# MRI turbulence and thermal instability in accretion discs

Johnathan Ross,[1][*] Henrik N. Latter[1] and Michael Tehranchi[2]

[1]*DAMTP, University of Cambridge, CMS, Wilberforce Road, Cambridge CB3 0WA, UK*
[2]*Statistical Laboratory, DPMMS, University of Cambridge, CMS, Wilberforce Road, Cambridge CB3 0WA, UK*

## ABSTRACT

A long-standing puzzle in the study of black hole accretion concerns the presence or not of thermal instability. Classical theory predicts that the encircling accretion disc is unstable, as do some self-consistent magnetohydrodynamic simulations of the flow. Yet observations of strongly accreting sources generally fail to exhibit cyclic or unstable dynamics on the expected time-scales. This paper checks whether turbulent fluctuations impede thermal instability. It also asks if it makes sense to conduct linear stability analyses on a turbulent background. These issues are explored with a set of MRI simulations in thermally unstable local boxes in combination with stochastic equations that approximate the disc energetics. These models show that the disc's thermal behaviour deviates significantly from laminar theory, though ultimately a thermal runaway does occur. We find that the disc temperature evolves as a biased random walk, rather than increasing exponentially, and thus generates a broad spread of outcomes, with instability often delayed for several thermal times. We construct a statistical theory that describes some of this behaviour, emphasizing the importance of the 'escape time' and its associated probability distribution. In conclusion, turbulent fluctuations on their own cannot stabilize a disc, but they can weaken and delay thermal instability.

**Key words:** accretion, accretion discs – instabilities – MHD – turbulence – X-rays: binaries.

## 1 INTRODUCTION

The assumption that the turbulent stress is proportional to pressure in accretion discs (the $\alpha$-model) is a fundamental ingredient of classic accretion disc theory (Shakura & Sunyaev 1973). It has been partially justified by successful application to quasi-steady systems, such as the thermal spectrum of dwarf novae and X-ray binaries (e.g. Warner 1995; Gierliński & Done 2004); but when applied to more delicate time-dependent dynamics, such as instabilities, the model has encountered difficulties. For instance, the modelling of dwarf nova outbursts requires different alphas for the high and low states (Smak 1984). Another example involves radiation-pressure-dominated accretion flows which the alpha model predicts are subject to thermal and viscous instability (Lightman & Eardley 1974; Shakura & Sunyaev 1976). X-ray observations, however, fail to find variability on the time-scales expected (Gierliński & Done 2004), with only the exceptional luminous source GRS 1915+105 and the intermediate black hole HLX-1 exhibiting anything like cyclic behaviour driven by thermal instability (Belloni et al. 1997; Done, Wardziński & Gierliński 2004; Sun et al. 2016; Wu et al. 2016). While it is possible the discs are stabilized by an additional but unknown cooling mechanism, it may be that the heating depends on temperature in a weaker way than the alpha model assumes. For example, the turbulent stresses may be proportional to gas pressure

rather than total pressure, or not on pressure at all (Gierliński & Done 2004).

Of course, one can bypass the alpha model (and its assumptions) and simulate the magnetohydrodynamic (MHD) turbulence in these discs self-consistently, with the turbulence then supplied directly by the magnetorotational instability (MRI; Balbus & Hawley 1991). And in fact early work indicated that radiation-pressure-dominated flows were thermally stable (Hirose, Krolik & Blaes 2009), in agreement with most observations. However, recent local and global simulations *do* exhibit thermal runaways (Jiang, Stone & Davis 2013; Mishra et al. 2016; Sadowski 2016), though these expose additional complications that may suppress instability, such as numerical effects (especially box size) and the impact of a mean magnetic flux. Obviously, both observations and simulations indicate that the onset and development of thermal instability is far less straightforward than that predicted by the classical laminar theory.

One very clear complication is the fact that the stress–pressure relationship can change, for both numerical and physical reasons, yielding instability or stability depending on conditions in the disc and in the simulation. A quite separate issue concerns the assumption that thermal instability can be appropriately defined at all, at least when dealing with quasi-steady turbulent states. The alpha theory treats the turbulence as a static eddy viscosity, and hence the equilibrium state as laminar. However, if the state hosts vigorous fluctuations it may not be well defined, or even make sense, to add a small linear perturbation on top of the stochastic background field, and subsequently calculate a growth rate. One envisages that, at the

very least, non-model, non-exponential behaviour ensues. Indeed, Jiang et al. (2013) report delayed runaway and algebraic growth rather than exponential growth in their MHD simulations, while Janiuk & Misra (2012) show via a stochastic 1D model that fluctuations induce random luminosity variations rather than the regular outbursts expected. It is to this aspect of the problem that this paper is devoted, focusing on the constructive and destructive interference of turbulence on thermal instability.

In order to isolate the essence of the problem, we employ an idealized model of MRI turbulence and of thermal instability. Unstratified shearing box simulations are performed using the code RAMSES on a state that is MRI turbulent and thermally unstable (at least according to the laminar alpha theory). Note that radiation pressure is omitted and the gas cools due to a simple cooling function. We find that the turbulent fluctuations induce thermal behaviour substantially different to that expected from laminar theory. In particular, the evolution of the temperature resembles more a biased random walk than an exponential runaway, with a wide range of trajectories possible: the temperature in some simulations departs from the laminar equilibrium relatively rapidly, whereas in others it can 'hang around' for several thermal times.

This motivates a probabilistic interpretation of instability, and we develop a simple statistical framework based on the model of geometric Brownian motion (GBM). A key idea is that of the 'escape time' $t_{esc}$ (which replaces the e-folding time). It describes how long it takes for the system to deviate significantly from the equilibrium. Reduced models involving both white noise and the power spectrum of the MRI show that the probability distribution of $t_{esc}$ possesses a long tail. Thus, there is a reasonable chance in any given simulation that thermal runaway is delayed. It should be stressed that ultimately realistic models of discs still undergo thermal runaways: turbulent fluctuations can impede instability but they cannot destroy it. The stabilization witnessed in Janiuk & Misra (2012) we attribute to the peculiarities of their stochastic model and a very large noise amplitude.

Another feature of our MHD simulations is thermal fragmentation when the cooling rate is too small, and hence the laminar thermal instability time-scale is too short. The disc then breaks up into hot and cold clouds. This occurs when the thermal mixing (by turbulence or radiative diffusion) is inefficient compared to thermal instability. On a sufficiently long length-scale, this is always presumably the case, but how this relates to the onset of instability in hot accretion flows is unclear. Estimates of both radiative diffusion and turbulent mixing suitable for the inner regions of X-ray binaries indicate that fragmentation is a marginal possibility.

The paper is organized as follows. In Section 2, we discuss a number of issues pertinent to thermal instability, stochastic fluctuations and the limits of the alpha theory (not all of which we take up in this paper). The third and fourth sections contain our numerical MHD model and the corresponding results, respectively. We then explore stochastic models in Section 5 and construct a statistical theory to help explain the MHD simulations. Our conclusions are then presented in Section 6.

## 2 THEORETICAL ISSUES

### 2.1 Stress–pressure relationship

The exact dependence of the turbulent stress, $\Pi_{xy}$, on pressure is key to the onset of thermal stability in radiation-pressure-dominated flows. The instability occurs when $\Pi_{xy} \propto (P_{gas} + P_{rad})$, but does not occur when $\Pi_{xy} \propto P_{gas}$ (Piran 1978). Here, $P_{gas}$ and $P_{rad}$ denote

gas and radiation pressure, respectively. The failure to observe signatures of thermal instability in most X-ray observations has been attributed to the stress depending on gas pressure alone or possibly the geometric mean of gas and radiation pressure (Gierliński & Done 2004). It has also been speculated that only exceptionally luminous flows could lead to a situation where $\Pi_{xy} \propto P_{rad}$, explaining the outbursts of GRS 1915+105 (though what exactly causes this shift in the stress's behaviour is unclear; Gierliński & Done 2004, but see also King & Ritter 1998). Numerical simulations of the MRI have since complicated this picture, as they indeed exhibit instability (Jiang et al. 2013), and we are left with the task of numerically tracing out the non-straightforward behaviour of $\Pi_{xy}$ in different conditions.

While earlier unstratified shearing box simulations without radiation pressure found only a weak dependence of stress on pressure (Sano et al. 2004), recent work has shown that $\Pi_{xy} \propto P_{gas}$ when the following conditions are satisfied: the computational domain is sufficiently large, explicit dissipation is included, and there is no net magnetic field (Ross, Latter & Guilet 2016, hereafter RLG16). Small boxes restrict the size of the turbulent eddies and prevent them from fully responding to an increase in pressure (be it gas or radiation). This restriction no doubt played a role in the failure of early radiation-pressure-dominated simulations to show thermal instability (Hirose et al. 2009): the radial domain was too small, the eddies unnaturally confined and, as a result, the stress unable to respond to changes in the total pressure. When larger radial boxes are used, as in recent work, the instability does in fact materialize (Jiang et al. 2013).

Another numerical effect uncovered by RLG16 was a sensitivity to the grid in simulations with no net magnetic field and no explicit diffusion. As in isothermal runs (Fromang & Papaloizou 2007), the stress is proportional to grid size and this leads to a significantly weaker stress–pressure relationship. Local boxes are not the only domains that exhibit the effect; recent vertically stratified simulations also show that the stress depends on the grid length (Ryan et al. 2017). The weaker dependence, both from numerical dissipation and from the box size, leads to artificially more stable systems (as can be shown from dimensional analysis). It is therefore necessary for both of these numerical complications to be considered when simulating thermal instability involving gas pressure. It is likely that global disc simulations of the MRI also suffer from strong numerical effects, though these have yet to be fully explored.

Another intriguing result from RLG16 is that the stress–pressure relationship depends on the existence and strength of any imposed magnetic flux. The stronger a mean toroidal flux, the weaker the relationship. If this behaviour generalizes to other field configurations and to the radiation-dominated regime, then one might speculate that highly magnetized flows are less prone to thermal instability. Indeed global simulations of the MRI suggest that it is impeded by strong magnetic fields (Sadowski 2016), possibly because they weaken the connection between the stress and pressure. This raises the interesting prospect that to assess susceptibility to thermal instability, we must also account for the build-up and evacuation of large-scale magnetic flux, in addition to the turbulent dynamics (e.g. Guilet & Ogilvie 2012, 2013).

### 2.2 Time lags

The alpha theory is a turbulence closure model, supplying a simple eddy viscosity in place of the complicated and chaotic time-dependent dynamics of the flow. On time-scales and length-scales much longer than the characteristic scales of the turbulence, this

approximation provides an adequate description, but its performance worsens the shorter the scales of interest. Since thermal instability can possess growth rates of tens of orbits (or less), then the detailed turbulent dynamics could potentially interfere with its onset.

Simulations, as expected, show that on shorter time-scales, the stress–pressure relationship is more complicated than a mere proportionality. One interesting feature is a time lag of a few orbits between the stress and pressure. Moreover, it is the pressure that follows the stress on shorter times, rather than the other way around. And so the dependence is opposite to that assumed by the alpha theory: bursts in stress are followed by jumps in pressure (Hirose et al. 2009). What is happening here is that the bursts in stress drive fluctuations in the heating rate (once their associated energy has reached the dissipation scale) and hence cause bursts in pressure a short time later.

The effect of this time delay on a stable thermal equilibrium was first explored by Hirose et al. (2009), who argued that if the direction of causation was from the stress to the pressure then thermal instability may not work. The argument fails, however, to acknowledge that the stress–pressure dynamics exhibit different time-scales, with longer time-scales (>10 orbits) characterized by a pressure-dependent stress (as in the classical theory), while the short time-scales show the lags described above (Latter & Papaloizou 2012, RLG16).

Follow-up work by Lin, Gu & Lu (2011) and Ciesielski et al. (2012) presents linear instability analyses of an alpha disc model with a time-delayed alpha. These show that the simulated delay of 1–10 orbits is insufficiently long to stabilize a thermally unstable state. They also reveal potential inconsistencies in such simple models: for example, infinite growth rates are possible for certain time-lags, a physical impossibility. Of course, these models are also incomplete, as they include only either the short-term or long-term dependences, but not both concurrently. Lastly, it is possible that the flow exhibits an additional time delay affecting the long-term stress-pressure relation, now with the stress lagging behind the pressure. This has not been verified by simulations yet, but if present may weaken thermal instability.

### 2.3 Temporal fluctuations

MRI turbulence has strong variation over a range of time-scales, from tens of orbits to a few shear times (Sano et al. 2004; Lesur & Ogilvie 2008). In what sense does a thermal equilibrium exist in such a system? Time-varying perturbations are constantly emerging which lead to a shifting balance of heating and cooling that the system continually responds to. On the other hand, if we assume that there is well-defined mean equilibrium, then it is awash in finite amplitude fluctuations. How can one then undertake a linear instability analysis? Is it meaningful to add a tiny perturbation on top of a sea of finite amplitude perturbations and check if it grows or not? On long length and time-scales this might work, but certainly not on shorter scales.

Putting aside the difficulty of interpreting linear stability analyses, a stochastic system exhibits a range of complicated and sometimes unexpected behaviour. A classic example is the destabilization of fixed points deemed stable by laminar theory. Originally studied in biological population dynamics (e.g. Levins 1969; May 1973), this feature of noisy systems appears in numerous applications, such as atmospheric modelling (e.g. De Swart & Grasman 1987), where the unresolved short time and lengths dynamics are represented by stochastic terms (see Majda, Timofeyev & Vanden-Eijinden 1999,

2003). On the other hand, the influence of stochasticity on an otherwise *unstable* fixed point has been studied in financial mathematics, where GBM can be used to model volatile stock prices in a rising market. Despite the mean trend of increasing prices, stochasticity can depress the price of some stock dramatically, if not stabilizing the fixed point then delaying a runaway in price for some period of time.

The very last example is perhaps the most relevant for our study, as it possesses the key ingredients of (a) an unstable fixed point (according to a deterministic or 'laminar' theory) and (b) stochastic fluctuations. The competition between them gives rise to behaviour one might liken to a biased random walk. In between the kicks delivered by the turbulence, the system drifts according to the deterministic unstable dynamics. One can then imagine certain limits: when the characteristic frequency of the turbulence is much greater than the thermal instability growth rate, then we may expect an unbiased random walk, and the system will only weakly sense the underlying thermal physics. In the opposite limit, when the growth rate is much greater than the turbulent frequency, the deterministic laminar dynamics should be reproduced. It is in the intermediate regime, explored in this paper, that interesting non-trivial behaviour manifests. There are also other key ratios, such as the size of the kicks relative to the magnitude of the fixed point or the initial condition. If these are too small, then we return to the laminar case. But for intermediate values, as exhibited by our MRI simulations, system trajectories can deviate markedly from both the laminar behaviour and a simple unbiased random walk.

### 2.4 Spatial fluctuations

In the previous subsection, we considered only temporal fluctuations on the system variables, implicitly regarding them as 'box averaged' or mean quantities. Indeed, the $\alpha$ model assumes a homogenized temperature over $\sim H$, the disc scaleheight. Turbulent heating, however, is spatially inhomogeneous with strong dissipation occurring in current and vorticity sheets and minimal dissipation in the surrounding regions. These spatial fluctuations also complicate the picture of thermal instability, especially when the instability growth rate is large.

Some form of thermal mixing is necessary to homogenize the temperature of the fluid. This can take multiple forms, such as turbulent advection, radiative diffusion or thermal conduction. When thermal instability is present, the assumption of a uniform temperature is reasonable as long as the instability time-scale is longer than the mixing time-scale, $t_{inst} > t_{mix}$. For less efficient mixing or stronger instability, there exists a maximal thermal coupling length-scale, $l \sim v_{mix} t_{inst}$. Regions separated by more than this only weakly interact thermally during an e-folding time. This implies that regions of a disc separated by more than $l$ can undergo thermal runaways independently, and the disc fragments into cold and hot clouds. Regions of strong kinetic and magnetic dissipation are likely to heat catastrophically, while those with weak dissipation will cool catastrophically. How relevant this scenario is in realistic discs is unclear, though perhaps marginally possible in X-ray binaries. It is certainly possible in numerical simulations as we show later.

## 3 NUMERICAL TOOLS AND SETUP

### 3.1 Formulation

We wish to explore the essential features of thermal stability and its onset in MRI-driven turbulence and so we choose an idealized set-up to isolate it. We adopt the local shearing box model

(Goldreich & Lynden-Bell 1965). To prevent complications such as buoyancy, mass-loss and disc expansion, we consider the unstratified case. With this model, MRI-driven turbulence can be obtained when a Keplerian flow profile is assumed (Hawley, Gammie & Balbus 1995). As is conventional, $x$, $y$, $z$ are the radial, azimuthal and vertical spatial variables and $\hat{\boldsymbol{e}}_x$, $\hat{\boldsymbol{e}}_y$, $\hat{\boldsymbol{e}}_z$ are the corresponding unit vectors. This frame of reference corotates with the disc at some radius with angular frequency $\boldsymbol{\Omega} = \Omega \hat{\boldsymbol{e}}_z$. The ideal compressible MHD equations are hence

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{v}) = 0, \tag{1}$$

$$\rho \frac{\partial \boldsymbol{v}}{\partial t} + \rho (\boldsymbol{v} \cdot \nabla) \boldsymbol{v} = -2 \rho \boldsymbol{\Omega} \times \boldsymbol{v} + 3 x \rho \Omega^2 \hat{\boldsymbol{e}}_x - \nabla P$$
$$+ (\nabla \times \boldsymbol{B}) \times \boldsymbol{B} \tag{2}$$

$$\frac{\partial \boldsymbol{B}}{\partial t} = \nabla \times (\boldsymbol{v} \times \boldsymbol{B}), \tag{3}$$

$$\frac{\partial \varepsilon}{\partial t} + \boldsymbol{v} \cdot \nabla \varepsilon = -P \nabla \cdot \boldsymbol{v} + Q - \Lambda, \tag{4}$$

where $\rho$ is the mass density, $\boldsymbol{v}$ is the velocity, $P$ is the gas pressure, $\boldsymbol{B}$ is the magnetic field and the internal energy is denoted by $\varepsilon$. Heating is represented by $Q$ and cooling by $\Lambda$. This set of equations is then closed by relating the internal energy to the pressure by assuming an ideal gas so that

$$\varepsilon = P/(\gamma - 1), \tag{5}$$

where $\gamma$ is the adiabatic index, taken to be 7/5. The sound speed is then given by $c_s = (\gamma P/\rho)^{1/2}$ and the pressure scaleheight by $H = (2/\gamma)^{1/2} c_s/\Omega$.

Ideally, we would include viscosity and Ohmic diffusion and so $Q$ would be given by the sum of the physical dissipative processes; however, to resolve the diffusion length-scales requires a higher resolution than is practical. Instead, we rely on numerical dissipation for heating. By solving equations (1)–(4) in conservative form, the kinetic and magnetic energy dissipated by the grid is converted to internal energy. Energy that is extracted from the background shear is converted to internal energy and ultimately removed via cooling.

Finally, for our cooling function, we take a power law of pressure,

$$\Lambda = \theta\, P^m, \tag{6}$$

where both $\theta$ and $m$ are constants. Though this choice is mainly for convenience, it might crudely approximate an optically thin medium.

### 3.2 Numerical methods

All of the simulations that we perform are carried out using RAMSES, a finite-volume Godunov code based on the MUSCL–Hancock algorithm (Teyssier 2002; Fromang et al. 2006). The HLLD Riemann solver (Miyoshi & Kusano 2005), and the multidimensional slope limiter described in Suresh (2000) are used in all the simulations presented in this paper.

Rather than solving for the total $y$-momentum, we evolve the equivalent conservation law for the angular momentum fluctuation $\rho v_y' = \rho(v_y - v_K)$, with $v_K$ the Keplerian velocity. An upwind solver is used for solving the azimuthal advection arising from $v_K$. The tidal and Coriolis forces are treated as source terms and implemented following the Crank–Nicholson algorithm described in Stone & Gardiner (2010).

The algorithm solves for the total fluctuation energy $E' \equiv \varepsilon + \rho v'^2/2 + B^2/2$ and its conservation law is written as

$$\frac{\partial E'}{\partial t} + \nabla \cdot (E' \boldsymbol{v}' + \boldsymbol{v}' \cdot \boldsymbol{P}) = -v_K \frac{\partial E'}{\partial y}$$
$$+ \left( B_x B_y - \rho v_x v_y' \right) \frac{\partial v_K}{\partial x} - \Lambda, \tag{7}$$

where $\boldsymbol{P}$ is the total pressure tensor

$$\boldsymbol{P} = (P + B^2/2)\boldsymbol{I} - \boldsymbol{B}\boldsymbol{B}. \tag{8}$$

The left-hand side of equation (7) comprises the usual energy conservation law, which we solve using the MUSCL–Hancock algorithm. The treatment of the two terms on the right-hand side have been modified: the azimuthal advection of energy is solved with an upwind solver, and the second term involving the Maxwell and Reynolds stresses is added as a source term.

For the set of simulations shown in this paper, we used a box size of $(L_x, L_y, L_z) = (4, 5, 4)H_0$ with a resolution of $(N_x, N_y, N_z) = (128, 100, 128)$. $H_0$ is a reference scaleheight that is close to but not exactly the same as the scaleheight at the start of a simulation. It will be defined in detail later. The grid scale is defined to be $\Delta = L_x/N_x$. We set $\Omega = 10^{-3}$ and $c_{s0} = 10^{-3}$ in code units, where $c_{s0}$ is the initial sound speed. The resolution is low in comparison to other MHD shearing box simulations; however, it is sufficient to capture the basic properties of the MRI, in particular the turbulent fluctuations. Importantly, it is computationally inexpensive allowing the simulations to run for $\gtrsim 500$ orbits.

### 3.3 Thermal equilibrium

In our shearing box model, the energy is injected into the computational domain by the second term on the right-hand side of equation (7) which represents the liberation of shear energy by the total stress:

$$\Pi_{xy} = B_x B_y - \rho v_x v_y'. \tag{9}$$

Simulations have found that the box-averaged stress $\langle \Pi_{xy} \rangle$ is roughly proportional to the box-averaged gas pressure to a given power (Sano et al. 2004; RLG16), though be aware of the caveats given in Section 2. Therefore, in numerical simulations, the averaged heating rate may be approximated by

$$\langle Q \rangle \propto \langle \Pi_{xy} \rangle = \tilde{\alpha} \langle P \rangle^q, \tag{10}$$

where $\tilde{\alpha}$ and $q$ may be calculated from the simulation. Note that the former is *not* the same as the $\alpha$ parameter. The exponent $q$ depends on the field geometry as well as the numerical parameters (RLG16). For our set-up, $q = 0.5$, as long as the pressure is sufficiently small (see later).

Combining this approximation with our cooling prescription, the evolution of the volume-averaged pressure, $\langle P \rangle$, is determined by

$$\frac{\mathrm{d}\langle P \rangle}{\mathrm{d}t} \approx (\gamma - 1)\frac{3}{2}\Omega \langle \Pi_{r\phi} \rangle - (\gamma - 1)\theta \langle P \rangle^m \tag{11}$$

$$\cong a_1 \langle P \rangle^q - a_2 \langle P \rangle^m, \tag{12}$$

where

$$a_1 = \frac{3}{2}\Omega(\gamma - 1)\tilde{\alpha}, \qquad a_2 = (\gamma - 1)\theta. \tag{13}$$

In the formulation of equation (11), we have made the approximation that $\langle P^m \rangle \cong \langle P \rangle^m$, which is reasonable as long as the variance
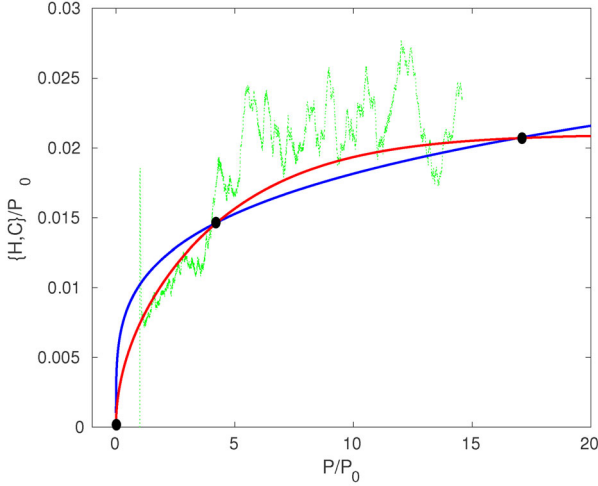
**Figure 1.** For illustrative purposes, the red curve shows the fitted heating rate calculated from a zero net-flux simulation with resolution $\frac{1}{64} H_0 = \Delta$ from RLG16 (shown in green). The blue line is an example cooling function $\propto P^{0.25}$. The fixed points are shown by black dots.

of $P$ within the box at fixed time is not too large. This system has two fixed points

$$P_1 = 0, \tag{14}$$

$$P_2 = \left(\frac{a_1}{a_2}\right)^{1/(m-q)} = \left(\frac{3\Omega\tilde{\alpha}}{2\theta}\right)^{1/(m-q)}. \tag{15}$$

In shearing box simulations, the thermodynamics have an additional complication: stress is independent of pressure once the scaleheight is larger than the radial and vertical box sizes (RLG16). The exponent $q$ may then be viewed as a function of the disc temperature, equal to 0.5 for cool discs, while asymptoting to 0 as the gas heats and the scaleheight $H$ equals the box size $L_z$. This behaviour introduces the possibility of an additional stable fixed point if $m < \max\{q\}$. In Fig. 1, we plot (a) the heating rate as a function of pressure from an $L = 4H_0$, $\Delta = 1/64$ simulation (appearing in RLG16), (b) a smooth fit to this curve and (c) overlay an $m = 0.25$ power-law cooling. Where the latter two curves intersect give three thermal equilibria. These are the $P_1$ and $P_2$ fixed points, described above, in addition to a third equilibrium, $P_3$, which arises from the finite size of the box. Note that a similar artificial fixed point will also appear in vertically stratified simulations.

The above volume-averaged analysis is deterministic, but, turbulent flows are not. The stress fluctuates around its mean value as a result of the formation and break-up of coherent structures within the flow. This means that $Q$ can no longer be expressed in as simple a form as equation (10). Instead, the stress is determined by the sum of a deterministic term and a fluctuating term. By using equation (11), we can obtain an estimate of the equilibrium pressure that the system feels at any given time

$$P_{\exp} = \left(\frac{3\Omega\langle\Pi_{xy}\rangle}{2\theta}\right)^{1/m}. \tag{16}$$

### 3.4 Thermal instability

According to a linear analysis of equation (4), the thermal instability criterion is

$$\frac{dQ}{dP} > \frac{d\Lambda}{dP} \tag{17}$$

which, given our power-law expressions for $Q$ and $\Lambda$, leads to the simple condition $q > m$ for $P_2$, and the opposite stability for $P_1$ and $P_3$.

We calculate the growth rate as

$$s_g = (q - m)(\gamma - 1)\left(\frac{3\tilde{\alpha}\Omega}{2}\right)^{(1-m)/(q-m)} \theta^{(q-1)/(q-m)}, \tag{18}$$

which can be used for comparison with the MHD simulations. The associated instability time-scale we define to be $t_{\text{inst}} = 1/s_g$.

If the system was truly laminar, the instability criterion would be fully determined by $p$ and $q$ and the instability time-scale from equation (18). Fluctuating systems, however, exhibit non-exponential and indeed non-monotonic behaviour that is poorly approximated by the laminar model. In particular, $t_{\text{inst}}$ may be an unsuitable measure for the instability time-scale. In its place, we introduce the 'escape time' $t_{\text{esc}}$ of a simulation, which is defined to be the last instance that the system lies within a pressure interval containing the fixed point. Mathematically, it may be defined via

$$t_{\text{esc}} = \max\left\{t > 0 : |P(t) - P_{\text{eqm}}| = \delta\right\}, \tag{19}$$

where $P_{\text{eqm}}$ is one of the three equilibria introduced earlier, and $\delta$ is the interval size, outside of which we consider the system to have unequivocally departed from the fixed point. We are free to specify the size of $\delta$, and it might reflect the particular problem of interest. For disc transitions between states differing by many orders of magnitude in temperature, we might be generous with $\delta$, permitting it to be up to 10 times the fixed point pressure. For smaller transitions, then $\delta$ must be smaller. Note that the system may stochastically dip in and out of this interval, but $t_{\text{esc}}$ will capture the time when the system finally leaves it forever. If the disc is laminar, then

$$t_{\text{esc}} = t_{\text{inst}} \ln\left(\frac{\delta}{|P_{\text{eqm}} - P_{\text{init}}|}\right), \tag{20}$$

where $P_{\text{init}}$ is the initial pressure of the system.

### 3.5 Initial conditions

If we want a thermal equilibrium to be achievable within our box and for the stress to depend appreciably on pressure, the initial conditions must be chosen carefully. An initial state that is too hot means the box size will unduly influence the stress and weaken $Q$'s dependence on $P$. As a consequence, thermal instability will fail to occur.

Stress is only observed to be a strong function of pressure ($\Pi_{xy} \propto P^{1/2}$) when $\Delta \ll H < L$ (Sano et al. 2004; RLG16). Therefore, $P_2$ must be sufficiently low so that the stress is increasing with pressure, but sufficiently high so that the characteristic length-scale of the turbulence is not on (or below) the grid. We choose

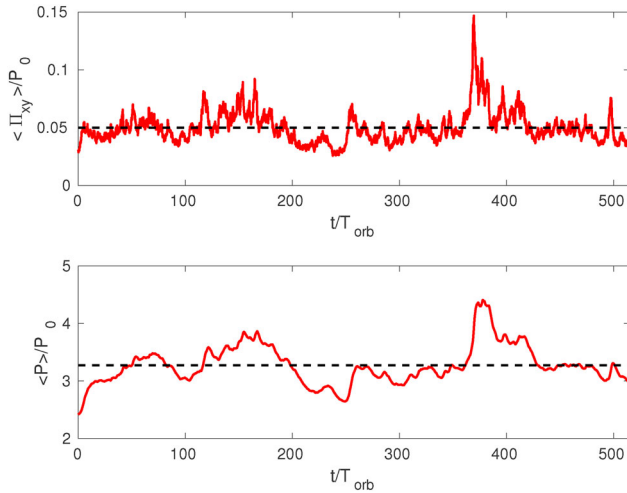$$\frac{\sqrt{P_2}}{\sqrt{P_{\text{box}}}} = \frac{H_2}{L} \approx 0.5, \tag{21}$$

**Figure 2.** The evolution of the box-averaged pressure and stress from simulation R1.



**Figure 3.** A zoom-in of the pressure and stress evolution in Fig. 2. The blue dashed vertical line indicates the time at which the stress achieves its maximum in this interval.

where $P_{box}$ is the pressure at which the scaleheight equals the box size and $H_2$ is the scaleheight when $\langle P \rangle = P_2$. From this, we can obtain the required value of $\theta$ for a given choice of $m$:

$$\theta = \frac{(3/2)\tilde{\alpha}}{(0.25 P_{box})^{(m-q)}}. \qquad (22)$$

To achieve a turbulent state, we first initialize a zero net-flux simulation with an initial field of $\boldsymbol{B} = B_0 \sin(2\pi x)\hat{\boldsymbol{e}}_z$ and no cooling. In code units, $B_0 = \sqrt{2/\beta}$ which we set with $\beta = 10^3$. Once a turbulent state is reached, we switch on cooling with $\theta = 10$ and $m = 2$. This choice of $m$ is to obtain a $P_2$ which we expect to be stable, as discussed in Section 3.4. These parameters lead to $P_2 \approx 3.2 P_0$, where $P_0$ is the initial pressure, having used that $P_{box} \approx 16 P_0$. During this steady state, we calculate $\tilde{\alpha} \approx 2.3 \times 10^{-5}$. This fully turbulent state in thermal equilibrium is used as our initial condition. The parameters $m$ and $\theta$ are then changed as appropriate.

## 4 RESULTS

### 4.1 Stable system

To fix ideas, we first consider the stable thermal equilibrium associated with the initial condition, $\theta = 10$, $m = 2$ (simulation R1). We run the simulation for over 500 orbits and the resulting stresses and pressures are shown in Fig. 2. As predicted from the laminar linear analysis, the equilibrium is 'stable' in the sense that both the stress and pressure fluctuate within some interval enclosing $P_2$. Though the stress shows substantial variation during the simulation (with a maximum value of $\approx 3$ times its mean), there is no runaway or mean drift during the 500 orbits. The variability in the pressure is less extreme, with a maximum value of $\approx 1.5$ times its mean value. The turbulence administers random 'kicks' to the system, but the deterministic physics always draws it back to the vicinity of $P_2$. The stochastic fluctuations in stress are a result of dynamo cycles and the formation and break-up of coherent structures. The energy in the flow then cascades down to the dissipation scale, where the magnetic and kinetic energies are converted to thermal energy leading to changes in pressure.

A short-time delay of a few orbits ($t_{orb}$) between the stress and pressure is clearly visible in Fig. 3, a result of the finite time taken
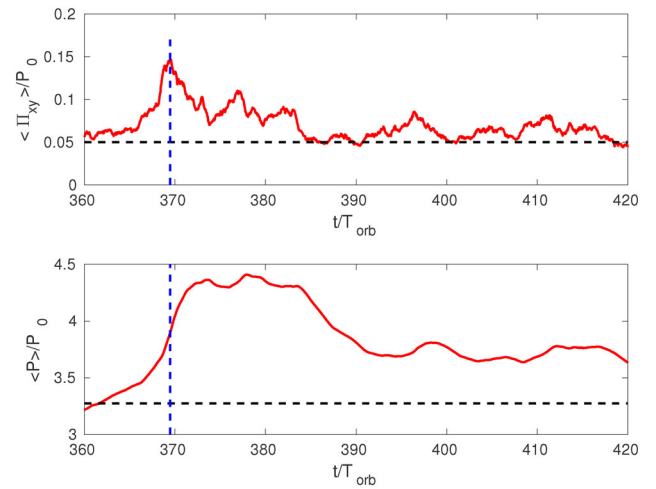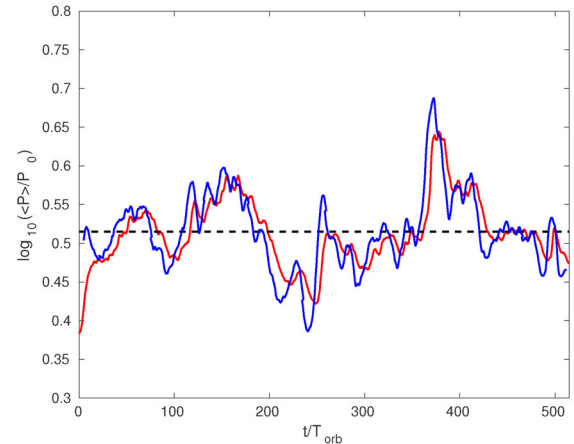


**Figure 4.** The evolution of the logarithm of the box-averaged pressure (red) along with the logarithm of the instantaneous expected equilibrium pressure calculated using an average stress over the previous $9 t_{orb}$ (blue) for the thermally stable simulation R1. The black dashed line indicates the mean equilibrium pressure.

for kinetic and magnetic energies to reach the dissipation scale from the injection scales (Hirose et al. 2009). The variations in pressure are smoother and longer than those in the stress. However, this short-time dependence of pressure on stress is distinct to the longer time dependence of stress on pressure that drives instability/stability (Latter & Papaloizou 2012).

Though the pressure seems to be drawn back to the mean of $P$, in actual fact the equilibrium balance that the system feels at any given instance is changing with time. A crude approximation to this is the instantaneous fixed point $P_{exp}$, calculated using equation (16). But because the thermal time-scale is longer than the turbulent time-scale, a better approximation is the average of $P_{exp}$ over the last thermal time. Using equation (18), we estimate this time to be $\approx 9$ orbits. The resulting smoothed and time-dependent equilibrium is plotted in Fig. 4 in blue, alongside the actual pressure of the system $\langle P \rangle$ in red. It is clear that $\langle P \rangle$ follows the 'equilibrium' with a time lag of $\lesssim 10$ orbits. Despite the fluctuations, the system senses
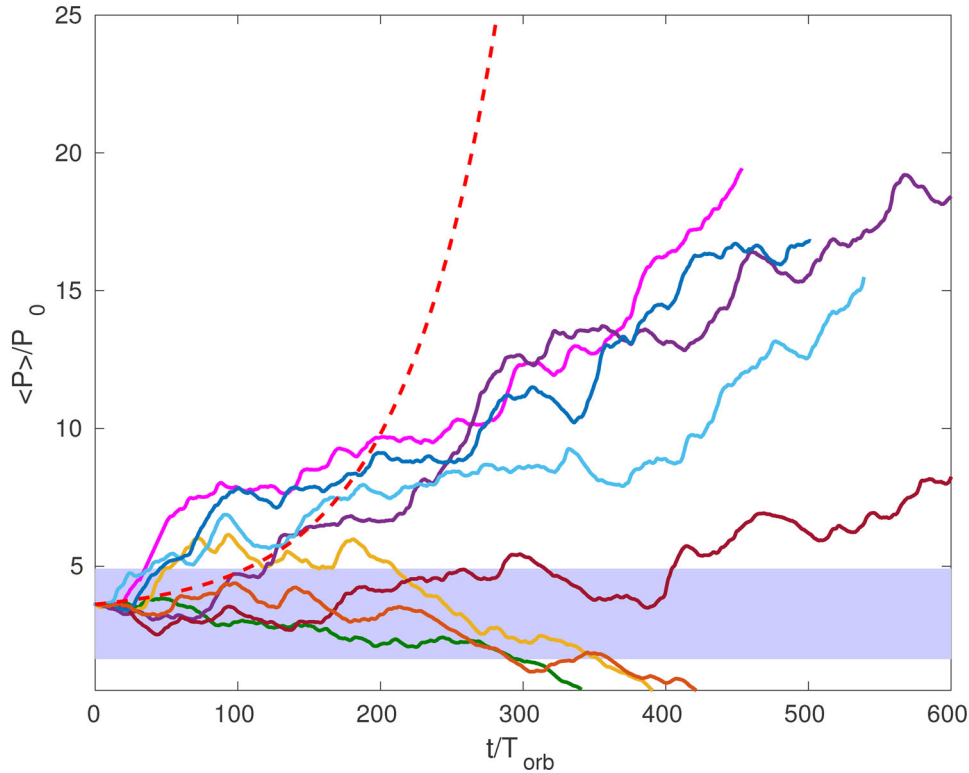
**Figure 5.** Pressure versus time for the unstable simulations R2a–R2h shown by solid curves. The dashed red line is linear evolution. The filled area indicates the fixed point region $0.5P_{eqm} < P < 1.5P_{eqm}$.

the stable fixed point and is attracted towards it, though because of those same fluctuations it can never come to rest upon it.

## 4.2 Unstable systems

Having explored the stable case we now proceed to an unstable thermal equilibrium. For this set of simulations, we choose $m = 0.25$ which we expect to be unstable based on the argument in Section 3.4.

We present eight simulations (R2a–R2h) that have the same fully turbulent initial state. In order to vary the initial condition between runs, these simulations each have slightly different $\theta \approx 1.31 \times 10^{-9}$ and so have slightly different $P_2$. Each of these is within a few per cent of the initial pressure. In practice, this means that though each simulation starts from the same initial condition, each corresponds to a slightly different perturbation from equilibrium.

In Fig. 5, we show the evolution of the box-averaged pressures alongside the trajectory derived from the laminar linear theory. Unsurprisingly, the turbulent fluctuations lead to a diversity of outcomes but do not indefinitely prevent thermal runaway. For example, a large kick can cause the system to escape from the fixed point on a shorter time-scale than the laminar time-scale, or the system may remain close to the fixed point for extended periods of time, longer than $t_{inst}$. Our most 'stable' simulation remains close to equilibrium for $400t_{orb}$. None of the simulations can be well modelled by the laminar theory. For $P \gtrsim 5P_0$, the behaviour is closer to algebraic growth than exponential runaway. In fact, the behaviour of the system is strongly influenced by the fluctuations that occur on time-scales comparable to the instability time-scale, $\sim 70 t_{orb}$ for these parameters. A more apt description of the system could be a biased random walk, when strong fluctuations in stress repeatedly perturb

the system, while in between kicks the system drifts according to the deterministic physics.

If we consider some characteristic interval around the fixed point then we can find the very last time, $t_{esc}$, the system was within this band, equation (19). If we define the interval to be rather narrow, $0.5P_{eq} < P < 1.5P_{eq}$, then this 'escape time' can be compared to $t_{inst}$. We find a wide range of escape times in our MHD simulations, from $0.25 t_{inst}$ to $5 t_{inst}$. Ideally, we would calculate a probability distribution function for the escape time but this would require substantially more simulations, which, at this time, is impractical. The choice of band is arbitrary, but it should be chosen to suit the problem.

Once $P > 10P_0$, the simulations become cooler than the laminar model prediction. This might be attributed to the box size beginning to influence the evolution. At this point, the dependence of stress on pressure decreases (see Fig. 1) and hence the laminar model is an overestimate. This numerical effect introduces a third equilibrium point, $P_3 > P_2$, as discussed in Section 3.3. We do not observe a plateau in pressure associated with the system being attracted to $P_3$, but we expect that if run for a sufficient duration then a plateau would appear.

That a runaway heating is limited by the box size is an important problem, both for these simulations and stratified radiation MHD simulations. To explore this further, we rescale the simulations by choosing a larger value of $\theta$, and initialize a simulation with the same initial turbulent state as in R2a–R2h. However, now the initial condition is much further away from the putative equilibrium, by a factor of some 3, and yet box effects remain negligible, because equation (21) is still satisfied. In these runs, the thermal runaway was somewhat faster than witnessed in Fig. 5 for the hotter systems. These few simulations illustrate the numerical limitations inherent
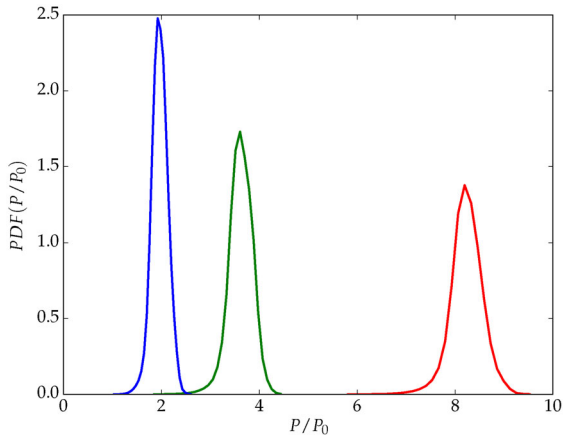
**Figure 6.** The pressure probability distribution function $PDF(P/P_0)$ from a heating (red) and a cooling run (blue) at $t = 340\,t_{\rm orb}$. The green curve shows the initial distribution.



**Figure 7.** The pressure probability distribution function from a fragmenting run at $t = 40\,t_{\rm orb}$. The green, red and green curves show the distribution at $t = 0, 4, 14\,t_{\rm orb}$.

in any simulation of catastrophic heating undertaken in a finite domain.

### 4.3 Thermal fragmentation

In a turbulent system, heat is not deposited uniformly throughout the box, but rather is localized in coherent structures such as current and vorticity sheets. If the instability time-scale, $t_{\rm inst}$, is large compared to the mixing time-scale, $t_{\rm mix}$, then the heating inhomogeneity will have little effect as temperature fluctuations will be smoothed out. Conversely, if $t_{\rm mix} > t_{\rm inst}$ then regions of fluid can evolve independently of each resulting in localized runaway. In Fig. 6, we plot the pressure probability distribution function, $PDF(P)$, for two R2 simulations at $t = 340\,t_{\rm orb}$. Inefficient turbulent mixing would result in the spreading of $PDF(P)$ during runaway. The figure exhibits minimal evidence of spreading in the heating run, and the pressure ensemble evolves with a well-defined and relatively narrow shape. The pressure in the cooling simulation is attracted to $P_1 = 0$ and in fact we see a further narrowing of the distribution as the equilibrium is approached.

To show localized runaway and fragmentation, we consider a case that we expect to be very unstable, choosing $m = 0.1$ and $\theta \approx 1.84 \times 10^{-10}$ (simulation R3). In Fig. 7, we plot $PDF(P)$ at three instances of time. The width of the distribution quickly increases, indicating localized thermal runaway: initially $\mathrm{Var}(P/\langle P \rangle) = 0.0045$, but after 4 and 14 orbits this grows to 0.019 and 0.057, respectively. During this time $\langle P \rangle$ itself varies little, which emphasizes that the box-averaged properties no longer give a satisfactory description of the state of the system. Soon after the final snapshot, very small pressures occur resulting in the termination of the simulation, preventing further exploration. In Fig. 8, we show $x - z$ slices in pressure before and during thermal fragmentation. Prior to thermal fragmentation, there are strong acoustic waves propagating in the radial direction. These appear to break up into patches that undergo rapid thermal runaway independent of each other. Because this behaviour appears to be significantly non-linear and disordered, we do not attribute it to the action of a linear instability mode with non-zero $k_x$ and $k_z$.

Though it is straightforward enough to achieve fragmentation in an unstratified local box, how likely is this in the inner regions of an X-ray binary? In this context, the two major contributors to mixing are radiative diffusion and turbulent advection. We first consider
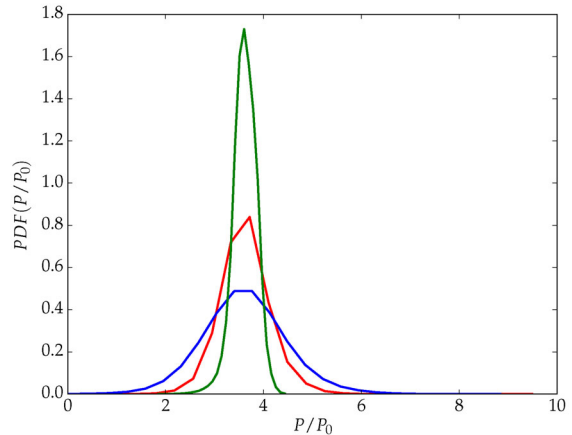
radiative diffusion. In the hot dense gas, the opacity is largely dominated by Thomson scattering with opacity $\kappa_T = 0.33\,{\rm cm^2\,g^{-1}}$. The radiative diffusion time across a length of $l$ is then $t_{\rm rad} \sim l^2/c\lambda$, where $c$ is the speed of light and $\lambda$ is the mean free path. The instability time-scale we estimate to be of order, but bounded below, by the thermal time-scale so that $t_{\rm inst} \gtrsim t_{\rm th} \sim (\alpha\Omega)^{-1}$. If $t_{\rm rad} \approx t_{\rm th}$, we have the following condition on $l$:

$$\frac{l}{H} \sim \left( \alpha \Sigma \kappa \frac{c_{\rm s}}{c} \right)^{-1/2}, \tag{23}$$

$$\sim 0.1 \left( \frac{\alpha}{0.1} \right)^{-1/2} \left( \frac{\Sigma}{10^5\,{\rm g\,cm^{-2}}} \right)^{-1/2} \left( \frac{T}{10^7\,{\rm K}} \right)^{-1/4}, \tag{24}$$

where $\Sigma$ is surface density. Regions separated by more than $l$ will be unable to mix sufficiently well on the instability time-scale. Typical values for an X-ray binary indicate that regions $0.1\,H$ apart may in fact thermally fragment.

What about turbulent mixing? We assume that the turbulent transport of heat by the MRI is similar in efficiency to its transport of angular momentum (though this is a point that has not been studied in detail). If we are permitted this assumption, then the turbulent diffusion time-scale is $t_{\rm turb} \sim l^2/(\alpha c_{\rm s} H)$. If we next assume that the relevant eddies are of size $H$ (plausible if some form of MHD convection is operative), then $t_{\rm turb} \sim t_{\rm inst}$ on these outer scales. These very rough scalings indicate that turbulent heat transport is somewhat more efficient than radiative transport, and moreover that it may be sufficient to preclude fragmentation – though a simple order of magnitude treatment is unable to determine precisely when this might occur. Only realistic simulations themselves can decide on this issue, and in fact Jiang et al. (2013) do not find fragmentation. Our simulations are marginally susceptible, in agreement with the above argument, but they omit important physical effects such as buoyancy, which may be crucial here, and enhanced compressibility effects in radiation-dominated flow.

## 5 REDUCED STOCHASTIC MODELS

Because it is impossible to run a sufficient number of simulations to build reliable statistics, especially regarding the distribution of $t_{\rm esc}$, we turn to simpler approximate models that illustrate more fully the effects of the fluctuations on stability and also permit analytical results.
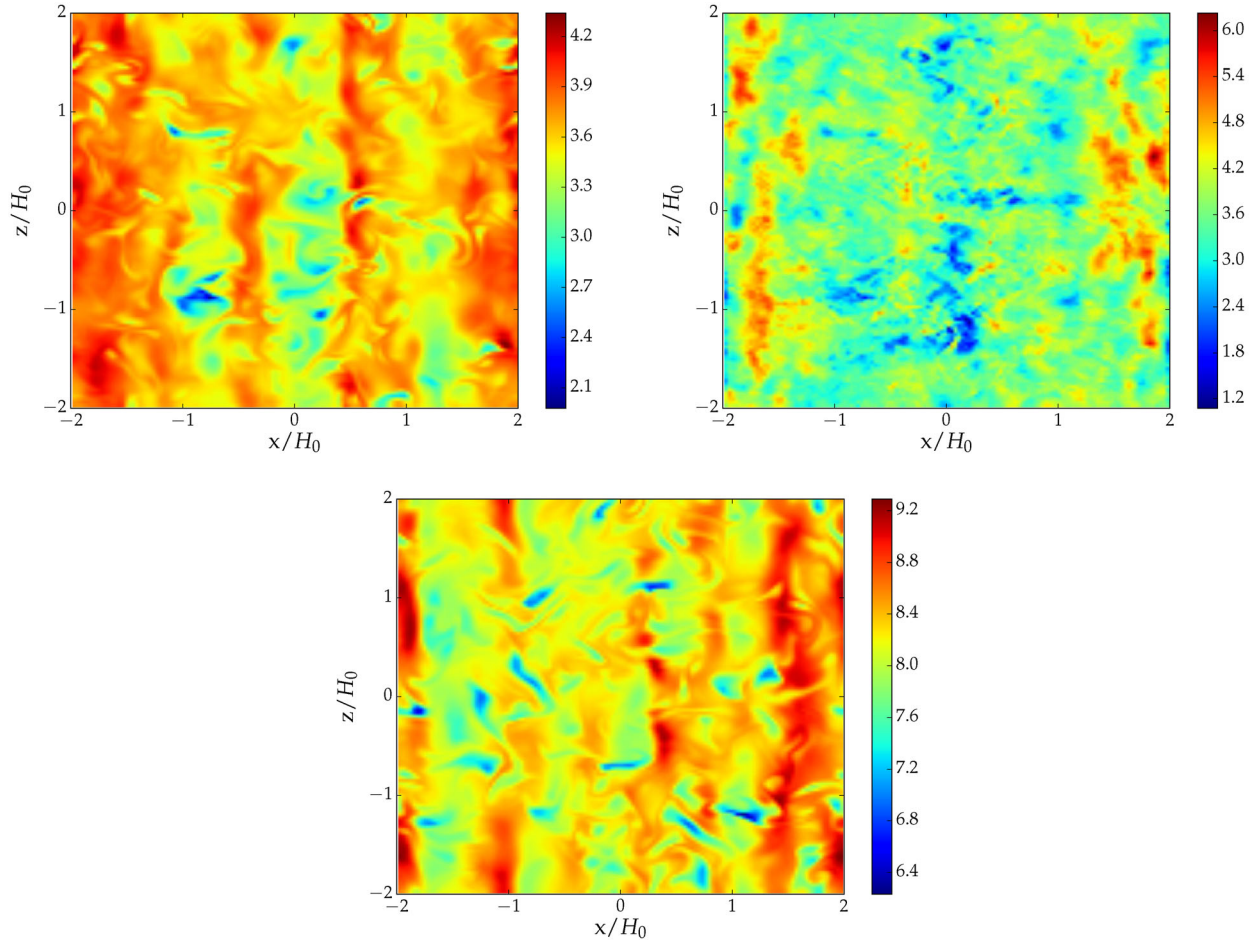
**Figure 8.** Heat maps of $P/P_0$ in common $x - z$ slices. The top-left panel is from the initial turbulent state, the top-right is from the fragmenting simulation R3 at $t = 4\,t_{\mathrm{orb}}$. In the latter, the difference in pressure between the coolest and hottest blobs is a factor of 6. For comparison, we show a R2 simulation undergoing a slow heating runaway at $t = 340\,t_{\mathrm{orb}}$ in the bottom panel. Here, the pressure difference is merely a factor of 1.5.

We work primarily with the averaged energy equation (12) but model the fluctuating turbulent stress via a random function $\zeta$. Our model is related to the logistic equation, and may be written as

$$\frac{\mathrm{d}x}{\mathrm{d}\tau} = [1 + \zeta(\tau)]\,x^q - \Theta x^m. \tag{25}$$

This can be derived from (12) by an appropriate rescaling, with $x$ and $\tau$ representing pressure/temperature and time, respectively. Constant parameters are $q$, $m$ and $\Theta$. To simplify the analysis while not losing much generality, we set $q = \Theta = 1$ in much of what follows.

Equation (25) admits the trivial steady state $x = 0$ and the more interesting equilibrium $x_{\mathrm{eq}} = 1$. In the 'laminar' case of $\zeta = 0$, this equilibrium is unstable when $m < 1$ with modes possessing the growth rate $1 - m$.

For $\zeta(\tau)$, a random but *continuous* function, (25) is the Bernoulli equation with analytic solution

$$\frac{x}{x_{\mathrm{eq}}} = \phi(\tau)\left[\left(\frac{x_0}{x_{\mathrm{eq}}}\right)^{1-m} - (1-m)\int_0^\tau \phi(s)^{m-1}\mathrm{d}s\right]^{1/(1-m)}, \tag{26}$$

where

$$\phi(\tau) = \exp\left\{\tau + \int_0^\tau \zeta(s)\mathrm{d}s\right\}, \tag{27}$$

and $x_0$ is the initial value of $x$. As it stands, the analytic solution is too unwieldy to be useful, even for basic prescriptions for $\zeta$, but it does illustrate clearly the competition between instability and stochasticity. These manifest as the two terms in the exponent of $\phi(\tau)$. The first describes the deterministic exponential runaway, while the second stochastic term potentially impedes this tendency. In fact, if $\int_0^t \zeta(s)\mathrm{d}s$ behaves like a random walk, then its standard deviation will be proportional to $\sqrt{\tau}$, and so on short-to-intermediate times the second stochastic term can outcompete the first instability term. On longer times, however, $\tau$ will always defeat $\sqrt{\tau}$ and the system will approach the unstable laminar solution, given by

$$\frac{x}{x_{\mathrm{eq}}} = \left\{\left[\left(\frac{x_0}{x_{\mathrm{eq}}}\right)^{1-m} - 1\right]\mathrm{e}^{(1-m)\tau} + 1\right\}^{1/(1-m)}. \tag{28}$$

### 5.1 Geometric Brownian motion

Before presenting an analysis of the full equation (25), it is worthwhile examining the simpler case of $\Theta = 0$ and $q = 1$. The resulting system isolates cleanly all the main characteristics of more realistic
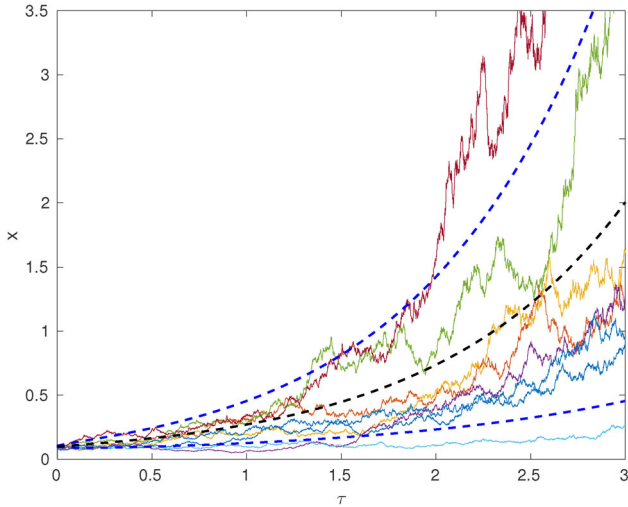
**Figure 9.** The solid curves are sample trajectories of GBM with $x_0 = 0.1$ and $\sigma = 0.5$. The dashed black line is the laminar path and the blue dashed curves are the 10th and 90th percentiles.

systems – an unstable fixed point ($x = 0$ now) and stochastic noise – while being analytically tractable. Equation (25) becomes

$$\frac{dx}{d\tau} = [1 + \zeta(\tau)] x \tag{29}$$

with initial condition $x_0 > 0$. For smooth random $\zeta$, the solution is

$$x(\tau) = x_0 \phi(\tau). \tag{30}$$

When studying stochastic dynamical systems, white noise is a convenient choice for modelling the variability. For white noise to be a good approximation, the fluctuation time-scales of the system should be much less than the characteristic time of interest. For MRI-driven turbulence, this choice is not ideal given that the spectrum of $\zeta$ has preferential frequencies. However, we use this as our starting point as it makes a number of results especially clear. With this choice of $\zeta$, equation (29) must be written in differential form

$$dx = x\,d\tau + \sigma x\,dW, \tag{31}$$

where $\sigma$ is the volatility coefficient (or noise amplitude) and $dW$ is white noise. Here, for simplicity, we have interpreted the calculus in the Ito sense. Equation (31) actually describes GBM and is frequently used in financial modelling. Its solution is

$$x(\tau) = x_0 \exp\left\{ \left[1 - \frac{\sigma^2}{2}\right] \tau + \sigma W(\tau) \right\}. \tag{32}$$

Given that $x_0 > 0$, the solution remains strictly positive as a result of the multiplicative form of the noise. We plot sample trajectories in Fig. 9 along with the 10th and 90th percentiles. When calculating these trajectories, we use the Euler–Maruyama method (Kloeden & Platen 1992). A feature of this collection of sample paths is the wide variation between them, an attribute that is shared with the simulations shown in Fig. 5. They are also non-monotonic; fluctuations 'kick' the system towards or away from the fixed point. The light blue curve is particularly striking, exhibiting a trajectory that remains close to equilibrium, $x(\tau) < 0.25$, up to time $\tau \approx 3$. Note that at $\tau = 3$, a purely deterministic model would have predicted $x$ to be $\approx 2$, an order of magnitude greater. The stochastic term in equation (32) has 'balanced out' the deterministic drift, at least on these shorter times.

The reader may note that when $\sigma > \sqrt{2}$, the stability of $x = 0$ switches. It becomes an attractor, and the system is stabilized. We stress, however, this effect is an artefact of multiplicative white noise in combination with the Ito calculus, and is not to be expected in real turbulent systems. For instance, the stabilization vanishes in the Stratonovich calculus and/or with noise models with memory and which are not multiplicative. We certainly do not expect MHD turbulence to exhibit the combination of special features that leads to this stabilization. Indeed, it makes little physical sense that 'shaking' an unstable system more vigorously ultimately leads to zero fluctuations. Moreover, the required amplitude of the fluctuations must be extremely large, in our case this would require negative $\alpha$ which is impossible. It is worth pointing out that the autoregressive stochastic model employed by Janiuk & Misra (2012) shares the same stabilizing property as white noise in the Ito calculus; consequently, we view their stabilization of thermal instability as an artefact of their model and not representative of a real fluctuating disc system.

The probability distribution of the solution trajectories may be obtained by solving the associated Fokker–Planck equation. If $f(x, \tau)dx$ is the probability of finding a path lying between $x$ and $x + dx$ at time $\tau$, then

$$f(\tau, x) = \frac{1}{\sigma x \sqrt{2\pi\tau}} \exp\left\{ -\frac{(\log(x/x_0) - \tau)^2}{2\sigma^2\tau} \right\}. \tag{33}$$

The expectation, $\mathbb{E}$, and variance, Var, can also be calculated as

$$\mathbb{E}(x) = x_0 e^\tau, \tag{34}$$

$$\text{Var}(x) = x_0^2 e^{2\tau} \left( e^{\sigma^2\tau} - 1 \right). \tag{35}$$

The mean is independent of the volatility and hence agrees with the laminar model. However, the variance contains a $(e^{\sigma^2\tau} - 1)$ factor which grows exponentially. This means that as time progresses, the expectation value grows less and less meaningful because the distribution becomes increasingly wide and flat. Overall, trajectories move away from the unstable equilibrium point, but individual trajectories can deviate from the laminar model dramatically.

Next, we turn to the statistics of the escape time (or 'last hitting time'), which can be defined as follows:

$$t_{\text{esc}} = \max\{\tau \geq 0 : x(\tau) = a\}, \tag{36}$$

where $a > x_{\text{eq}} = 0$. This gives us the last time that a sample path is within the interval $[0, a]$. Thus, $t_{\text{esc}}$ provides a measure for the effective instability time-scale, more accurate than the inverse of the laminar growth rate. If the system was laminar ($\zeta = 0$), however, the escape time would be simply $t_{\text{esc}}^{\text{lam}} = \log(a/x_0)$.

For GBM, it is possible to derive the probability distribution of $t_{\text{esc}}$ analytically (Kennedy 2010; Profeta, Roynette & Yor 2010). Because this is not a standard calculation, we go through its details in the appendix. Denoted by $g(\xi)$ where $\xi = t_{\text{esc}}/t_{\text{esc}}^{\text{lam}}$, it is a modified form of the Rayleigh distribution:

$$g(\xi) = \frac{1}{\sqrt{2\pi d^2 \xi}} \exp\left\{ -\frac{1}{2d^2} \left( \xi^{1/2} - \xi^{-1/2} \right)^2 \right\}, \tag{37}$$

where the parameters $\sigma$, $x_0$ and $a$ have conveniently combined into $d = \sigma \log(a/x_0)^{-1/2}$. In Fig. 10, we plot $g(\xi)$ for a range of noise amplitudes.

The mean and variance of $g$ can be calculated analytically

$$\mathbb{E}(\xi) = 1 + d^2 \tag{38}$$
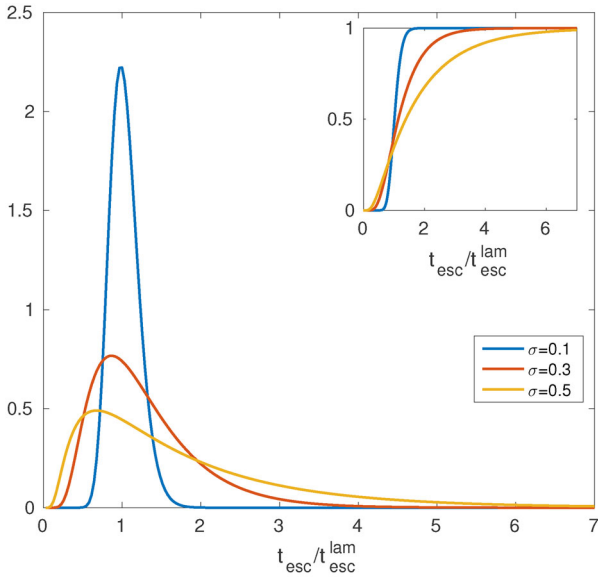
$$\text{Var}(\xi) = d^2(1 + 2d^2). \tag{39}$$

**Figure 10.** Probability distribution function of the escape times for GBM with $x_0 = 0.1$ and $a = 1.5$.

It is clear that the mean of the distribution does not depart greatly from the laminar escape time. Though the variance does increase with the volatility of the noise, the spread is not especially dramatic. Perhaps a more illuminating quantity is the kurtosis:

$$\text{Kurt}(\xi) = 3 \frac{20d^2 + 9d^2 + 1}{(2d^2 + 1)^2}, \qquad (40)$$

which varies from 3, when $d = 0$, to 9, when $d = 1$. A Gaussian possesses a kurtosis of 3, so our escape time distribution can in fact be exceptionally 'fat-tailed', or 'leptokurtic', meaning it generates a significant number of outliers. Fig. 10 illustrates this point, with larger $\sigma$ giving rise to very skewed and broad distributions. Inserted in the upper right of the figure, we also plot the cumulative distribution for various values of $d$. For $d = 1$, in particular, we see that only 30 per cent of systems possess an escape time equal to or less than the laminar escape time ($\xi = 1$), while over 10 per cent of systems possess an escape time of four times or more the laminar escape time. In fact, the probability that a trajectory possesses a $t_{esc}/t_{esc}^{lam}$ greater than some value $\xi_0$ may be approximated by the expression

$$\mathbb{P}(\xi > \xi_0) \approx \frac{d}{\sqrt{2\pi\xi_0}} \exp\left( \frac{2}{d^2} - \frac{\xi_0}{2d^2} \right), $$

for large $\xi_0$. When $d = 1$ and $\xi_0 = 10$, the probability is only about 1 per cent. However, this rises to 10 per cent when $d = 2$, indicating the strongly non-linear dependence of the statistics on noise amplitude. In summary, through the statistics of the escape time, one can observe stochasticity impeding instability over some initial period of the evolution.

### 5.2 Random logistic equation

We return to the more general problem by reintroducing the cooling term with $\Theta = 1$

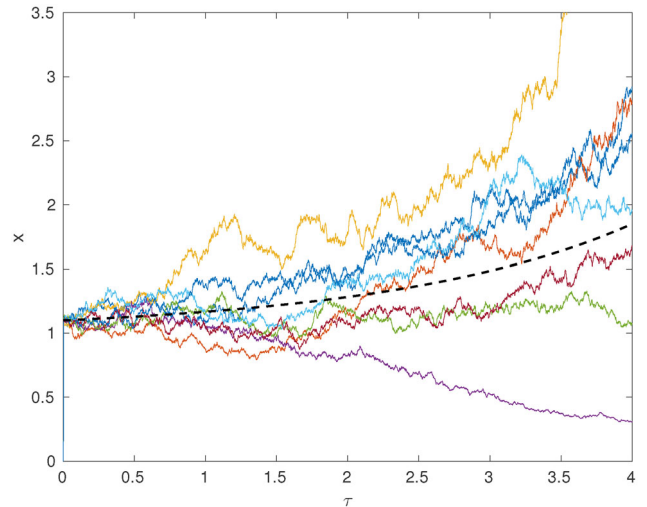$$\frac{dx}{d\tau} = [1 + \zeta(\tau)] x^q - x^m. \qquad (41)$$



**Figure 11.** Trajectories for the logistic model, $\sigma = 0.2$ with $x_0 = 1.1$.

Again assuming white noise for $\zeta$ and setting $q = 1$, we rewrite the ODE in differential form

$$dx = (x - x^m)\, d\tau + \sigma x \, dW. \qquad (42)$$

The equation was solved numerically, and sample trajectories are shown in Fig. 11. With a fixed initial condition, the system can either undergo runaway heating or cooling. Note that the initial perturbation does not give a good indication of which direction the sample paths eventually go. The trajectories in Fig. 11 qualitatively resemble those in Fig. 5 from the MHD simulations. Both show trajectories that remain close to their equilibrium values for multiple thermal time-scales along with trajectories that escape from the equilibrium faster than the laminar model. One substantial difference exists however; Fig. 11 shows no indication that the trajectories will be in general slower than the laminar model as is the case in Fig. 5.

Unlike GBM, it is difficult to analytically calculate the probability density function (PDF), and instead this is accomplished by solving the Fokker–Planck equation numerically

$$\frac{\partial f(x, \tau)}{\partial \tau} = -\frac{\partial}{\partial x} \left[ (x - x^m)\, f(x, \tau) \right] + \frac{\sigma^2}{2} \frac{\partial^2}{\partial x^2} \left[ x^2 f(x, \tau) \right]. \quad (43)$$

In order to do this, an approximation for the initial distribution must be made. Rather than using a Dirac–$\delta$ function as the initial condition, it is necessary to use a somewhat smoother function, a Gaussian distribution with variance 1000. An example solution for the probability distribution $f$ is plotted in Fig. 12. As earlier, the distribution becomes increasingly wide and flat as time progresses, indicating the broad range in possible evolutionary paths.

With the introduction of cooling comes the chance of a sample path going to zero; hence, our escape time definition must be modified to include a lower boundary

$$t_{esc} = \sup \{\tau \geq 0 : (x = a) \text{ or } (x = b)\}, \qquad (44)$$

where $a > x_{eq} > b$. For a given sample path, the escape time is the last instance when the solution is within the interval $[a, b]$, containing $x_{eq}$. In Fig. 13, we show the estimated PDF $g$ of $t_{esc}$ for a range of $\sigma$. These are calculated from 100 000 sample paths for each $\sigma$. As expected, for small $\sigma$ we approach the laminar escape time, while as $\sigma$ increases, $g$ undergoes asymmetric broadening, shifting its maximum to lower $t_{esc}$ and developing a tail at long $t_{esc}$. Qualitatively, this behaviour mirrors that shown for GBM, Fig. 10.
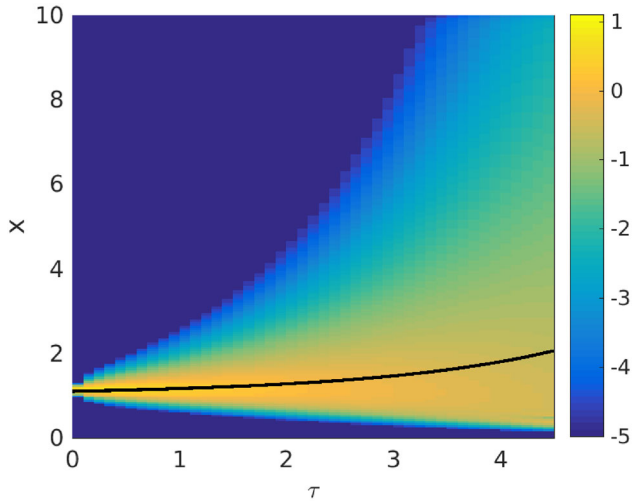
**Figure 12.** Time evolution of the logarithm of the PDF $f(\tau, x)$ for the logistic equation with $x_0 = 1.1$, $\sigma = 0.1$.
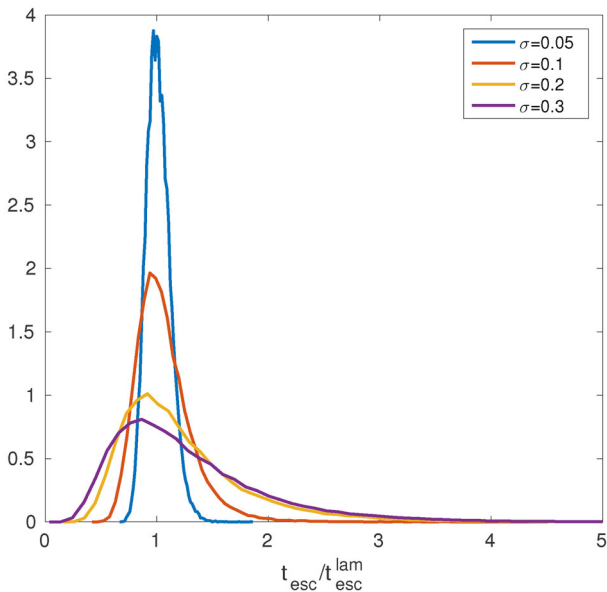


**Figure 13.** Probability distribution of the escape time $g$ for the logistic model, with $x_0 = 1.4$ with $a = 4$ and $b = 0.25$.

And so our conclusions for GBM carry over to the more realistic logistic case.

### 5.2.1 Variable q and m

In equation (25), $q$ and $m$ are free parameters which can be chosen to fit the system of interest. This flexibility allows it to approximately describe a range of different scenarios. For instance, the classical Shakura & Sunyaev disc (Shakura & Sunyaev 1973) can be represented with $q = 2$ and $m = 1$. Alternatively, $q$ and $m$ could be chosen so as to model the vertically stratified, radiation-pressure-dominated simulations of Jiang et al. (2013) and Jiang, Stone & Davis (2016). In this case, a large surface density disc yields $q = 1.6$ and $m = 0.98$, while a less dense disc gives $q = 1.9$ and $m = 0.9$. Finally, when the temperature of the gas ensures the opacity is influenced by the 'iron bump', the scaling of cooling with central pressure is found to greatly exceed that of the classical Shukura & Sunyaev model
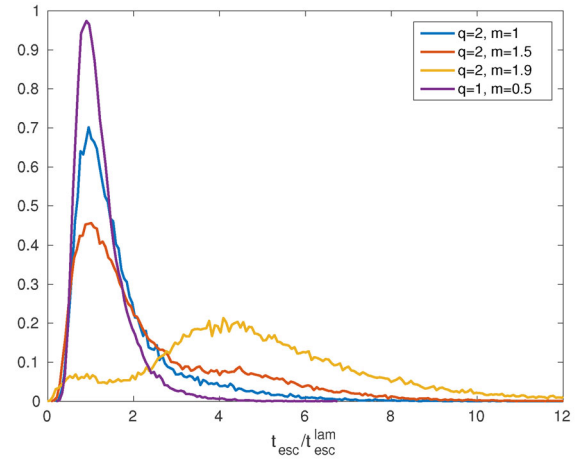


**Figure 14.** Probability distribution of the escape time $g$ for the logistic model with various $q$ and $m$ but keeping fixed: $x_0 = 1.4$, $\sigma = 0.2$, $a = 4$ and $b = 0.25$.
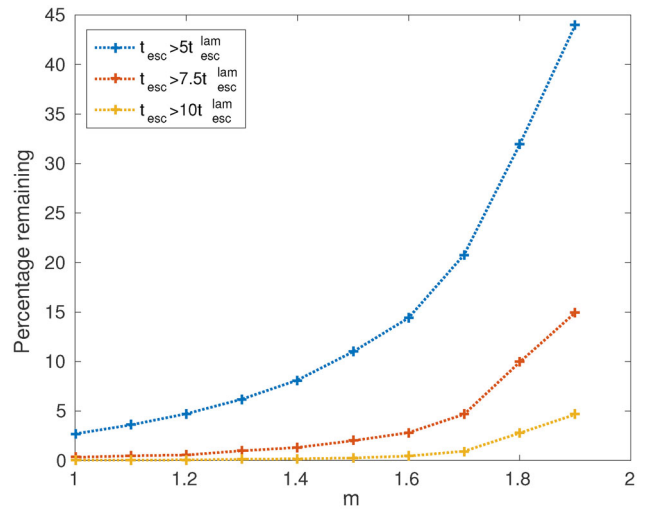


**Figure 15.** The probability that $t_{esc}$ is longer than a multiple of $t_{esc}^{lam}$ as a function of $m$ for fixed $q = 2$.

and $m = 1.89$ (Jiang et al. 2016). In Fig. 14, we plot the probability distribution function of the escape time $g$ for various choices of $q$ and $m$. As $m$ approaches $q$ from below, the escape time distribution becomes increasingly elongated to large $t_{esc}/t_{esc}^{lam}$. The deterministic component weakens and the stochastic drift becomes more important, dominating the results.

To better quantify this effect, we compute the probability that a given trajectory possesses a $t_{esc}$ greater than various multiples of $t_{esc}^{lam}$ and plot these probabilities as a function of $m$ for fixed $q = 2$. These results appear in Fig. 15. The curves show that as $m$ approaches $q$, the escape time can be significantly enhanced. For $m = 1.9$, there is approximately 45 per cent chance that it is five times the laminar prediction, and 5 per cent chance that it is 10 times greater. It should be noted that as $m$ approaches $q$, the laminar instability time-scale itself can be significantly longer than the thermal time $(\alpha\Omega)^{-1}$, which further separates the expected turbulent $t_{esc}$ from the thermal time. The stability uncovered in Jiang et al. (2016), on the time-scales of their simulations, can be easily explained by the enhanced delay witnessed in such marginally unstable systems.
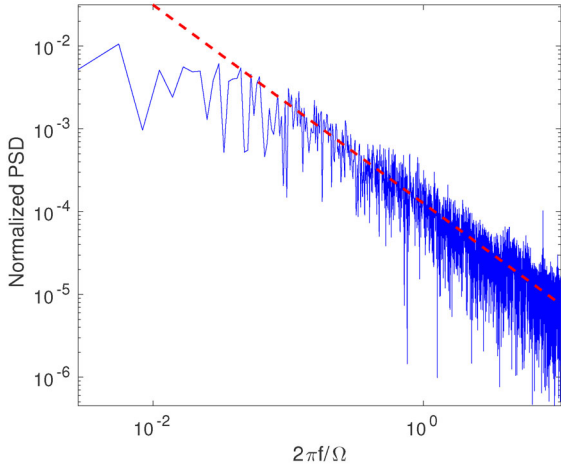
**Figure 16.** Time-averaged PSD normalized by $P_0$ from simulation R1. Slope $\propto f^{-1.2}$.

### 5.3 MHD power spectral density

The next step is to replace the white noise with the power spectral density (PSD) of $\Pi_{xy}$ calculated from the MHD simulations of Section 4. In Fig. 16, we show the PSD of $\Pi_{xy}$ from simulation R1. We then represent the fluctuating term in equation (25) by

$$\zeta(\tau) = \frac{\sigma}{N} \Sigma_{i=1}^{200} \alpha_i(f_i) \cos\left(2\pi f_i + \phi_i\right), \tag{45}$$

$$N = \left(\Sigma_{i=1}^{200} \alpha_i(f_i)^2\right)^{1/2}, \tag{46}$$

where $f_i$ are the frequencies, logarithmically spaced between $[0.05, 750]$, and $\phi_i$ are phase shifts, chosen randomly from a uniform distribution on the interval $[0, 2\pi]$. Finally, the constant amplitudes $\alpha_i(f_i)$ are calculated from a two-slope power law that fits our MHD simulations (constant for $2\pi f/\Omega < 0.1$ and $\propto f^{-1.2}$ otherwise), which is then multiplied by a random number generated from a uniform distribution on the interval $[0, 1]$.

To improve the comparison with the simulations R2a–R2h, we set $q = 0.5$ and $m = 0.25$ and thus our model equation is

$$\frac{dx}{d\tau} = [1 + \zeta(\tau)] x^{0.5} - x^{0.25}, \tag{47}$$

where $\zeta(\tau)$ is defined in equation (45). This equation is evolved forward in time 50 000 times for each choice of $\sigma$ in order to derive adequate statistics, especially for the escape time.

We show the resulting escape time distribution in Fig. 17. We find similar behaviour to both the GBM and the random logistic equation. As $\sigma$ increases, the distribution broadens and its tail at large $t_{esc}$ increases. We estimate that $\sigma \sim 0.015$ gives approximately the correct fluctuation amplitude when compared to the stress in Fig. 2. For this choice, Fig. 17 shows a clear tail reaching $t_{esc}/t_{esc}^{lam} \sim 2.5$–3.

We conclude that fat-tailed distributions are a generic feature of the escape time in turbulent but thermally unstable systems. Such systems produce a broad range of outcomes, and instability can be delayed for several instability time-scales. Being fat-tailed they also exhibit significant outliers – systems that 'hang around' for surprisingly long times before wandering away.

## 6 DISCUSSION AND CONCLUSIONS

We have performed a set of idealized shearing box simulations of the MRI in order to explore the effects of turbulent variability on
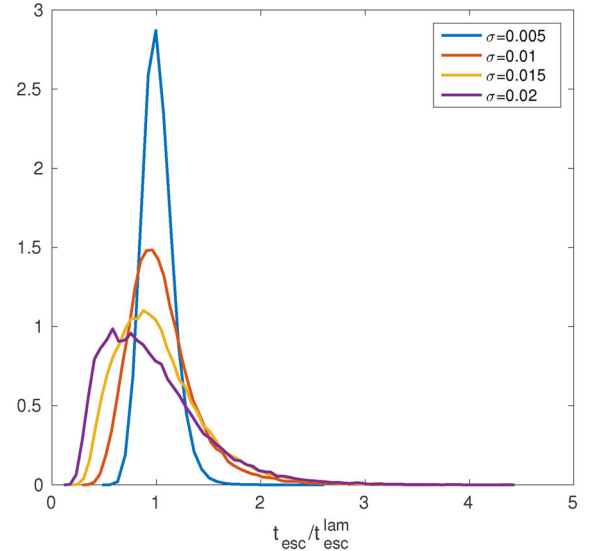


**Figure 17.** Escape time distribution from our PSD model with $q = 0.5$ and $m = 0.25$.

thermal instability. Our main aim was to check if turbulence interferes with the thermal runaway predicted by laminar theory. In our simulations, heating comes from numerical dissipation while a power-law cooling imitates optically thin radiative cooling. Relatively large computational domains are used in order to ensure a strong dependence of stress on pressure.

Simulations with an expected stable thermal equilibrium are found to fluctuate around their fixed points. When the cooling power-law exponent is decreased, and the laminar analysis predicts instability, our simulations indeed show thermal runaways. However, the system trajectories deviate from the laminar theory. The turbulence can stall the onset of instability in a large number of cases, for multiple thermal times. To better account for the instability time-scale, we introduce the concept of the escape time which we define to be the last time the system leaves an interval encapsulating the equilibrium in phase space. Our simulations show a large range of escape times ranging from $\sim 1/4$ to $\sim 5$ times the laminar thermal time-scale, for relatively narrow intervals.

Further reducing the cooling power-law exponent results in disc fragmentation. This is due to localized imbalances between heating and cooling; the instability time-scale is shorter than the mixing time-scale and hence distant pockets of fluid evolve independently. Very rough estimates indicate that thermal fragmentation on lengthscales larger than $H$ are at best marginally possible in the inner regions of X-ray binaries.

To better understand our results, we construct a probabilistic theory centred on simple stochastic equations that approximate the box-averaged thermal equation. These present us with a much larger sample of possible system outcomes to analyse than the MHD simulations can afford. First, we analyse GBM, which possesses the main ingredients of interest (an unstable fixed point and stochastic variations), while remaining analytically tractable. The distribution function of $t_{esc}$ exhibits a variance proportional to the square of the noise amplitude and considerable kurtosis. In general, the distribution is 'fat-tailed', permitting many instances of delayed thermal instability, and outliers for whom the escape time can be $\sim 10$ thermal times. Our second model introduces a logistic term to incorporate a power-law cooling and a third model replaces the white noise with the PSD of the stress from a thermally stable shearing box

simulation. In both cases, we obtain qualitatively similar behaviour to before, which instils confidence that its behaviour is generic to systems sustaining noise and instability. GBM may be thought of as the model equation for such systems.

In our GBM model, very large amplitude noise can stabilize the thermal instability, but we believe this is physically implausible. In fact, this behaviour arises from the special combination of Ito calculus and multiplicative white noise. It is not generic. While GBM is a convenient model for 'unresolved' turbulence, this dynamical peculiarity must be discarded when applying results to real systems. The stochastic process employed by Janiuk & Misra (2012) shares the same properties and thus suffers the same dynamical artefact. The stabilization witnessed in their simulations we hence view as unphysical.

The bulk of X-ray binary observations show no indication of limit cycles that could correspond to radiation-pressure-induced thermal instability (Gierliński & Done 2004). Notable exceptions are GRS 1915+105 (Done et al. 2004) and HLX-1 (Sun et al. 2016). We show that it is unlikely that turbulent fluctuations alone stabilize these discs. However, turbulence can delay and weaken instability. This weakening might be significant in combination with additional stabilizing mechanisms, for example: energy lost in the disc corona or by outflows, magnetic buoyancy effects, alteration of the pressure–stress relationship by strong magnetic fields, or opacity shifts near the iron bump (Svensson & Zdziarski 1994; Jiang & Stone 2016; Sadowski 2016). Finally, separate and important physical effects may arise from the global nature of the flow, especially from accretion. The viscous time-scale is probably longer than a typical $t_{\rm esc}$, as measured in this paper. But if further delayed by additional physics, $t_{\rm esc}$ could in some circumstances approach the accretion time. If this occurs, then the classic limit cycle behaviour expected from thermal instability could well be impeded, and/or pushed to smaller radii.

## ACKNOWLEDGEMENTS

## REFERENCES

Balbus S. A., Hawley J. F., 1991, ApJ, 376, 214
Belloni T., Méndez M., King A. R., van der Klis M., van Paradijs J., 1997, ApJ, 488, L109
Ciesielski A., Wielgus M., Kluźniak W., Sądowski A., Abramowicz M., Lasota J.-P., Rebusco P., 2012, A&A, 538, A148
De Swart H. E., Grasman J., 1987, Tellus A, 39, 10
Done C., Wardziński G., Gierliński M., 2004, MNRAS, 349, 393
Fromang S., Papaloizou J. C. B., 2007, A&A, 476, 1113
Fromang S., Hennebelle P., Teyssier R., 2006, A&A, 457, 371
Gierliński M., Done C., 2004, MNRAS, 347, 885
Goldreich P., Lynden-Bell D., 1965, MNRAS, 130, 125
Guilet J., Ogilvie G. I., 2012, MNRAS, 424, 2097
Guilet J., Ogilvie G. I., 2013, MNRAS, 430, 822
Hawley J. F., Gammie C. F., Balbus S. A., 1995, ApJ, 440, 742
Hirose S., Krolik J. H., Blaes O., 2009, ApJ, 691, 16
Janiuk A., Misra R., 2012, A&A, 540, A114
Jiang Y. F., Stone S. M., Davis S. W., 2013, ApJ, 778, 65
Jiang Y. F., Stone S. M., Davis S. W., 2016, ApJ, 778, 10
Kennedy D., 2010, Stochastic Financial Models. Chapman & Hall/CRC, London
King A. R., Ritter H. R., 1998, MNRAS, 293, L42
Kloeden P. E., Platen E., 1992, Numerical Solution of Stochastic Differential Equations. Springer-Verlag, Berlin
Latter H. N., Papaloizou J. C. B., 2012, MNRAS, 426, 1107
Lesur G., Ogilvie G. I., 2008, A&A, 488, 451
Levins R., 1969, Proc. Natl. Acad. Sci. USA, 62, 1061
Lightman A. P., Eardley D. M., 1974, ApJ, 187, 13
Lin D.-B., Gu W.-M., Lu J.-F., 2011, MNRAS, 415, 2319
Majda A. J., Timofeyev I., vanden- Eijnden E., 1999, PNAS, 96, 14687
Majda A. J., Timofeyev I., Vanden-Eijinden E., 2003, J. Atmos. Sci., 60, 1705
May R. M., 1973, Stability and Complexity in Model Ecosystems. Princeton Univ. Press, Princeton, NJ
Mishra B., Fragile P. C., Johnson L. C., Kluźniak W., 2016, MNRAS, 463, 3437
Miyoshi T., Kusano K., 2005, J. Comput. Phys., 208, 315
Piran T., 1978, ApJ, 221, 652
Profeta C., Roynette B., Yor M., 2010, Option Prices as Probabilities: A New Look at Generalized Black-Scholes Formulae. Springer-Verlag, Berlin
Ross J., Latter H. N., Guilet J., 2016, MNRAS, 455, 1, 526 (RLG16)
Ryan B. R., Gammie C. F., Fromang S., Kestener P., 2017, ApJ, preprint (arXiv:1702.00777)
Sadowski A., 2016, MNRAS, 459, 4397
Sano T., Inutsuka S., Turner N. J., Stone J. M., 2004, ApJ, 605, 321
Shakura N. I., Sunyaev R. A., 1973, A&A, 24, 337
Shakura N. I., Sunyaev R. A., 1976, MNRAS, 175, 613
Smak J., 1984, PASP, 96, 5
Stone J., Gardiner T., 2010, ApJS, 189, 142
Sun M., Gu W.-M., Yan Z., Wu Q., Liu T., 2016, MNRAS, 463, L99
Suresh A., 2000, SIAM J. Sci. Comput., 22, 1184
Svensson R., Zdziarski A. A., 1994, ApJ, 436, 599
Teyssier R., 2002, A&A, 364, 337
Warner B., 1995, Catacylsmic Variables. Cambridge Univ. Press, Cambridge
Wu Q. et al., 2016, ApJ, 833, 79

## APPENDIX A: ESCAPE TIME DISTRIBUTION FOR GBM

Following Kennedy (2010), we derive the PDF for the escape time, or last hitting time, for GBM. To achieve this, we calculate the PDF of the last hitting time of Brownian motion with drift and then perform a change of variables to obtain the last hitting time for GBM.

In the appendix, $W_\tau$ will refer to a standard Brownian motion, the amplitude of the fluctuations can be represented by a pre-factor of $\sigma$, which we refer to as the 'volatility'. To begin, consider the *first* hitting time of a random function $x(\tau)$, denoted by $M_a$ and defined by

$$M_a = \inf\{\tau \geq 0 : x(\tau) = a\}. \tag{A1}$$

(Here, the infimum can be thought of as the minimum.) Physically, this represents the first time the function $x$ reaches the value $a$. In the special case of $x$ corresponding to standard Brownian motion with drift, $x(t) = W_\tau + \epsilon\tau$, the PDF, $f_{M_a}$, of $M_a$ is

$$f_{M_a}(\xi) = \frac{a}{\sqrt{2\pi\xi^3}} \exp\left\{-\frac{1}{2}\left(\epsilon\sqrt{\xi} - a/\sqrt{\xi}\right)^2\right\}. \tag{A2}$$

See Kennedy (2010) for a derivation. With this result, we can then determine the *last* hitting time of GBM with drift.

Let $T_{-\mu\tau+b}$ denote the last time that the standard Brownian motion hits the line $(-\mu\tau + b \geq 0)$, for some constant $b$. Then, the PDF of $T_{-\mu\tau+b}$ is given by

$$f_{T_{-\mu\tau+b}}(\xi) = \frac{\mu}{\sqrt{2\pi\xi}} \exp\left\{-\frac{1}{2}\left(\mu\sqrt{\xi} - b/\sqrt{\xi}\right)^2\right\}. \tag{A3}$$

To see why this is true, consider the probability that $T_{-\mu\tau+b} > \xi$, for some $\xi$:

$$\mathbb{P}\left(T_{-\mu\tau+b} > \xi\right) = \mathbb{P}\left(W_s = b - \mu s, \text{ for some } s > \xi\right)$$
$$= \mathbb{P}\left(sW_{1/s} = b - \mu s, \text{ for some } s > \xi\right)$$
$$= \mathbb{P}\left(W_u = bu - \mu, \text{ for some } u < \frac{1}{\xi}\right)$$
$$= \mathbb{P}\left(M_{-\mu} < \frac{1}{\xi}\right).$$

In the second equality, we have used the fact that $\{sW_{1/s}\}$ is also a Brownian motion. The probability distribution function of $T_{-\mu\tau+b}$ is given by the $\xi$ derivative of the above expression. By considering the probability distribution function of $M_{-\mu}$ where the drift is $-b$ and the threshold $-\mu$, we conclude that

$$f_{T_{-\mu\tau+b}}(\xi) = \frac{1}{\xi^2} f_{M_{-\mu}}\left(\frac{1}{\xi}\right) \tag{A4}$$

from which the result follows. Notice that $f_{T_{-\mu\tau+b}}$ is identical to the PDF of the last time Brownian motion with drift $\mu$ hits the line $b$.

Let us now finally define the last hitting time, or escape time, for a random function $x(\tau)$:

$$T_a = \sup\{\tau \geq 0 : x(\tau) = a\}, \tag{A5}$$

for some threshold $a$. If $x$ corresponds to Brownian motion with drift $x(\tau) = \nu\tau + W_\tau$, for constant $\nu > 0$, then the probability distribution of $T_a$ is

$$f_{T_a}(\xi) = \frac{\nu}{\sqrt{2\pi\xi}} \exp\left\{-\frac{1}{2}\left(\nu\sqrt{\xi} - a/\sqrt{\xi}\right)^2\right\}. \tag{A6}$$

If however, $x(\tau)$ is a GBM

$$x(\tau) = x_0 e^{\sigma W + \mu\tau}, \tag{A7}$$

and $x_0$ a constant, then the distribution of $T_b$, for some constant threshold $b$, is easy to obtain from equation (A6). Consider the probability for GBM that $T_a < \xi$, for some $\xi$. This corresponds to

$$\mathbb{P}(T_a \leq \xi) = \mathbb{P}(x(\tau) \geq b, \forall \tau \geq \xi)$$
$$= \mathbb{P}\left(\mu\tau + \sigma W_\tau \geq \log\left(\frac{b}{x_0}\right), \forall \tau \geq \xi\right)$$
$$= \mathbb{P}\left(\frac{\mu}{\sigma}\tau + W_\tau \geq \frac{1}{\sigma}\log\left(\frac{b}{x_0}\right), \forall \tau \geq \xi\right)$$

which is precisely the probability for Brownian motion with drift, but with $\nu = \frac{\mu}{\sigma}$ and $a = \frac{1}{\sigma}\log(b/x_0)$. This supplies us with an expression for the distribution of $T_a$ for GBM. A rescaling of $\tau$ and a renormalization obtains equation (37).

This paper has been typeset from a TeX/LaTeX file prepared by the author.