Astronomy & Astrophysics

Linear analysis of the vertical shear instability: outstanding issues and improved solutions
(Research Note)

O. M. Umurhan[1,2], R. P. Nelson[3], and O. Gressel[4]

[1] Space Sciences Division, NASA Ames Research Center, Moffett Field, CA 94035, USA
  e-mail: orkan.m.umurhan@nasa.gov
[2] SETI Institute, 189 Bernardo Way, Mountain View, CA 94043, USA
[3] Queen Mary University of London, School of Physics and Astronomy, 327 Mile End Road London, E1 4NS, UK
[4] Niels Bohr International Academy, The Niels Bohr Institute, Blegdamsvej 17, 2100 Copenhagen Ø, Denmark

ABSTRACT

Context. The vertical shear instability is one of several known mechanisms that are potentially active in the so-called dead zones of protoplanetary accretion disks. A recent analysis of the instability mechanism indicates that a subset of unstable modes shows unbounded growth – both as resolution is increased and when the nominal lid of the atmosphere is extended. This trend suggests that, possibly, the model system is ill-posed.
Aims. This research note both examines the energy content of these modes and questions the legitimacy of assuming separable solutions for a problem whose linear operator is fundamentally inseparable.
Methods. The reduced equations governing the instability are revisited and the generated solutions are examined using both the previously assumed separable forms and an improved non-separable solution form that is introduced in this paper.
Results. Reconsidering the solutions of the reduced equations by using the separable form shows that, while the low-order body modes have converged eigenvalues and eigenfunctions (for both variations in the model atmosphere's vertical boundaries and radial numerical resolution). It is also confirmed that the corresponding high-order body modes and the surface modes indeed show unbounded growth rates. The energy contained in both the higher order body modes and surface modes diminishes precipitously due to the disk's Gaussian density profile. Most of the energy of the instability is contained in the low-order modes. An inseparable solution form is introduced to filter out the inconsequential surface modes, leaving only body modes (both low- and high-order ones). The analysis predicts a fastest growing mode with a specific radial length scale. The growth rates associated with the fundamental corrugation and breathing modes match the growth and length scales observed in previous nonlinear studies of the instability.
Conclusions. Linear stability analysis of the vertical shear instability should be done assuming non-separable solutions, especially for settings involving boundaries in the radial direction. We also conclude that the surface modes are relatively inconsequential because of the little energy they contain, and are artifacts of imposing specific kinematic vertical boundary conditions in isothermals disk models.

Key words. protoplanetary disks – instabilities – turbulence – waves – methods: analytical

1. Introduction

The vertical shear instability (VSI; Urpin 2003; Urpin & Brandenburg 1998; Arlt & Urpin 2004; Nelson et al. 2013; McNally & Pessah 2015; Stoll & Kley 2015), also known as the Goldreich Schubert Fricke instability (Goldreich & Schubert 1967; Fricke 1968) is a linear instability of axisymmetric inertial modes that relies on the vertical shear of the basic near-Keplerian flow state. This instability may be active in non-magnetized parts of protoplanetary accretion disks and is perhaps discernible in their dead zones (Turner et al. 2014). Nelson et al. (2013, NGU13 hereafter) and Stoll & Kley (2014) demonstrate that the instability can generate a modest amount of turbulence, with effective disk α ranging somewhere between 4×10^{-4} and 10^{-3}. In both studies, the basic background setting is that of a locally isothermal disk with a radial temperature gradient arising either from some external imposition (NGU13) or appearing naturally, as a result of radiative transfer effects (Stoll & Kley 2014).

A satisfactory linear stability analysis is still missing for the disk setting. While the basic essence of the instability has been sketched out using a local point analysis (Goldreich & Schubert 1967; Fricke 1968; Urpin 2003), the way the instability manifests itself in a global or semi-global disk setting is difficult to assess because the basic linear stability problem is non-separable[1] even in the simplest model reduction (NGU13; Barker & Latter 2015, BL15 hereafter). NGU13 and BL15 present a similar linear stability analysis using a reduced model set, and they show that, while the low-order modes that go unstable are consistent with the time scales of the instability seen in the numerical experiments, there are some serious shortcomings associated with the analysis that cast doubt on whether it accurately describes the physical manifestation of the VSI, especially when analyzed in the locally isothermal disk setting.

---

[1] In the usual sense where a function $f(x, y)$ may or may not be written as a product of functions purely of x and purely of y.

The mode analyses done by NGU13 and BL15 show that if one assumes radially propagating traveling waves, there are two classes of modes loosely referred to as body modes and surface modes. The surface modes come into existence if one imposes no-flow boundary conditions like an impenetrable lid at positions above and below the disk midplane (usually at least a few local disk scale heights or higher). The body modes are present irrespective of the kinematic conditions in the vertical, so long as their kinetic energies decay far enough away from the midplane (see below). BL15 also point out that the surface modes mulitply as the radial disturbance wavelengths become shorter. BL15 and NGU13 show that, for a given value of the radial wavenumber, there is a mode with the fastest growth rate, which corresponds, generally, to a surface mode.

However, there are three troubling features:

1. As the nominal lid of the atmosphere is extended to infinity, the fastest growing eigenmodes have growth rates that become similarly unbounded, growing like $\sqrt{m}$ for integer m representing the number of vertical nodes in the disturbance (BL15).

2. Where no-flow boundary conditions are imposed in the vertical direction, the number of unstable surface modes (with increasingly finer length scales) increases with higher radial resolution, possibly suggesting that the fundamental problem in the model VSI setup itself could be ill-posed – at least with respect to these surface modes (BL15).

3. As the wavelength of the radially propagating traveling wave becomes larger, the growth rate similarly increases in an unbounded way.

With regards to the third feature, both numerical simulations of NGU13 and Stoll & Kley (2014) indicate that a radial scale of maximum linear growth exists, yet neither of the analyses of the asymptotically reduced equations examined by NGU13 or BL15 allow for such a trend. Is it possible that the reason for this is the breakdown in the validity of the reduced equations, which hinges on the assumption of radial geostrophy in the dynamics, or might this be a problem with the assumption of radial traveling waves? These questions are reviewed in more detail in Sect. 2.

In Sect. 3, we argue that the first two of the troubling features listed does not lead to any serious deficiency in either the reduced set of equations or in the robustness/validity of the VSI itself. This is because both the surface modes and the other high nodal modes (i.e. high m) carry very little of the total vertical kinetic energy of the system. For the third feature, we consider this pathology to be a shortcoming of adopting a traveling wave-like solution and not a deficiency of the reduced set of equations. This, in turn, is intimately related to incorrectly assuming separable solution forms for a problem that is inherently inseparable[2]. We present an improved approximation in Sect. 4, where we adopt a relatively tractable, non-separable solution form and reanalyze the reduced equations. We find that there are maximally growing disturbances at some finite radial length scale and that they, in turn, match the growth rates and fastest growing radial scales reported in NGU13. In Sect. 5 we briefly discuss our findings.

---

[2] Another way of viewing this is to say that adopting boundary conditions that permit simple traveling waves solutions are probably not physically relevant to a disk.

2. Background

In both NGU13 and BL15, the following asymptotic reduced set of equations governing the dynamics of the VSI was shown to be appropriate for describing the linear development of the instability of axisymmetric disturbances:

$$0 = 2\Omega_0 v - \frac{\partial \tilde{\Pi}}{\partial x}, \tag{1}$$

$$\frac{\partial v}{\partial \tau} = -\frac{1}{2}\Omega_0 \tilde{u} - \frac{\Omega_0}{2} qzw, \tag{2}$$

$$\frac{\partial w}{\partial \tau} = -\frac{\partial \tilde{\Pi}}{\partial z}, \tag{3}$$

$$0 = \frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} - zw, \tag{4}$$

with Ω_0 the local rotation rate of the disk section at a distance R_0 from the parent star, the above set was obtained assuming that the spatial and temporal scales of motion are related to one another according to the following: temporal dynamics are given by $O(1/\epsilon\Omega_0)$, radial dynamic scales x are $O(\epsilon^2 R_0)$, and the vertical scales z are on the scale height $H_0 = O(\epsilon R_0)$, in which the small parameter $\epsilon \equiv H_0/R_0$ measures the relative thinness of the disk (which is usually taken to be approximately 0.05 in most theoretical studies, including the ones cited above). The scaled radial and vertical velocities are u and w, respectively, while v is the deviation azimuthal velocity with respect to the background near-Keplerian flow, and $\tilde{\Pi}$ is the scaled pressure perturbation. These equations model disk inertial modes with very short radial wavelengths. The degree of the vertical shear, which varies with disk height z, is controlled by the parameter q (where no vertical shear is equivalent to $q = 0$). In both NGU13 and Stoll & Kley (2014), the value of $q = -1$ is adopted. Equation (1) states that the disturbances are largely in radial geostrophic balance. Equations (2), (3) are the azimuthal and vertical momentum equations while Eq. (4) is the anelastic equation of state. See both NGU13 and BL15 for further details regarding the derivation of this set of reduced equations. Because the equations have been appropriately non-dimensionalized, we henceforth set $\Omega_0 = 1$ on all of the Coriolis and vertical shear terms appearing in Eqs. (1), (2).

This simplified model may then be combined into a single partial differential equation (PDE) for the scaled pressure perturbation

$$-\frac{\partial^2}{\partial t^2}\frac{\partial^2 \tilde{\Pi}}{\partial x^2} + \frac{\partial^2 \tilde{\Pi}}{\partial z^2} + \left(1 + q\frac{\partial}{\partial x}\right)z\frac{\partial \tilde{\Pi}}{\partial z} = 0. \tag{5}$$

Assuming normal mode solutions $\tilde{\Pi} = \hat{\Pi}(x,z)\,e^{-i\omega t} + \text{c.c.}$ transforms the above PDE into the simpler one:

$$\omega^2 \frac{\partial^2 \tilde{\Pi}}{\partial x^2} + \frac{\partial^2 \tilde{\Pi}}{\partial z^2} + \left(1 + q\frac{\partial}{\partial x}\right)z\frac{\partial \tilde{\Pi}}{\partial z} = 0. \tag{6}$$

The task still remains to construct solutions to this system and determine the eigenvalue ω that, in turn, determines the temporal response. Inspection of the above form shows that this system is inseparable when $q \neq 0$, which introduces a number of problems with regards to linear stability analyses that we discuss briefly below.

One might continue to analyze this system assuming an approximate separable form[3]. For instance, as was done by BL15, one can consider the relatively tractable traveling-wave ansatz as

$$\hat{\Pi} = Z(z) e^{ikx}. \tag{7}$$

Equation (6) can then be simplified further to

$$-k^2 \omega^2 Z + \frac{\partial^2 Z}{\partial z^2} + (1 + iqk) z \frac{\partial Z}{\partial z} = 0, \tag{8}$$

where $Z(z)$ is an, as yet undetermined, vertical structure function. The above equation, which is explicitly the same form as appear in BL15, is in the form of Hermite's equation. One solution of this system that admits tractable analytic results is to allow perturbations to show (at most) algebraic growth as $|z| \to \infty$, in which case

$$Z(z) = He_m(z) \tag{9}$$

is an acceptable solution, where m is a positive index and He_m is the Hermite polynomial of order m. The index m counts the number of vertical nodes in the pressure eigenfunction, and is thus referred to as indicating this quality throughout the rest of this Research Note. When inserted into Eq. (6), we find that this solution form is an actual solution, provided the following relationship holds:

$$\omega = \frac{\sqrt{m}}{k} \sqrt{1 + ikq}. \tag{10}$$

As discussed in BL15, the growth rate associated with this form increases without bound as $k \to 0$. We agree that this is pathological for a number of reasons, the following two being most prominent: (i) the approximation of radial geostrophy most likely breaks down when the horizontal wavelengths become large; and (ii) because, as we show in the next section, the traveling wave ansatz is also deeply flawed. This solution is also problematic because the growth rates increase without bound, both as the integer m increases and as k decreases. Nonetheless, this simplest solution indicates that the system is ill-posed as higher-order vertical modes (increased m) are included in the analysis.

We note that this solution ansatz cannot recover surface modes for the obvious reason that no kinematic boundary conditions are either imposed or enforced in the vertical. We also observe the following: that since both the atmosphere density drops off as a Gaussian (i.e. $\rho_0 \sim \exp -z^2/2$) and that eigenmodes have z^m structure, the effect of adopting a free vertical boundary condition is to imply that kinetic energies of all modes decay to zero as $z \to \pm\infty$.

Another approach is the one taken by NGU13 and BL15 in which Eq. (8) is solved with no-normal-flow boundary conditions at the vertical boundaries at $z = \pm H$ (we note that H refers to the height of the solution domain and H_0 refers to the disk

---

[3] If the calculation domain is on a uniform rectangular grid, a sure way to guarantee an unambiguous determination of eigenvalues and eigenmodes is to apply a finite difference discretisation of Eq. (6). With N_z and N_x discretisation points in the vertical direction and radial direction, the resulting eigenvalue system requires inversion of $(N_x N_z) \times (N_x N_z)$ sized (relatively) sparse matrices. High resolution in both the radial and vertical direction are desirable and, therefore, $N_z \approx 150$, $N_x \approx 100$ are preferred resolutions, which means constructing matrices that are prohibitively large to invert and challenging to reconfigure into known sparse matrix formulations.
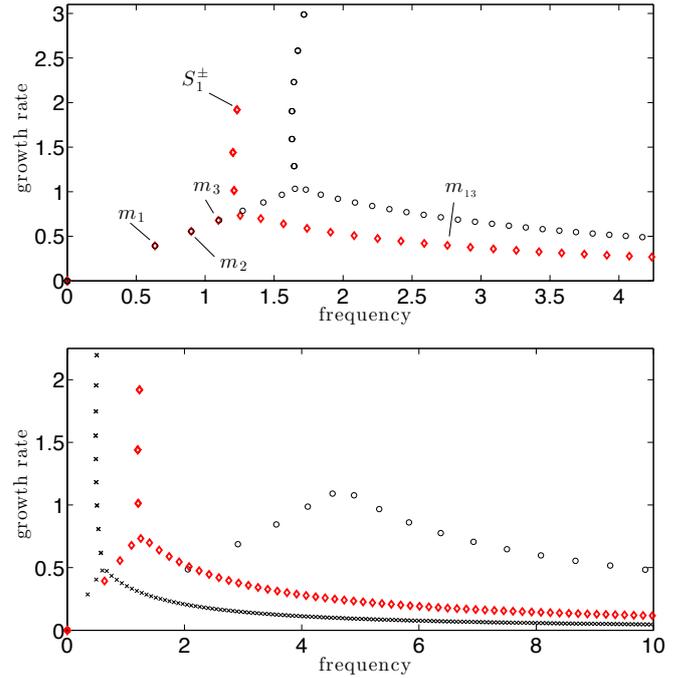

Fig. 1. Growth rates, $\mathrm{Im}(\omega)$, and frequencies, $\mathrm{Re}(\omega)$, of solutions of Eq. (8) subject to no-normal-flow boundary conditions at $z = \pm H$, where H is in units of scale heights. *Top panel*: distribution shown for $k = 2$, and $H = 5$ (diamonds), and $H = 7$ (open circles). As H increases, more surface modes become activated and high-order body modes have increased growth rates. In both cases shown, the frequency and growth rates of low-order body modes (labeled m_1, m_2, m_3) remain unchanged. The surface modes generally appear in pairs as indicated by labeling the topmost surface mode with the superscript $S_1^\pm$. This panel confirms the trends reported by BL15. *Bottom panel*: distribution of the complex frequencies shown for differing values of k with fixed $H = 5$: $k = 5$ (crosses), $k = 2$ (diamonds), $k = 0.5$ (open circles). The growth rates increase without bound as k is decreased, with the same trend identified in the problem with no vertical boundaries as found in Expression (10). As k is increased, the number of surface modes increases, including the maximum growth rates which also confirms the results reported in BL15.

scale height in this paper), which amounts to stating that $\partial_z \tilde{\Pi} = 0$ at $z = \pm H$ provided $\omega \neq 0$, which follows from the normal mode form of Eq. (3). The general solution of (8) is given by

$$Z(z) = A \, He_\lambda(z) + B_\lambda \, {}_1F_1\left(-\frac{\lambda}{2}, \frac{1}{2}, \frac{z^2(1 + iqk)}{2}\right)$$

where ${}_1F_1$ is a confluent hypergeometric function of the first kind and where λ is the usual separation constant which is determined through the application of boundary conditions. The second of these special functions does not offer very much in the form of analytical insight (NGU13) except in the guise of certain asymptotic limits (BL15), and as such, it is more convenient to numerically solve Eq. (8) directly.

The numerical eigenvalues determined by this procedure recover the surface modes as well as the body modes of the VSI. The low-order body modes (the fundamental and first overtone breathing and corrugation modes) are also recovered with eigenvalues consistent with the numerical results of NGU13. However, this system introduces apparent pathologies, which are shown in Fig. 1. There are generally three branches of solutions: one associated with low-order body modes with relatively low frequencies, another branch of body modes with

higher frequencies and a third short branch consisting of surface modes. The high-frequency branch of body modes shows decreasing growth rates as the mode frequency increases, while the low-order body modes show an increasing growth rate with increasing frequency.

However, as BL15 demonstrate, when the location of the vertical height is increased, the following occurs: (a) The growth rates of the high-frequency branch increases; (b) the growth rates of the surface body modes also increase; while (c) the growth rates and frequencies of the low-order body modes remain unchanged, and (d) as the lid of the atmosphere is raised to $\pm\infty$, the low- and high-order body modes line up with the frequencies and growth rates expressed in Eq. (10). This corresponds to the response predicted assuming the ansatz found in Eq. (7). While it would seem that imposing vertical no-flow boundary conditions removes the ill-posed nature of the model problem, the system's ill-posedness appears to manifest itself as $H \to \infty$, since the growth rates in the high-order body modes and surface modes correspondingly increase, once again seemingly without bound. This is problematic and suggests that the model problem may be ill-posed after all.

The situation deteriorates even further if we consider the behavior of the surface modes, as BL15 indicate: an increase in the radial resolution (larger k) causes the number of surface modes attached to the upper and lower no-flow boundaries to proliferate, as the second panel of Fig. 1 clearly illustrates. Raising the position of the lid in the model also increases the growth rate of the fastest-growing surface modes, which is also indicative of an ill-posed model.

Moreover, another serious shortcoming implied by the results of the linear stability solutions developed in NGU13 and BL15, is that they do not predict finite, non-zero maximally growing radial wave disturbances, something that is observed in the numerical experiments of NGU13 and Stoll & Kley (2014). This suggests that assuming wavelike modes in the radial direction is either flawed or, at the very least, not physically relevant.

3. Mode kinetic energies

In spite of these troublesome features, these approximate theoretical solutions reveal a lot about the physical nature of the developing instability, especially in relation to the high frequency body modes and the proliferating surface modes. For example, they provide information on the relative energy content for each mode. The low-order body modes carry most of the inertia of the disk disturbances since their amplitudes are greatest near $z = 0$ (NGU13, BL15). Because these are also locations where most of the disk mass is concentrated, the energy contained in these low-order body modes dominate the corresponding energy contained in the higher order body and surface modes. This has a direct consequence on the interpretation of the VSI, even in the context of this somewhat incomplete analysis.

To illustrate this quantitatively, Fig. 2 shows a comparison of the relative energy densities contained in representative modes, labeled in the top panel of Fig. 1, which corresponds to solutions of Eq. (8) with $H = 5$ and $k = 2$. From Eq. (3) it follows that each eigenmode $Z(z)$ generates a corresponding vertical velocity eigenfunction $w = i\omega^{-1}\partial_z Z$. We normalize each vertical velocity eigenfunction so that

$$\int_{-H}^{H} |w| dz = 1. \tag{11}$$

Since the reduced equations represent an isothermal atmosphere, the steady-state density is given by $\rho_0 = \exp(-z^2/2)$. We also
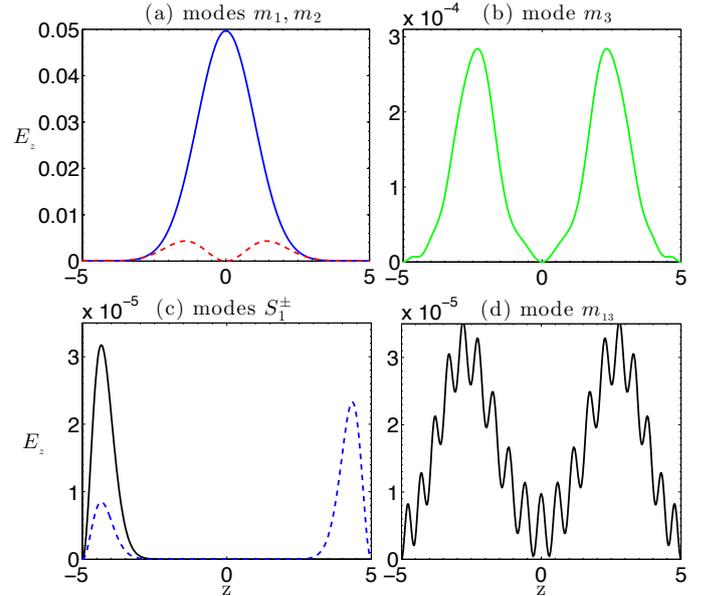

Fig. 2. Vertical kinetic energy density plots, E_z, plotted for the modes labeled in the top panel of Fig. 1, which corresponds to solutions of Eq. (8) with $H = 5$ and $k = 2$. Each corresponding vertical velocity eigenfunction w is normalized, such that $\int_{-5}^{5} |w| dz = 1$. Shown here: **a)** low-order body modes m_1, m_2 (solid and dashed lines respectively); **b)** low-order body mode m_3; **c)** fastest growing surface modes, S_1^+, S_1^- (solid and dashed lines respectively), and **d)** a high-order body mode, m_{13}. The lowest order body mode, m_1 is the fundamental corrugation mode that dominates the energy density, contained in the high-order body modes and the surface modes by, at least a factor of 10^3. The energy density contained in the fundamental breathing mode (m_2) is a factor 10 less than the mode m_1.

recall that this instability is one in which the perturbation vertical kinetic energy density dominates over the radial and azimuthal kinetic energy densities (NGU13, Stoll & Kley 2014). As such, we consider the vertical kinetic energy density E_z of each of the corresponding modes defined by $E_z(z) \equiv 0.5\rho_0|w|^2$. "$m_1$" labels the fundamental corrugation mode (FCM), while "m_2" labels the fundamental breathing mode (FBM) so that, for example, the expression $E_z(z, m_1)$ corresponds to the energy density of the FCM, and so on. For each mode we also define a total vertically integrated density $\overline{E}_z \equiv \int_H^H 0.5\rho_0|w|^2 dz$ (i.e., a surface energy density). When we refer to the surface energy density of a particular mode, we write, for example, $\overline{E}_z(m_1)$ to indicate the surface energy density of the FCM, and so on.

Figure 2 plots the energy densities $\overline{E}_z$ for the various modes admitted by the system with parameter values $k = 2, H = 5$, in which each vertical velocity normal mode is normalized according to Eq. (11). Here, we see that the relative energy density content is greatest with FCM and that it dominates the FBM by a factor of 10. The energy density distribution in the other higher order body modes are reduced by at least a factor of 100 compared to the FCM. The energy density contained in the two fastest growing surface modes is diminished by a factor of 1000 compared to the FCM.

A comparison of the total vertically integrated energies in these various modes emphasizes further the relative unimportance of the high-order body and surface modes. To express this quantity, relative to the vertically integrated energy density of the FCM ($\overline{E}_z(m_1)$), the following selected modes are used: for the FBM, $\overline{E}_z(m_2) \approx 0.11\overline{E}_z(m_1)$; for the first overtone corrugation
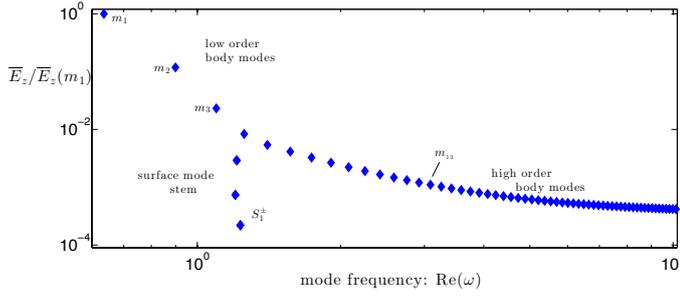
Fig. 3. Total vertically integrated vertical kinetic energies $\overline{E}_z$ as a function of mode frequencies, $\mathrm{Re}(\omega)$, which correspond to solutions of Eq. (8) with $H = 5$ and $k = 2$. $\overline{E}_z$ is shown, normalized to the corresponding vertically integrated energy of the fundamental corrugation mode, i.e. $\overline{E}_z(m_1)$. The relative weakness in the power of the surface modes located in the frequency windows 1.2 and 1.3. The labeled modes displayed in Fig. 2 are also labeled here.

mode, $\overline{E}_z(m_3) \approx 8.3 \times 10^{-3} \overline{E}_z(m_1)$; the selected high-order body mode m_{13}: $\overline{E}_z(m_{13}) \approx 1.3 \times 10^{-3} \overline{E}_z(m_1)$; and for the two surface body modes, both of which have the same amount of vertically integrated energy contained within, $\overline{E}_z(S_1^\pm) \approx 2.2 \times 10^{-4} \overline{E}_z(m_1)$. Together with the first 50 vertical eigenmodes, the resulting trends are plotted in Fig. 3, which clearly shows that the energy contained in the modes drops with increased values of m. We similarly plot the relative vertically integrated kinetic energy densities for models where $H = 7, k = 2$ (Fig. 4) and $H = 8, k = 5$ (Fig. 5), and we see how, by increasing the resolution (going to larger k) and by extending the atmosphere lid, the low-order modes become increasingly populated (as BL15 point out) and how the $\overline{E}_z$ energies, contained in the corresponding surface and high-order body modes, diminish even further.

Most importantly we confirm the trend, reported by BL15m in which the energy in the low-order body modes remains steady as H is increased. This is especially true for the FCM and FBM but also becomes a characteristic feature of increasing overtones as H increased. With reference to Fig. 3, all the body modes to the left of the triple junction, where the surface mode branch meets the low-order and high-order body modes, display energies that remain unchanged as H increases. Increasing H, however, moves the location of the triple junction toward higher order body modes, whereas the energies of the low-order body modes (i.e., those left of the triple junction) do not change with increased H.

The relatively weak energy carrying potential of both the high-order body modes and the branch of surface modes comes about because the corresponding kinetic energy associated with them is severely diminished as a result of the Guassian drop-off that is associated with the mean density field ρ_0. This is despite their relatively large velocity amplitudes for large values of $|z|$.

4. An improved approximate solution

While we cannot address all of the concerns listed in Sect. 2, we offer an improved solution ansatz to the eigenvalue problem posed by Eq. (6). The simulations presented in NGU13 and Stoll & Kley (2014) employ a numerical set-up in which radial boundaries are enforced. These boundaries introduce effects that alter the growth rates and character of the low-frequency body modes, which are the very ones that have been observed to carry the instability into the nonlinear regime. We demonstrate here that (a) the unbounded growth predicted by assuming wavelike
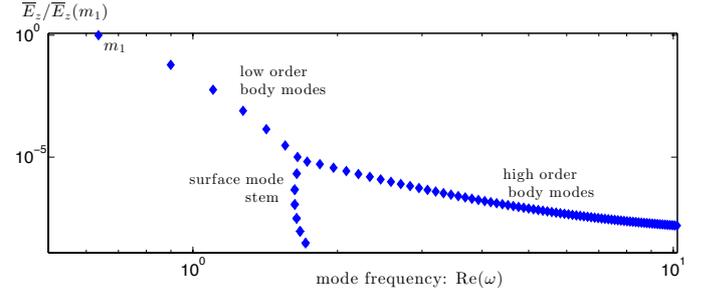

Fig. 4. As in Fig. 3 except with $H = 7$. The power contained both in the surface and high-order body modes is diminished as the atmosphere lid is set farther away. Note that the energy in the low-order body modes remain unchanged, especially the FCM and FBM.
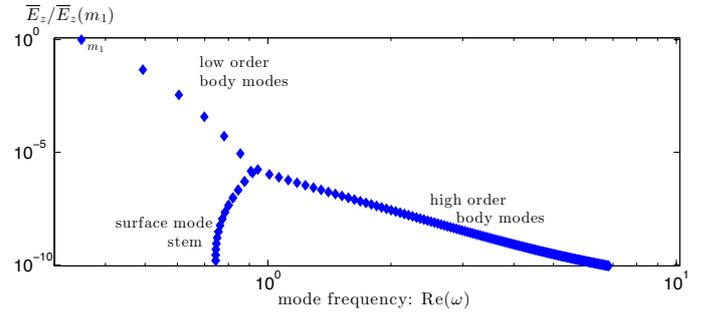

Fig. 5. As in Fig. 3, except that $H = 8$ and $k = 5$. The diminished energy-carrying capacity of both the surface modes and the high-order body modes when the lid of the atmosphere is made larger and when higher radial resolution modes are considered. The energy of the low-order body modes, especially the FCM, FBM and the first overtone corrugation mode remain unchanged compared to their energies for lower values of H, depicted in the two previous figures.

disturbances in the radial direction, is an artifact created by assuming radially-traveling wave solutions and (b) that this pathology is removed by the imposition of some kind of fixed radial boundary condition. Furthermore, the imposition of boundary conditions selects the fastest mode with radial length scales and growth rates that match those found in the aforementioned numerical experiments.

To represent the effect of radial boundaries, the following non-separable ansatz is assumed:

$$\hat{\Pi} = P_m(z, x) = \sum_{n=0}^{m} P_{n,m}(x) z^n m \qquad (12)$$

where m is a positive integer, and $P_{n,m}(x)$ a set of unknown functions of x[4]. This solution form is one borrowed from singular value decomposition methods, and has been used in other disk studies (e.g., Lubow & Pringle 1993). We note already that the ansatz found in Eq. (12) builds unbounded algebraic spatial growth into the solutions as $|z| \to \infty$, which will predict the same kind of unbounded growth in which $\mathrm{Im}(\omega) \sim \sqrt{m}$, just as the simple solution shown in Eq. (10). But, as we already note in Sect. 2, this means that these solutions are ones where the kinetic energies always decay as $z \to \pm\infty$. On a positive note, these solutions are not burdened by the introduction of surface modes.

---

[4] This ansatz form is non-separable for all integer values of $m \geq 2$, and it is separable only for the $m = 1$ mode.

Based on the above, we adopt this solution form and insert it into Eq. (6). Separating out like-orders of powers of z turns this system into $O\left(\frac{m+1}{2}\right)$ nested ODEs for the unknown functions $P_{n,m}$. For $n = m$ we have the so-called "top" equation

$$\omega^2 \frac{\partial^2 P_{m,m}}{\partial x^2} + m\left(P_{m,m} + q\frac{\partial P_{m,m}}{\partial x}\right) = 0, \tag{13}$$

while for $0 \le n < m$ we have the remaining so-called "slaved" equations

$$\omega^2 \frac{\partial^2 P_{n,m}}{\partial x^2} + n\left(P_{n,m} + q\frac{\partial P_{n,m}}{\partial x}\right) = -(n+2)(n+1)P_{n+2,m}. \tag{14}$$

The above system has even and odd symmetries associated with it, so that there are so-called breathing modes (even m) and corrugation modes (odd m). The fundamental corrugation mode (FCM) corresponds to $m = 1$ while the fundamental breathing mode (FBM) is associated with $m = 2$.

For this particular demonstration, we assume no-normal-flow boundary conditions at particular inner and outer radial positions, i.e., $u = 0$ at $x = \pm L_x$, where $L_x > 0$. In terms of the variables we use, the expression of this condition is found by rewriting Eq. (2) in terms of the normal mode ansatz

$$-\mathrm{i}\omega \frac{\partial \tilde{\Pi}}{\partial x} + \frac{qz}{\mathrm{i}\omega}\frac{\partial \tilde{\Pi}}{\partial z} = 0. \tag{15}$$

Given the solution form (12) and since the polynomials z^n are linearly independent with respect to one another (for integer n), each function $P_{n,m}$ must separately satisfy

$$\frac{\partial P_{n,m}}{\partial x} + \frac{qn}{\omega^2} P_{n,m} = 0 \quad (\text{at } x = \pm L_x). \tag{16}$$

The method for solving the full problem is now straightforward: (i) solve the "top" Eq. (13) subject to boundary conditions and then (ii) solve the "slaved" Eqs. (14) subject to the boundary conditions expressed in Eq. (16) for each $P_{n,m}$. This should be done by decreasing values of n (by 2) until one terminates either at $n = 0$ (breathing modes) or $n = 1$ (corrugation modes). A detailed depiction of the full solution will be the subject of a future study. Of concern to us here, however, is the fact that the top equation yields the eigenvalue ω. In fact, it is straightforward to show that

$$P_{m,m} = \left(A_j \sin k_j x + B_j \cos k_j x\right) \exp\left(-\frac{qm}{2\omega^2}x\right), \tag{17}$$

is a solution to Eq. (13) provided $\omega = \omega(k_j)$ satisfies

$$\frac{\omega^2}{m} = \frac{1 \pm \sqrt{1 - k_j^2 q^2}}{2k_j^2} \tag{18}$$

together with $k = k_j \equiv j\pi/2L_x$, where j is any integer including zero. When j is an odd integer, then $A_j = k_j$, $B_j = -qm/2\omega^2$. However, when j is an even integer (including zero) $A_j = -qm/2\omega^2$, $B_j = -k_j$.

Since all the k_j are real and their multitude are controlled by L_x, we can consider the set of k_j as part of a continuum of real values given as k, and we can analyze these results accordingly. It follows that unstable solutions exist only if $|kq| > 1$, and after a little algebra it implies that the growth/decay rate is given by

$$\mathrm{Im}(\omega) = \pm \sqrt{m}\frac{\sqrt{|kq| - 1}}{2k}. \tag{19}$$

This solution states that there is a wavelength of maximal growth $k_{\max}$ and corresponding growth rate $\sigma_{\max}$, which are given by

$$|k_{\max}| = \frac{2}{|q|}, \qquad \sigma_{\max} = \sqrt{m}\frac{|q|}{4}. \tag{20}$$

Restoring these results in terms of the physical scalings of the disk, this indicates a maximally growing wavelength $\Lambda_{\max}$ with corresponding growth rate $\Sigma_{\max}$ expressed as

$$\Lambda_{\max} = \pi\epsilon^2|q|R_0 = \pi|q|\frac{H_0^2}{R_0},$$
$$\Sigma_{\max} = \epsilon\sqrt{m}\frac{|q|}{4}\Omega_0 = \sqrt{m}\frac{|q|}{4}\frac{H_0}{R_0}\Omega_0, \tag{21}$$

of which the latter is given as $\Sigma_{\max} = 0.5\epsilon\sqrt{m}|q|\pi$ orb^{-1}, expressed in units of local disk orbit times (orb $= 2\pi/\Omega_0$).

5. Discussion

The analysis developed in Sect. 4 is an improvement over previous ones reported in the literature (namely NGU13 and BL15). Below we itemize some relevant observations regarding this analysis.

1. In the numerical experiments reported in NGU13, it is shown that in model disks where $\epsilon = 0.05$ and $q = -1$, the growth rate of the perturbation kinetic energy during the early phase (between 10 and 25 orbits of the inner disk) of the growing VSI is about 0.25 orb^{-1} (see righthand panel of Fig. 1 of NGU13). Closer inspection of the dynamical response during this phase (see Fig. 3 of NGU13) primarily shows a breathing mode character in the vertical velocity. The radial wavelength of the response near the left boundary indicates a size of about $0.009R_0$. (This corresponds to approximately 17 grid points, resolving the fastest growing radial mode.) According to the theory developed in the previous section, the radial scale and growth rate of the fundamental breathing mode (FBM, $m = 2$) is given by Eq. (21), which predicts $\Lambda_{\max} \approx 0.0079R_0$, together with $\Sigma_{max} \approx 0.11$ orb^{-1}. However, the growth rate in the kinetic energy is equal to $2\Sigma_{\max} \approx 0.22$ orb^{-1}. These predictions, based on this improved approximation, compare favorably with the results of the numerical simulations.

2. Similarly, for the same simulation reported in NGU13, after about 25 orbit times the simulations indicate slower growth, and the corresponding figures indicate that, in this latter phase, the disk response is primarily that of the FCM. For the FCM with k as given, the theory predicts a growth rate in the kinetic energy of about $2\Sigma_{\max}(m = 1) \approx 0.15$ orb^{-1}, which approximately matches the decelerated growth seen in the kinetic energy growth rates displayed in Fig. 1 of NGU13.

3. The analysis developed here, using the solution ansatz in Eq. (12), only captures the essence of the low-order body modes and cannot say anything about the surface modes. This is because no vertical boundary conditions are applied. However, based on our reflections in Sect. 3 regarding the kinetic energy density content of these modes, the surface

modes are probably ephemeral, with no significant dynamical effect upon the development of the VSI in the bulk interior, where most of the disk inertia is contained.

4. Despite the improved theoretical construction embodied in the ansatz of Eq. (12), especially in relation to the correct behavior predicted for a fastest growing radial mode, the theory still indicates unbounded growth as the vertical node number s*m* increases. Is this indicative of a profound flaw in the ansatz or is it a real effect? Based on our reflections upon the lack of energy contained in high-order body modes and surface modes (Sect. 3), it is possible that the theoretical predictions are actually valid, and that, despite the unbounded growth predicted for increasing *m*, the main instability and turbulent development in nonlinear calculations is driven primarily by the fundamental breathing and corrugation modes. The concomitant fast-growing high-order body modes and surface modes likely have little effect upon the overall development of the VSI primarily because they contain so little energy in comparison to the fundamental modes.

5. We have compared the energy contained in each mode assuming a white-noise spectrum in the initial velocity field. Because the density field decays like a Gaussian in the vertical, the relative energy content in modes with many nodes in the vertical is shown to be very small. Just as the system evolves into the turbulent regime, so too will the corresponding energy/velocity spectrum and the relative energy content contained in each mode. Given the behavior observed in the numerical simulations reported in the literature to date, whereby energy appears to drain into low-order (vertical) *m* modes through nonlinear processes, the assumption and discussion of the modal energy content taking on a white spectrum in the velocity field is a conservative assumption and gives us an upper bound as to the real energy content in a simulation of well-developed turbulence. Of course, the meaning and physical structure of the linear modes will also probably change into the nonlinear regime, since the basic state upon which they are constructed also changes as the system evolves into the nonlinear regime.

6. Related to the previous point, we believe that the VSI develops robustly and independently of the high-order and surface modes − even if artificial boundary conditions are emplaced on the upper and lower parts of a numerically modeled disk. This will not be true if, by conspiracy, most of the initial energy is placed in these same high-order modes and surface modes. This would be like stirring the low density parts of the disk with very large velocity perturbations, a scenario hard to imagine for a realistic disk. We agree with BL15 when they say that adding a bit of artificial viscosity, or possibly a sponge (an artifice commonly used in atmosphere GCMs) should either erase or strongly diminish these modes from a numerical calculation. We note that in both numerical experiments of NGU13 and Stoll & Kley (2014), the velocity fields are shown as timed snapshots of the instability as it develops, and they show a strong initial development in the velocity field near the boundaries that is primarily due to the fast growth rate of the surface modes. Nevertheless, the kinetic energy densities contained in disturbances high up in the atmosphere are diminished by a factor of e^{-12}, as a result of the vanishingly small densities up there. It is hard to imagine that the ensuing instability within the bulk of the disk is dependent upon these surface modes and we view them as inconsequential as far as the long-term development of

the VSI is concerned, including its aggregate turbulent transport. This can be verified by new numerical experiments in which, for instance, one replaces the upper hard boundary by a sponge or something similar.

7. In simulations, the VSI manifests itself as very short radial wavelength disturbances. BL15 have suggested that this trend may be interpreted as the radial scale of the instability is being controlled by the action of viscosity. If this is the case, then the length scale of the fastest growing mode ought to change when the actual physical viscosity is lowered or when the effective numerical viscosity is reduced. We ran a very high resolution simulation that replicated the results displayed in Fig. 11 of NGU13, but with three times as many grid points in the radial direction (all other quantities remained fixed). These simulations were controlled by numerical viscosity alone and, because of this, an increased resolution effectively lowers the associated numerical dissipation. Essentially, we find no difference between these high resolution runs and the aforementioned results in Fig. 11 of NGU13. Furthermore, we confirm that during the early phases of the linear instability, the FBM gets expressed first, as predicted by theory.

8. The results of the previous section show that enforcing boundary conditions at the inner and outer radial positions controls the growth of the VSI and selects for a fastest growing radial structure. As such, we see the unbounded growth rates resulting from the radial traveling wave ansatz (Sect. 2) as the result of the ansatz itself, which (we contend) is inappropriate to describe the dynamics in simulations with boundaries, and is not a result of a breakdown in the validity of the assumptions underpinning the equation set (1) to (4), namely the approximation of radial geostrophy of Eq. (1).

This last point naturally begets the following question: what are the appropriate boundary conditions to use in such disk models? How robust is the VSI to different radial boundary conditions, and how do these vouch for the true conditions of protoplanetary disk dead zones?

As observed by Urpin (2003) and Arlt & Urpin (2004), the local growth rates of the VSI scale approximately by $\epsilon\Omega_0$. While it has not yet been theoretically established to what degree the VSI is active in disk regions that also support the MRI, it is worthwhile noting that the growth rate of the VSI (when active in the absence of the MRI) is a factor of ϵ slower than the MRI underscoring the differing timescales of the two instabilities. Furthermore, there is also a disparity in the length scales on which the two instabilities operate, since the MRI is dynamically more potent on vertical disturbances with length scales that are comparable to the scale height H_0, together with very little if no radial structure, whereas the VSI requires very short radial scales, $\sim\epsilon H_0$. It is quite conceivable that the two may operate concurrently in a given disk section, but so far there are no theoretical studies illustrating how the one modifies the other when they transition into the nonlinear regime, either separately or together.

The character of the turbulent state achieved by the VSI is different than in traditional turbulent cascades. Preliminary indications of the turbulent development of the VSI (NGU13, Stoll & Kley 2014, aforementioned unpublished studies) point to a complex expression including quasi-steady vertical vortices, as well as radially alternating vertical velocity fields with aperiodic interfacial roll-up, as well as attendant small scale unsteady fluctuations concentrated at about one scale height from

the midplane. An analysis of the achievable outcome states of the VSI remains to be examined.

References

Arlt, R., & Urpin, V. 2004, A&A, 426, 755
Barker, A. L., & Latter, H. N. 2015, MNRAS, 450, 21 (BL15)
Fricke, K. 1968, Z. Astrophys., 68, 317
Goldreich, P., & Schubert, G. 1967, ApJ, 150, 571
Lubow, S. H., & Pringle, J. E. 1993, ApJ, 409, 360
McNally, C. P., & Pessah, M. 2015, ApJ, 811, 121
Nelson, R. P., Gressel, O., Umurhan, O. M. 2013, MNRAS, 435, 2610 (NGU13)
Stoll, M. H. R., & Kley, W. 2014, A&A, 572, A77
Turner, N. J., Fromang, S., Gammie, C., et al. 2014, Protostars and Planets VI, 411
Urpin, V. 2003, A&A, 404, 397
Urpin, V., & Brandenburg, A. 1998, MNRAS, 294, 399