit time, and denoted by τ . The vertical and azimuthal velocities are assumed to scale by c_s and the radial velocity is scaled by $c_s(H_0/R_0)$, and denoted, respectively, by \tilde{u} , \tilde{v} , \tilde{w} . We dynamically follow the density/pressure response as disturbances of the function $\Pi \equiv \ln \rho$, so that the linearization of this quantity $\Pi' = \rho/\rho_0$. Formally, in the expansion procedure Π' scales as a factor of H_0/R_0 of the non-dimensionalized vertical and azimuthal velocities and is represented in the calculation of $\tilde{\Pi}$. These assumed relative scalings are implemented into the equations

of motion which are then solved order by order in powers of H_0/R_0 until a non-trivial reduction is achieved. The fruit of this procedure results in the simpler system expressed in equations (A15)–(A18) which contains the aforementioned radial geostrophy vertical hydrostasy. These equations may be combined into a single equation for the pressure perturbation $\tilde{\Pi}$:

$$\frac{\partial^2}{\partial \tau^2} \frac{\partial^2 \tilde{\Pi}}{\partial x^2} = -\frac{\partial^2 \tilde{\Pi}}{\partial z^2} + \left(1 + q \frac{\partial}{\partial x}\right) z \frac{\partial \tilde{\Pi}}{\partial z}. \quad (38)$$

It is worth emphasizing here that a point analysis of equation (38), i.e. assuming that $z = z_0$ is fixed and making a wave ansatz and proceeding similarly to Section 6.1 above, recovers the content of the asymptotic growth rates contained in equation (36), indicating the consistency of the scaling arguments we have exploited to obtain this equation with the analysis in GS67.

Unfortunately, a summary examination of equation (38) indicates it to be challenging to solve despite it appearing so simple. This is because it is fundamentally inseparable between the x and z coordinates owing to the $q\partial/\partial x$ term on the RHS of the expression. However, we generate approximate solutions by assuming that disturbances have the form

$$\tilde{\Pi} = \Pi(z)e^{ikx + \sigma\tau} + \text{c.c.}$$

turning (38) into a boundary value problem for the unknown eigenvalue σ

$$\frac{d^2 \Pi}{dz^2} - (1 + ikq)z \frac{d\Pi}{dz} + \sigma^2 k^2 \Pi = 0, \quad (39)$$

allowing us to explore the vertical structure of the associated eigenmodes. The above is supplemented by the boundary condition that there is no vertical flow at the top and bottom of the disc. Since the bulk of the simulations of this study have implemented no-flow boundaries at five scaleheights, we impose that

$$\frac{d\Pi}{dz} = 0, \quad \text{at } z = \pm 5.$$

So long as the eigenvalue $\sigma \neq 0$, the vertical velocity vanishes at $z = \pm 5$ if the vertical gradient of Π vanishes there too (see equation A17). Solutions of the ordinary differential equation (39) subject to the above boundary conditions involve manipulating Hermite and hypergeometric functions of complex arguments which offer very little in the form of insight or ease of analysis. Instead, we opt for solving the above system on a Chebyshev collocation grid ($N = 177$ points) and solving the resulting linear matrix eigenvalue problem for σ .

Generally speaking (except see below), the eigenvalues take on the general form

$$\sigma(k) = \pm(|\sigma_r| + \text{sgn}(k)|\sigma_i|),$$

so that for each left or right going wave, there are two responses of the system, one being decaying while the other is growing. The response of the system is composed primarily of *body modes* which correspond to ‘corrugation’ or ‘breathing’ modes depending upon the vertical symmetry of the corresponding pressure perturbation eigenfunction (odd for corrugation and even for breathing modes). In the figures and corresponding eigenvalue portraits found in Fig. 19, these modes are denoted by C_n and B_n , respectively, where the subscript n is a positive index starting from 1. C_1 and B_1 are, respectively, referred to hereafter as the ‘fundamental’ corrugation and breathing modes. All higher indices are referred to as ‘overtones’ indicating (in general) an increasing number of vertical nodes in the basic structure function. The body modes are

¹ Technically speaking one could satisfy (37) if u' were much smaller than $\mathcal{O}(w' H_0/R_0)$ but that would imply disturbances with no perturbed vertical momenta since the remaining terms could be combined into a single expression $\partial \bar{\rho} w' / \partial z = 0$, where $\bar{\rho}$ is the steady-state density profile. We revisit this in the subsequent subsections.

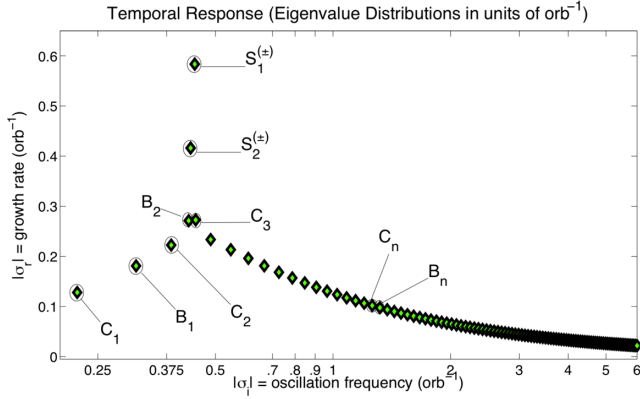


Figure 18. Distribution of growth rates $\sigma(k) = \pm(|\sigma_r| + \text{sgn}(k)|\sigma_i|)$ corresponding to wave solutions of equation (38), where $k = 1.74$ (corresponding to a disturbance of wavelength $\sim 0.009R_0$) and $q = -1$ with $H_0/R_0 = 0.05$. The growth rates and oscillation frequencies are given in units of local orbit times. The response is broadly grouped into (i) body modes constituting corrugation modes (odd symmetry with respect to the disc mid-plane, labelled C_n) and breathing modes (even symmetry, labelled with B_n); (ii) surface modes given by $S_{1,2}^{(\pm)}$. The fundamental corrugation mode C_1 has the fastest growth rate of $\sim 0.12 \text{ orbit}^{-1}$, which corresponds to an energy growth rate of $\sim 0.24 \text{ orbit}^{-1}$. The surface modes appear to correspond to the instability structures appearing in the beginning of the simulations.

classical inertial-gravity mode responses of an atmosphere. The other response (discovered rather unexpectedly) is the presence of a pair what we call *surface modes* (equivalent to the finger modes discussed earlier) which are labelled $S_{1,2}^{(\pm)}$. Unlike the body modes, the growth rates of a given pair of surface modes are not distinct: in other words, the growth rate σ associated with $S_1^{(+)}$ is the same as $S_1^{(-)}$ (but different from both $S_2^{(\pm)}$) although the structure of their eigenfunctions is different. This is why the surface mode eigenvalues are plotted exactly on top of one another in Fig. 18.

In Fig. 18, we depict the response of the system assuming a horizontal wavelength similar to the wavelength response seen in our simulations $\sim 0.009R_0$ for $H_0/R_0 = 0.05$ and $q = -1$ (in our system of study, this wavelength corresponds to a wavenumber $k \sim 1.74$). We note immediately that the growth rate of the fundamental corrugation mode, the mode that is dominant at late times in our simulations, is given to be $|\sigma_r| \sim 0.12 \text{ orbit}^{-1}$. If instead the growth rate is measured in terms of the kinetic energy, then the fundamental mode's growth rate would be $=2|\sigma_r| \sim 0.24 \text{ orbit}^{-1}$, which compares closely to what we measure in our highly perturbed simulations discussed in Section 5.1 where the corrugation mode is seen to quickly dominate the dynamics. We also note, however, that the growth rates in Fig. 18 indicate that the fundamental breathing mode has a growth rate (for the energy) of $\sim 0.4 \text{ orbit}^{-1}$ and the first overtone breathing mode has a growth rate of $\sim 0.6 \text{ orbit}^{-1}$. As discussed in Section 5.5, where models with smaller initial velocity perturbations were presented, the simulations seem to concur with the expectation that B_2 modes will appear at early times, followed by the B_1 modes and then the C_1 mode. One puzzle that we have not addressed in this paper is the fact that many higher order modes are also predicted to be unstable by the analysis, and these modes are not obviously observed in the simulations. Furthermore, low-order overtones of the corrugation modes are also not easily picked out by eye when plotting contours of v_z . Given the approximations inherent in our analysis, we will present a more complete exploration of these issues in a future paper where we consider solutions to the coupled radial and vertical eigenvalue problem that do not assume a

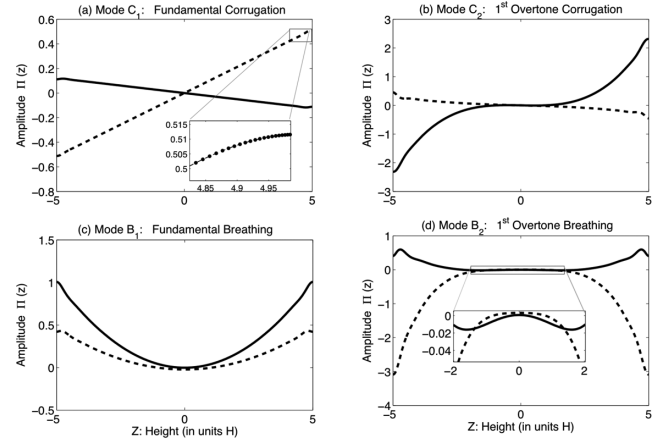


Figure 19. Corresponding depiction of body modes for the problem with the same parameters described in Fig. 18. The solid and dashed lines represent real and imaginary parts, respectively. Panel (a): the fundamental corrugation mode, (b) the first overtone corrugation mode, (c) the fundamental breathing mode and (d) the first overtone breathing mode. In panel (a), we show in the inset the distribution of the grid points as one approaches the domain boundary. The inset found in panel (d) depicts how close to zero the eigenfunctions get near the mid-plane ($z = 0$). All eigenfunctions drawn are normalized so that $\int_{-5}^5 |\Pi|^2 dz = 1$.

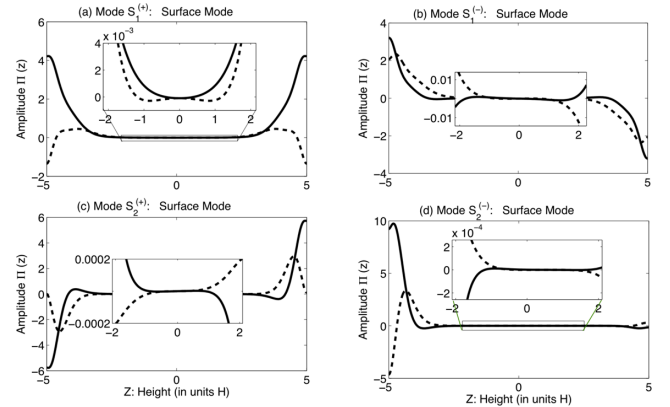


Figure 20. Same as Fig. 19 except the surface modes are depicted. Panel (a) shows $S_1^{(+)}$, (b) $S_1^{(-)}$, (c) $S_2^{(+)}$ and (d) $S_2^{(-)}$. The insets are provided to illustrate the extreme scale disparities exhibited by these functions, especially when comparing the amplitudes near the boundaries to those within two scaleheights of the mid-plane. Note the marked lack of mid-plane symmetry associated with mode $S_2^{(-)}$.

radial wave ansatz. Portraits of the eigenfunctions that correspond to the eigenvalues C_1, C_2, B_1 and B_2 in Fig. 18 are displayed in Fig. 19, where the solid lines depict the real parts of the eigenfunctions and the dashed lines show the imaginary parts.

Compared to all of the body modes, the two pairs of surface modes have the fastest growth rates of the collection. They grow at nearly four to six times the rate of the fundamental corrugation mode ($S_1^{(\pm)} \leftrightarrow |\sigma_r| \sim 0.58$ while $S_2^{(\pm)} \leftrightarrow |\sigma_r| \sim 0.42$). Moreover, Fig. 20 displays how, indeed, the power associated with these modes is concentrated near the boundaries. It is interesting to note that all the surface modes show clear expression of strictly even or odd symmetry with respect to the mid-plane except for $S_2^{(-)}$ which has power strongly concentrated on the top boundary. (The corresponding $S_2^{(-)}$ mode for $k = -1.74$ shows power concentrated at the bottom boundary, not shown in the figure.) This particular mode

is strongly asymmetrical. In all cases, the power of the surface modes within the inner two scaleheights of the disc is considerably diminished compared to the surface amplitude by anywhere from three to five orders of magnitude. Qualitatively speaking, these surface modes appear to be expressions of the fast growing boundary disturbances we witness during the outset of the numerical simulations. Although we plot the mode amplitudes in terms of the density/pressure perturbation $\Pi(z)$, we note that the perturbed vertical velocity is related to this quantity through equation (A17), which demonstrates that the velocity perturbation will similarly be confined near the disc surfaces.

We conclude this section with some cautionary comments. (1) These calculations were done assuming radial wave-like solutions and any true calculation will require a proper analysis of the fundamental partial differential equation (38) wherein the simple Fourier wave-mode assumption is replaced with a bona fide treatment in the radial direction including proper implementation of boundary conditions. (2) The asymptotically reduced equations are likely to become invalid once k becomes too large or too small. Similarly, a breakdown in validity is expected if the node number significantly increases. (3) What is important is to develop a theoretical construction in order to see and understand if the dependence of the growth rates can be tracked reliably as a function of the horizontal wavenumber. (4) We do not yet understand the origins of the surface modes nor have we developed a physically transparent mechanism explaining them. (5) Although the GSF instability requires $k_z/k_R \leq H_0/R_0$, we notice that for fixed k_R Fig. 18 shows the existence of a large number of unstable modes as the number of vertical nodes (and k_z) increases, and this issue needs to be explored using a more accurate calculation of the eigenfunctions and their associated growth rates. (5) Lastly, it is not yet clear why the fundamental corrugation mode appears to be the preferred vehicle of the growing instability at late times before saturation. This short list of outstanding issues comprises those that need addressing and will be considered in our follow-up studies.

7 DISCUSSION AND CONCLUSION

7.1 Application to protoplanetary discs

As our non-linear simulations have shown, the vertical shear instability is most likely to arise in discs that do not support strong levels of internal angular momentum transport. It is precisely the ability of fluid elements to travel small distances in the disc without changing their angular momenta that enables the vertical shear instability described originally by GS67 to operate by tapping into the angular momentum gradient generated by the combination of radial and vertical shear. We suggest that the most likely setting for the vertical shear instability is therefore in the magnetically decoupled dead zones in protoplanetary discs.

Before discussing the thermal time-scales expected in protoplanetary discs, we comment on the theoretically expected and observationally inferred temperature profiles in these discs. The passively irradiated disc model computed by Chiang & Goldreich (1997) leads to a radial temperature profile in the disc interior $T(R) = (R/\text{au})^{-3/7}$ 120 K, which lies in the range of radial temperature profiles that we have shown support the vertical shear instability. Submillimetre observations of circumstellar discs presented by Andrews & Williams (2005) indicate that the range of radial temperature power-law indices lies between -0.4 and -0.7 , which is again within the range of unstable profiles. Vertical temperature profiles are difficult to obtain from observations, but theoretical calculation of discs heated by their central stars has been presented by D'Alessio et al. (1998) and

Dullemond, van Zadelhoff & Natta (2002). An optically thick disc with grey opacities shows a characteristic two-component vertical temperature profile, with a warm surface layer and a cooler, vertically isothermal region near the mid-plane. A full radiative transfer calculation that adopts frequency-dependent opacities shows a moderate departure from this simple thermal structure, and the vertically isothermal inner region develops a modest temperature gradient, with temperature decreasing towards the mid-plane. A model calculated by Dullemond et al. (2002) for the outer disc at 17.2 au, for example, leads to a warm upper layer with $T \sim 85$ K and an interior region covering a few scaleheights where the temperature drops from $T \sim 40$ K to 20 K at the mid-plane. A positive thermal gradient implies that the vertical entropy profile will provide a stabilizing influence. Given our results on the thermal relaxation time required for the vertical shear instability to operate in vertically isothermal discs, this suggests that a short cooling time ≤ 0.01 local orbital periods is likely to be required for the vertical shear instability to operate under these conditions.

We now consider the radiative cooling times expected in the outer regions of protoplanetary discs. A similar discussion to the one that follows is found in Umurhan, Nelson & Gressel (2013). According to the standard theory of frequency-integrated radiative transfer with Rosseland mean opacities (Kippenhahn & Weigert 1990), the thermal time-scale due to radiative diffusion is

$$t_r = \Delta^2 (3C_p/4ac)(\rho^2 \kappa_R/T^3),$$

where κ_R is the Rosseland mean opacity and c , a and C_p are the speed of light, radiation constant and specific heat at constant pressure. Δ measures the perturbation length-scale of interest. Model values of T and ρ representing minimum mass solar models are drawn from the previously mentioned irradiated disc models of Chiang & Goldreich (1997), which compare favourably to the more detailed disc structure calculations of D'Alessio et al. (1998) and Dullemond et al. (2002). Using

$$\rho \approx \rho_{\text{mid}} = 2.7 \times 10^{-9} (\text{g/cm}^3) F \cdot (R/\text{au})^{-39/14}$$

and

$$T = (R/\text{au})^{-3/7} 120 \text{ K},$$

we find that the thermal time-scale is given approximately by

$$t_r/P_{\text{orb}} = 168 F^2 (\kappa_R/\text{cm}^2/\text{g}) (\Delta/R)^2 (R/20 \text{ au})^{-53/14},$$

$$P_{\text{orb}} \equiv 2\pi/\Omega_K, \quad (40)$$

where F represents the mass fraction of the protoplanetary disc relative to a minimum mass solar nebula model. Noting that the vertical wavelength of the instability $\lambda_z \sim H$ and that growing modes satisfy $k_R/k_z > R/H$, we adopt a radial length-scale $\Delta \equiv \lambda_R \approx H^2/R$. This gives $\Delta/R \approx 0.0025$. From the calculations of Henning & Stognienko (1996), opacities in the temperature range 50–150 K hover below $1\text{--}5 \text{ cm}^2 \text{ g}^{-1}$. Using the lower value as fiducial,² we find that at $R = 10 \text{ au}$, $t_r/P_{\text{orb}} = 0.015 F^2$ while at

² The frequency-averaged opacities calculated in Henning & Stognienko (1996) assume a grain mixture of amorphous silicates as typical of the interstellar medium. As part of the general paradigm of long-term disc evolution, the sizes of these grains grow quickly ($\sim 10^4$ yr) and achieve sizes of about $10^2\text{--}10^3 \mu\text{m}$ (Dullemond & Dominik 2005). With grain size growth comes a reduction of opacities (Pollack, McKay & Christofferson 1985). D'Alessio et al. (1998) argue that spectral energy distributions of T-Tauri stars support the presence of grains $\sim 3 \text{ mm}$. Thus, referencing figs 6 and 7 of Pollack et al. (1985) shows that grains with $300 \mu\text{m}$ radii correspond to Rosseland mean opacities of $0.2 \text{ cm}^2 \text{ g}^{-1}$ in the temperature range of interest; it is therefore fair to assume a fiducial upper bound for κ_R of $1 \text{ cm}^2 \text{ g}^{-1}$.

$R = 20 \text{ au}$, $t_r/P_{\text{orb}} = 0.001F^2$. To check that thermal time-scales based on the diffusion approximation are valid, we have calculated the optical depth in the radial direction across the mode length-scale assumed above and find that the optical depth is ~ 2 at 10 au. This justifies adoption of the diffusion approximation for radii less than or equal to this value. For larger disc radii, however, radial heat transport over length-scales characteristic of the unstable modes will be in the optically thin regime.

We conclude that for the perturbation length-scales on which the vertical shear instability is active, the thermal time-scales of minimum mass solar nebula discs are sufficiently short to accommodate the process, provided one is investigating disc dynamics out beyond $\sim 10 \text{ au}$ – a region that is expected to be within the dead zone. Furthermore, although the dead zone may transition to a region of active MRI turbulence further out in the disc beyond $\sim 20 \text{ au}$ (Armitage 2011), we note that there may also be a region towards the outer edge of the disc beyond $\gtrsim 50 \text{ au}$ where thermal time-scales are short and MRI turbulence is quenched by ambipolar diffusion. This may also be a region where the vertical shear instability can operate. On long evolution time-scales, the growth of dust grains reduces opacity and the disc surface density decreases due to accretion on to the star. Both of these processes lead to the vertical shear instability being more likely to manifest itself, provided that fully developed MRI turbulence does not arise in the disc because of these very same effects. Given the inherent uncertainties contained in the estimates derived above, the issues discussed here ultimately need to be investigated with more sophisticated disc models that include radiation transport, magnetic fields and a self-consistent ionization model.

7.2 Concluding remarks

We have presented results from a large suite of 2D axisymmetric and 3D simulations, using two independent hydrodynamic codes, that examine the stability and dynamics of accretion disc models in which the temperature or entropy is a strict function of the cylindrical radius. Such thermal profiles lead to equilibrium angular velocity profiles that depend on radius and height because of the baroclinic nature of the flow. We find that these disc models are unstable to the growth of two distinct types of modes. The first are seen to grow rapidly at high latitudes in the disc before descending down towards the mid-plane, and they have wavenumber ratio satisfying $|k_R/k_Z| \gg 1$. The second set of growing modes arise over slightly longer time-scales and occupy the main body of the disc, including the mid-plane regions. They manifest themselves early on as fundamental breathing modes, but as symmetry about the mid-plane is broken they become corrugation modes that cause the disc mid-planes to develop a pattern of vertical oscillations in the non-linear saturated state. These two mode types are only seen to grow when the thermal evolution time is comparable to or shorter than the dynamical time, and as such the conditions in the simulations match those that are required for operation of the GSF instability which predicts the occurrence of modes with $|k_R/k_Z| \gg 1$. The growth and dominance of corrugation modes, however, is a new result that has not been described in the literature previously.

The potential for the GSF (or vertical shear) instability to operate in accretion discs has been investigated previously by Urpin & Brandenburg (1998), Urpin (2003) and Arlt & Urpin (2004) using the Boussinesq approximation. We have confirmed this with our own analysis that applies to fully compressible flows for a locally isothermal equation of state. We have also extended the analysis by

introducing a number of assumptions motivated by the simulations, where we also relax the purely local assumption adopted previously in favour of one that accounts for the global nature of perturbations in the vertical direction. This simplified analysis leads to the prediction of two types of growing modes, one concentrated at high latitudes near the disc surface and the other occurring throughout the body of the disc, in agreement with the non-linear simulations. The growth rates derived from this analysis also match the simulations with good accuracy.

The non-linear saturated state in the 2D axisymmetric simulations arises from the interaction between the two types of growing modes. It appears to consist of a rather complex hydrodynamic flow that combines turbulence with a clear pattern of alternating vertical oscillations. A full 3D simulation computed using a locally isothermal equation of state indicates that the instability generates a turbulent flow that exhibits non-negligible transport of angular momentum through a Reynolds stress with corresponding viscous alpha value $\alpha \sim 10^{-3}$.

We have shown that the instability is damped in viscous flows with stress parameter $\alpha \geq 4 \times 10^{-4}$, so disc regions where the instability operates need to be stable against the MRI (the instability is not observed in global simulations of discs supporting fully developed MHD turbulence). Combining this constraint with the requirement for rapid thermal relaxation times suggests that the instability may operate in the outer regions of protoplanetary discs within the dead zone. A simple analysis based on thermal diffusion times across the characteristic length-scales of unstable GSF modes suggests that instability may arise in the dead zone at radii exterior to 10 au in discs with masses and surface densities similar to minimum mass solar nebula models. It is also likely that the MRI is quenched by ambipolar diffusion in the very outermost regions of protostellar discs (Armitage 2011), so these may also provide low-density, magnetically inactive regions where the vertical shear instability can operate.

The vertical shear instability is an example of a baroclinic instability because the required vertical shear in a disc only occurs if the pressure is a function of both density and temperature, $P(\rho, T)$. An important question that we have not addressed in this work is how a disc would evolve that is subject to both the vertical shear instability and the SBI studied by Petersen et al. (2007) and Lesur & Papaloizou (2010). The conditions required for the SBI to operate and be sustained strongly are a radial entropy gradient and a fairly rapid thermal relaxation time, and we have shown that the vertical shear instability operates under these conditions also. Both Petersen et al. (2007) and Lesur & Papaloizou (2010) report that a strongly sustained instability is obtained for thermal time-scales close to the local orbital period, leading to a highly complex flow in which long-lived vortices are formed, and an effective viscous stress of $\alpha \sim \text{few} \times 10^{-3}$ is maintained by the Reynolds stress in compressible flows. The 3D simulation we presented in Section 5.8 used a locally isothermal equation of state, and so is not subject to the SBI, but it seems likely that the combined action of the two instabilities in a disc with longer thermal evolution time and appropriate entropy gradient will generate a complex flow containing long-lived vortices and vertical motions that correspond to corrugation of the disc, accompanied by a Reynolds stress that leads to efficient outward angular momentum transfer. It is worth noting that the vertical shear instability is linear, whereas the SBI is a finite-amplitude instability, so it is possible that the SBI may be stimulated by perturbations generated during the development of the vertical shear instability. We will present a study of these two instabilities operating in tandem in a future publication to explore

these hypotheses. Given the role that vortices may play in the trapping of solids during planet formation (e.g. Barge & Sommeria 1995; Klahr & Bodenheimer 2003), and the influence of turbulence on the growth of dust (Dullemond & Dominik 2005) and planetesimals (Nelson & Gressel 2010), this is clearly an important issue for further investigation.

ACKNOWLEDGEMENTS

Part of this work used the NIRVANA-III code developed by Udo Ziegler at the Leibniz Institute for Astrophysics (AIP). Computations were performed on the QMUL HPC facility, purchased under the SRIF and CIF initiatives, and on the Complexity Cluster hosted by the University of Leicester on behalf of the DiRAC service. We acknowledge the hospitality of the Isaac Newton Institute for Mathematical Sciences, where much of the work presented in this paper was completed during the ‘Dynamics of Discs and Planets’ research programme. RPN acknowledges useful and informative discussions with Steve Lubow on the subject of corrugation modes in discs during an early phase of this project. We also acknowledge comments provided by the referee of an earlier version of this work that led to improvements in the paper. We acknowledge further comments received by a second referee, Jerome Guilet, that also improved the paper.

REFERENCES

- Andrews S. M., Williams J. P., 2005, *ApJ*, 631, 1134
 Arlt R., Urpin V., 2004, *A&A*, 426, 755
 Armitage P. J., 2011, *ARA&A*, 49, 195
 Balbus S. A., Hawley J. F., 1991, *ApJ*, 376, 214
 Barge P., Sommeria J., 1995, *A&A*, 295, L1
 Beckwith K., Armitage P. J., Simon J. B., 2011, *MNRAS*, 416, 361
 Cameron A. G. W., Pine M. R., 1973, *Icarus*, 18, 377
 Chiang E. I., Goldreich P., 1997, *ApJ*, 490, 368
 Cresswell P., Nelson R. P., 2006, *A&A*, 450, 833
 Cresswell P., Dirksen G., Kley W., Nelson R. P., 2007, *A&A*, 473, 329
 D’Alessio P., Canto J., Calvet N., Lizano S., 1998, *ApJ*, 500, 411
 D’Angelo G., Lubow S. H., 2010, *ApJ*, 724, 730
 Dullemond C. P., Dominik C., 2005, *A&A*, 434, 971
 Dullemond C. P., van Zadelhoff G. J., Natta A., 2002, *A&A*, 389, 464
 Fragner M. M., Nelson R. P., 2010, *A&A*, 511, A77
 Fricke K., 1968, *Z. Astrophys.*, 68, 317
 Fromang S., Nelson R. P., 2006, *A&A*, 457, 343
 Fromang S., Lyra W., Masset F., 2011, *A&A*, 534, A107
 Gardiner T. A., Stone J. M., 2008, *J. Comput. Phys.*, 227, 4123
 Goldreich P., Lynden-Bell D., 1965, *MNRAS*, 130, 125
 Goldreich P., Schubert G., 1967, *ApJ*, 150, 571 (GS67)
 Hawley J. F., Balbus S. A., 1991, *ApJ*, 376, 223
 Heinemann T., Papaloizou J. C. B., 2009, *MNRAS*, 397, 52
 Henning T., Stognienko R., 1996, *A&A*, 311, 291
 Johnson B. M., Gammie C. F., 2006, *ApJ*, 636, 63
 Kippenhahn R., Weigert A., 1990, *Stellar Structure and Evolution*. Springer-Verlag, Berlin
 Klahr H. H., Bodenheimer P., 2003, *ApJ*, 582, 869
 Kley W., D’Angelo G., Henning T., 2001, *ApJ*, 547, 457
 Lesur G., Papaloizou J. C. B., 2010, *A&A*, 513, A60
 Lin D. N. C., Papaloizou J., 1980, *MNRAS*, 191, 37
 Lin D. N. C., Pringle J. E., 1987, *MNRAS*, 225, 607
 Lubow S. H., Ogilvie G. I., 1998, *ApJ*, 504, 983
 Lubow S. H., Pringle J. E., 1993, *ApJ*, 409, 360
 Lynden-Bell D., Pringle J. E., 1974, *MNRAS*, 168, 603
 Nelson R. P., Gressel O., 2010, *MNRAS*, 409, 639
 Okazaki A. T., Kato S., Fukue J., 1987, *PASJ*, 39, 457
 Papaloizou J. C. B., Nelson R. P., 2003, *MNRAS*, 339, 983

- Papaloizou J. C. B., Pringle J. E., 1984, *MNRAS*, 208, 721
 Papaloizou J. C., Savonije G. J., 1991, *MNRAS*, 248, 353
 Petersen M. R., Stewart G. R., Julien K., 2007, *ApJ*, 658, 1252
 Pollack J. B., McKay C. P., Christofferson B. M., 1985, *Icarus*, 64, 471
 Ruden S. P., Papaloizou J. C. B., Lin D. N. C., 1988, *ApJ*, 329, 739
 Rüdiger G., Arlt R., Shalybkov D., 2002, *A&A*, 391, 781
 Ryu D., Goodman J., 1992, *ApJ*, 388, 438
 Shakura N. I., Sunyaev R. A., 1973, *A&A*, 24, 337
 Stone J. M., Norman M. L., 1992, *ApJS*, 80, 753
 Takeuchi T., Lin D. N. C., 2002, *ApJ*, 581, 1344
 Tassoul J.-L., 1978, *Theory of Rotating Stars*. Princeton Univ. Press, Princeton, NJ
 Teyssier R., 2002, *A&A*, 385, 337
 Toomre A., 1964, *ApJ*, 139, 1217
 Umurhan O. M., Regev O., 2004, *A&A*, 427, 855
 Umurhan O. M., Nelson R. P., Gressel O., 2013, *European Physical Journal Web of Conferences*, Vol. 46, Breathing Life Into Dead-Zones. EDP Sciences, Les Ulis, p. 3003
 Urpin V., 2003, *A&A*, 404, 397
 Urpin V., Brandenburg A., 1998, *MNRAS*, 294, 399
 van Leer B., 1977, *J. Comput. Phys.*, 23, 276
 Ziegler U., 2004, *J. Comput. Phys.*, 196, 393
 Ziegler U., 2011, *J. Comput. Phys.*, 230, 1035
 Ziegler U., Yorke H. W., 1997, *Comput. Phys. Commun.*, 101, 54
 Zingale M. et al., 2002, *ApJS*, 143, 539

APPENDIX A: EQUATIONS OF MOTION REVISITED AND STEADY STATES REDERIVED

The equations of motion for axisymmetric inviscid dynamics in a cylindrical geometry are given by

$$\left(\frac{\partial}{\partial t} + U\frac{\partial}{\partial R} + W\frac{\partial}{\partial Z}\right)U - \frac{V^2}{R} = -\frac{1}{\rho}\frac{\partial c_s^2 \rho}{\partial R} - \frac{\partial \Phi}{\partial R}, \quad (\text{A1})$$

$$\left(\frac{\partial}{\partial t} + U\frac{\partial}{\partial R} + W\frac{\partial}{\partial Z}\right)V + \frac{UV}{R} = 0, \quad (\text{A2})$$

$$\left(\frac{\partial}{\partial t} + U\frac{\partial}{\partial R} + W\frac{\partial}{\partial Z}\right)W = -\frac{1}{\rho}\frac{\partial c_s^2 \rho}{\partial Z} - \frac{\partial \Phi}{\partial Z}. \quad (\text{A3})$$

Note that the (R, ϕ, Z) velocity components are given here by (U, V, W) . We dispense with the subscripted scheme (v_r, v_ϕ, v_z) used in previous sections in order to simplify the notation. The corresponding equation of mass continuity is

$$\frac{\partial \rho}{\partial t} + \frac{1}{R}\frac{\partial R\rho U}{\partial R} + \frac{\partial \rho W}{\partial Z} = 0. \quad (\text{A4})$$

As mentioned above, we focus here on dynamics that are locally isothermal with an infinitely short cooling time ($\tau_{\text{relax}} \rightarrow 0$). This then is to be considered in the context of simulations T1R-0 to T4R-0 summarized in Table 1. Reciting therefore from Section 2, it means that the square of the sound speed is given by $c_s^2 = c_0^2(R/R_0)^q$, where R_0 is the fiducial reference disc position and c_0 is the scaled sound speed at that point. The gravitational potential emanating from the central object is $\Phi = -GM/(R^2 + Z^2)^{1/2}$.

The general equilibrium state solutions are found in equations (12) and (13) but, as we mentioned earlier, perturbations superposed on this base state are difficult to analyse because the resulting equations are fundamentally inseparable so that a typical normal-mode analysis is out of the question. In order to facilitate some kind of tractable analysis, we make the one and only approximation here: the radial and vertical gradients of the potential Φ are

expressed in terms of their corresponding first-order Taylor Series expansions, i.e.

$$\frac{\partial \Phi}{\partial R} \approx -\frac{GM}{R_0^2} \left(\frac{R_0}{R}\right)^2 = -\Omega_0^2 R_0 \left(\frac{R_0}{R}\right)^2,$$

$$\frac{\partial \Phi}{\partial Z} \approx -\frac{GM}{R_0^3} \left(\frac{R_0}{R}\right)^3 Z = -\Omega_0^2 \left(\frac{R_0}{R}\right)^3 Z,$$

in which $\Omega_0 = (GM/R_0^3)^{1/2}$ is the reference Keplerian rotation rate at radius R_0 . The mid-plane density is chosen to be of the form $\rho_{\text{mid}} = \rho_0 (R/R_0)^p$, where ρ_0 is the reference density and p is an arbitrary index (as referenced earlier). In the following analysis, it will be convenient for our discussion to refer to the natural logarithm of the density instead of directly to the density itself; thus, we define $\Pi \equiv \ln \rho$. Because we shall be concerned with perturbations around the steady states implied by the above equations, we shall represent these states by overbars. As such, we have that

$$\bar{\Pi} = \ln \rho_{\text{mid}} - \frac{1}{2} \frac{Z^2}{H_0^2} \left(\frac{1}{\mathcal{H}^2}\right), \quad (\text{A5})$$

$$\bar{V} = \Omega_0 R_0 \left(\frac{R}{R_0}\right)^{-3/2} \left[1 + \varepsilon(R) + \frac{q}{4} \left(\frac{H_0}{R_0}\right)^2 \left(\frac{Z}{H_0}\right)^2 \left(\frac{R}{R_0}\right)^{-2} \right], \quad (\text{A6})$$

where $H_0 \equiv c_0/\Omega_0$ is the local vertical scaleheight as referenced near the end of Section 2. The non-dimensionalized scaleheight \mathcal{H} is accordingly given via the relationship $\mathcal{H}^2 \equiv (R/R_0)^{3+q}$ as found in Section 2. The non-dimensional quantity ε is given by

$$\begin{aligned} \varepsilon &\equiv \left(\frac{H_0}{R_0}\right)^2 \mathcal{C}^2 \left(\frac{R}{R_0}\right) \left[R_0 \frac{\partial \ln \rho_{\text{mid}}}{\partial R} + R_0 \frac{\partial \ln \mathcal{C}^2}{\partial R} \right] \\ &= (p+q) \left(\frac{H_0}{R_0}\right)^2 \left(\frac{R}{R_0}\right)^{1+q}, \end{aligned}$$

where $\mathcal{C}^2 \equiv (R/R_0)^q$.

A1 Linearized perturbations and non-dimensionalization

We introduce perturbations by writing for each dependent quantity

$$U \rightarrow u', \quad V \rightarrow \bar{V} + v', \quad W \rightarrow w', \quad \Pi \rightarrow \bar{\Pi} + \Pi',$$

and inserting these into the governing equations (A1)–(A4). Linearizing results in the expressions

$$\begin{aligned} \frac{\partial u'}{\partial t} - 2 \frac{\bar{V}}{R} v' &= -c_s^2 \frac{\partial \Pi'}{\partial R}, \\ \frac{\partial v'}{\partial t} + u' \frac{1}{R} \frac{\partial R \bar{V}}{\partial R} + w' \frac{\partial \bar{V}}{\partial Z} &= 0, \\ \frac{\partial w'}{\partial t} &= -c_s^2 \frac{\partial \Pi'}{\partial Z} \end{aligned} \quad (\text{A7})$$

and

$$\frac{\partial \Pi'}{\partial t} = -\frac{1}{R} \frac{\partial R u'}{\partial R} + \frac{\partial w'}{\partial Z} - u' \frac{\partial \bar{\Pi}}{\partial R} - w' \frac{\partial \bar{\Pi}}{\partial Z}. \quad (\text{A8})$$

It will now be made more transparent if we non-dimensionalize the above equations according to the quantities appearing. We see that a natural time unit is given by the Keplerian rotation time Ω_0^{-1} . The

radial and vertical length-scales are naturally scaled by R_0 and H_0 , respectively. Thus, we write for these quantities

$$t \rightarrow \Omega_0^{-1} t, \quad Z \rightarrow H_0 z, \quad R \rightarrow R_0 r, \quad (\text{A9})$$

where t, r, z are the corresponding non-dimensionalizations of the independent variables representing, respectively, time, radius and height. It is very important to note that the radial and vertical length-scales are disparate with respect to each other by a factor of H_0/R_0 . Because in all of our simulations this ratio is quite small (~ 0.05), we shall treat this ratio as one of our ‘small parameters’ and formally represent it by $\epsilon \equiv H_0/R_0$ [not to be confused with $\varepsilon(R)$ defined earlier]. This disparity must be kept in mind when the scalings invoked to recover the GSF instability are formally made in the next section.

Judging from the dynamics observed in the simulations, the structures appearing tend to be radially and vertically constrained. These spatial constraints (especially the radial confinement) indicate that perturbation velocities ought not to exceed the sound speed (at least initially). This is typical of the scalings frequently used to derive equations appropriate to the dynamics in a small box of a disc (Goldreich & Lynden-Bell 1965; Umurhan & Regev 2004) although we are not, technically, considering the dynamics on such small scales yet. In sum, therefore, we scale the dependent perturbation velocities by

$$u' \rightarrow c_0 u, \quad v' \rightarrow c_0 v, \quad w' \rightarrow c_0 w,$$

where, as before, u, v, w represent the corresponding non-dimensionalized component velocities in the radial, azimuthal and vertical directions. Therefore, the perturbation equations now take on the following more transparent appearance:

$$\frac{\partial u}{\partial t} = 2 \frac{\bar{v}}{r} v - \epsilon r^q \frac{\partial \Pi'}{\partial r}, \quad (\text{A10})$$

$$\frac{\partial v}{\partial t} = -u \frac{1}{r} \frac{\partial r \bar{v}}{\partial r} - \epsilon w \left(\frac{1}{2} q z r^{-7/2} \right) \quad (\text{A11})$$

$$\frac{\partial w}{\partial t} = -r^q \frac{\partial \Pi'}{\partial z}, \quad (\text{A12})$$

and

$$\frac{\partial \Pi'}{\partial t} = -\epsilon \left(\frac{u}{r} + \frac{\partial u}{\partial r} \right) - \epsilon \frac{\partial \bar{\Pi}}{\partial r} u - \frac{\partial w}{\partial z} - w \frac{\partial \bar{\Pi}}{\partial z}, \quad (\text{A13})$$

where the non-dimensionalization of the mean azimuthal flow \bar{V} is given in terms of the other redefined variables

$$\begin{aligned} \bar{\Pi} &\equiv \frac{p}{r} - z^2 \frac{1}{2 \mathcal{H}^2}, \\ \bar{v} &\equiv \frac{\bar{V}}{\Omega_0 R_0} = r^{-3/2} \left(1 + \varepsilon(r) + \frac{q}{4} \epsilon^2 z^2 r^{-2} \right), \end{aligned} \quad (\text{A14})$$

with $\varepsilon(r) = \epsilon^2 (p+q) r^{1+q}$ and $\mathcal{H}^2(r) = r^{3+q}$. We have written (A11) in a seemingly curious way: the last term on the RHS of that equation is the product of the vertical gradient of the mean azimuthal flow term, i.e. $-w \partial \bar{v} / \partial z$. We have chosen to write it out explicitly in order to bring to the fore the leading-order scaling that sits in front of it as it will affect how we proceed towards the reduced model (see the next section).

A2 Asymptotic scalings and resulting reduced equations

The linearized equations of motion (A10)–(A13) are, despite our efforts to simplify, still inseparable between the r and z variables.

In order to proceed asymptotically, we must make further scaling choices. These are guided by both the results of the numerical solutions and the discussion at the beginning of Section 6. At this stage, we shall list them once again.

(i) As established by GS67, Urpin (2003) and Arlt & Urpin (2004), growth rates scale as $\sim q\Omega_0 H_0/R_0$ which is, in our non-dimensionalized time units, $\sim \epsilon q$.

(ii) For growth rates slow on the time-scale of the local disc rotation, the emerging structures have radial dimensions (ℓ_r) considerably smaller than the corresponding vertical dimensions (ℓ_z). That is to say, for $\epsilon q \ll 1$ the scaling analysis of both GS67 and Urpin (2003) indicate that $\ell_r/\ell_z \ll 1$.

(iii) The numerical solutions also clearly indicate that during the growth of the instability, the radial velocity fluctuations are significantly smaller than the corresponding meridional velocity fluctuations (see Fig. 1).

In the following, we describe scalings of (A10)–(A13) that simplify them into a set which is both more transparent and more amenable to further analysis while retaining the essential physical processes involved in the instability. We assume that $\epsilon \ll 1$ and treat q as an order 1 quantity (although a more general analysis can be done without this a priori assumption achieving, in the end, much of the same results discussed hereafter). Furthermore, we consider the analysis around the fiducial radius $r = 1$. Since interest is in radial scales that are much smaller than the vertical scales and recalling that the r scales dimensionally represent physical length-scales that are *longer* than the dimensional vertical scales z by a factor of ϵ^{-1} ($= R_0/H_0 \sim 20$), we consider radial disturbances

$$r - 1 = \epsilon^2 x,$$

where x is of order 1. We leave the z scales untouched as these are the de facto reference scales of the analysis. Because the growth rates are long by a factor of ϵ^{-1} , we introduce a new long-time variable τ given by

$$t = \tau \epsilon^{-1}.$$

With this long time-scale assumed, we find that in order to bring about non-trivial pressure balancing with the inertial term in the vertical momentum equation, it must follow that the pressure fluctuations must relatively scale by ϵ as well. This can be confirmed by examining (A12) and noticing that for w order 1 and the time derivative scaling as order ϵ , that the only way balance occurs is if the pressure is correspondingly small by a factor of ϵ . This means introducing a new pressure fluctuation reflecting this scaling through

$$\Pi' = \epsilon \tilde{\Pi},$$

where $\tilde{\Pi}$ is the scaled pressure.

Finally as we have just intimated, in addition to the vertical velocity being of order 1 we assume that the azimuthal velocity fluctuations are also unscaled (i.e. remaining order 1) in accordance with our numerical observations. We note here that scaling v of order 1 is also consistent with the pressure scalings assumed because it

leads to a balance between the radial pressure gradient and the Coriolis term in (A10).

However, we suppose that the *radial velocities are small* in comparison to the other velocity components and we propose that its relative smallness is similar to the pressure field's scaling, i.e.

$$u = \epsilon \tilde{u},$$

where \tilde{u} is the correspondingly scaled radial velocity. Applying these scalings assumptions to equations (A10)–(A13) results in the following equations at lowest order,

$$0 = 2v - \frac{\partial \tilde{\Pi}}{\partial x}, \quad (\text{A15})$$

$$\frac{\partial v}{\partial \tau} = -\frac{1}{2}\tilde{u} - \frac{1}{2}qzw, \quad (\text{A16})$$

$$\frac{\partial w}{\partial \tau} = -\frac{\partial \tilde{\Pi}}{\partial z}, \quad (\text{A17})$$

$$0 = \frac{\partial \tilde{u}}{\partial x} + \frac{\partial w}{\partial z} - zw, \quad (\text{A18})$$

with corrections to the above equations appearing at order ϵ^2 . In this form, these reduced equations contain insight with regard to two very important physical implications. The first of these follows from the interpretation of equation (A15) which says that the dynamics of the instability occur under *radially geostrophic conditions*, that is to say, that the processes develop under conditions in which radial Coriolis effects balance radial pressure gradients. The second observation is that the linear dynamics are anelastic rather than incompressible in character. By this we mean to say the following: since on these radial/vertical length-scales the mean (scaled) density profile has the form $\bar{\rho} = e^{-z^2/2}$, equation (A18) may be equivalently written as

$$\frac{\partial \bar{\rho} \tilde{u}}{\partial x} + \frac{\partial \bar{\rho} w}{\partial z} = 0. \quad (\text{A19})$$

The fact that the dynamics here are not incompressible in the usual sense is perhaps less surprising given that vertical stratification is non-negligible under these spatial constraints – had we been interested in vertical scales that were equally as short as the radial scales, stratification would not figure prominently.

The other two equations describing the vertical and azimuthal momentum balances retain their inertial terms and are largely unaffected (directly) by these scalings.

Before analysing the solutions to these equations, it is important to keep in mind that the essential effect giving rise to the instability is present in the guise of the final term on the RHS of equation (A16). Additionally, in reflecting upon these equations, it should be kept in mind that \tilde{u} and $\tilde{\Pi}$ indicate real quantities that are intrinsically smaller (but not zero) compared to the other terms.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.