P G Ivanov, A A Schekochihin, W Dorland, A R Field,
F I Parra

Zonally dominated dynamics and Dimitsthreshold in curvature-driven ITG turbulence

Zonally dominated dynamics and Dimitsthreshold in curvature-driven ITG turbulence

P G Ivanov, A A Schekochihin, W Dorland, A R Field, F I Parra

Zonally dominated dynamics and Dimits threshold in curvature-driven ITG turbulence

Plamen G. Ivanov[1,2,3]†, A. A. Schekochihin[1,4], W. Dorland[1,5], A. R. Field[3], and F. I. Parra[1,6]

[1]Rudolf Peierls Centre for Theoretical Physics, University of Oxford, Oxford OX1 3PU, UK

[2]St John's College, Oxford OX1 3JP, UK

[3]EURATOM/UKAEA Fusion Association, Culham Science Centre, Abingdon, OX14 3DB, UK

[4]Merton College, Oxford OX1 4JD, UK

[5]Department of Physics, University of Maryland, College Park, Maryland 20740, USA

[6]Worcester College, Oxford OX1 2HB, UK

The saturated state of turbulence driven by the ion-temperature-gradient instability is investigated using a two-dimensional long-wavelength fluid model that describes the perturbed electrostatic potential and perturbed ion temperature in a magnetic field with constant curvature (a Z-pinch) and an equilibrium temperature gradient. Numerical simulations reveal a well-defined transition between a finite-amplitude saturated state dominated by strong zonal-flow and zonal-temperature perturbations, and a blow-up state that fails to saturate on a box-independent scale. We argue that this transition is equivalent to the Dimits transition from a low-transport to a high-transport state seen in gyrokinetic numerical simulations (Dimits *et al.* 2000). A quasi-static staircase-like structure of the temperature gradient intertwined with zonal flows, which have patchwise constant shear, emerges near the Dimits threshold. The turbulent heat flux in the low-collisionality near-marginal state is dominated by turbulent bursts, triggered by coherent long-lived structures closely resembling those found in gyrokinetic simulations with imposed equilibrium flow shear (van Wyk *et al.* 2016). The break up of the low-transport Dimits regime is linked to a competition between the two different sources of poloidal momentum in the system — the Reynolds stress and the advection of the diamagnetic flow by the $\boldsymbol{E} \times \boldsymbol{B}$ flow. By analysing the linear ITG modes, we obtain a semi-analytic model for the Dimits threshold at large collisionality.

1. Introduction

Understanding the heat transport properties of magnetically confined plasmas is crucial for the design of successful tokamak experiments. Since the characteristic correlation length of the turbulence is small compared to the size of the tokamak, one normally assumes that the local heat transport depends only on local conditions, such as density, temperature, magnetic field and their gradients (this view has been challenged; see Dif-Pradalier *et al.* 2010). Existing research suggests that the dominant contribution to the heat flux in tokamaks arises from turbulence driven by microinstabilities, the most prominent of which is the ion-temperature-gradient instability (Waltz 1988; Cowley *et al.* 1991; Kotschenreuther *et al.* 1995a). We use "ITG" and "ITG turbulence" as shorthand terms for this instability and the turbulence driven by it, respectively. As the name suggests, it is controlled by the gradient of ion temperature, which is a source of

† Email address for correspondence: plamen.ivanov@physics.ox.ac.uk

free energy for unstable microscale perturbations. It is then natural to investigate the dependence of the heat flux carried by the perturbations on the temperature gradient that drives those perturbations. Knowing the relationship between them, one can invert this relationship and find the heating power needed to support a given temperature gradient.

Strongly driven ITG turbulence, i.e., ITG turbulence with a temperature gradient far above the linear-instability threshold, is believed to saturate via a "critically balanced" turbulent cascade (Barnes *et al.* 2011): free energy stored in the equilibrium gradient is injected into perturbations by the instability and nonlinearly transferred (cascaded) to smaller scales, where it is thermalised via collisions. This is governed by the same kind of processes as the Kolmogorov cascade in hydrodynamic turbulence (Frisch 1995). This strongly turbulent saturated state supports vigorous turbulent transport of energy, so increasing the temperature gradient in such a system requires very substantial increases in heating power.

Naïvely, one expects strong turbulence and high levels of transport to set in as soon as the temperature gradient exceeds the linear-instability threshold. However, there is numerical evidence for a low-transport regime with low levels of turbulence at temperature gradients larger than the linear threshold for the ITG instability but smaller than some nonlinear threshold above which strong turbulence and a high-transport state set in (Dimits *et al.* 2000). Simulations have shown that the low-transport state below this threshold (to which we refer as the "Dimits state" and "Dimits threshold", respectively) is dominated by strong zonal flows (ZFs) — Larmor-scale shear flows in the poloidal direction. These help regulate turbulence by shearing heat-carrying perturbations and hence reducing their amplitude. In this paper, we attempt to explain how the Dimits state is maintained and what leads to its eventual collapse.

Despite being fairly well studied, many aspects of ZF physics, e.g., generation of ZFs from turbulence, stability of zonal fields, dependence of experimentally important quantities, like the heat flux, on basic plasma parameters (density, temperature, magnetic field and their gradients) in zonally dominated plasmas, remain far from being settled. One of the established paradigms is the primary-secondary-tertiary instability scenario (Rogers *et al.* 2000; Rogers & Dorland 2005). Let us outline it here. The primary ITG instability feeds energy into a spectrum of linearly unstable modes that become nonlinearly unstable to zonal perturbations: this is the "secondary instability". Saturation is reached when the energy injection into ZFs is balanced by their slow viscous damping. Increasing the temperature gradient increases the primary drive, hence the secondary drive, hence the amplitude of ZFs. However, ZFs of large enough amplitude become nonlinearly unstable to a "tertiary instability", so they break up, transferring energy back into the ITG modes. The suppression due to zonal shear having been lost, fully developed turbulence ensues. In this scenario, the Dimits threshold is given by the threshold of the tertiary instability.

Even though we show that the tertiary instability determines important properties of the saturated state (e.g., poloidal spectra), we find that the Dimits threshold is not directly determined by the tertiary instability. The latter only works to excite turbulent perturbations that coexist with the ZFs. The way these perturbations interact with the ZFs via a mechanism akin to a generalised nonlinear secondary instability is what determines the Dimits threshold. In the low-collisionality regime, the interactions between turbulent perturbations and ZFs give rise to predator-prey-like oscillations familiar from past studies of ZF physics (see, e.g., Diamond *et al.* 2005; Ricci *et al.* 2006; Kobayashi & Rogers 2012).

Recent progress has suggested that an entirely different scenario might need to be developed for turbulence with imposed background flow shear, applicable to tokamak

plasmas made to rotate differentially. The work by van Wyk *et al.* (2017) has shown that close to marginality, the effect of the self-generated zonal shear is negligible compared to the equilibrium flow shear. The heat flux in this near-marginal state is dominated not by space-filling turbulence, but by localised, long-time-coherent, soliton-like, finite temperature and density perturbations travelling through the plasma (van Wyk *et al.* 2016). We call these structures "ferdinons", after Ferdinand van Wyk's name. As the temperature-gradient drive is increased, the number of ferdinons increases, they begin overlapping and interacting strongly, and the system enters a fully developed turbulent state. We do not investigate the case of imposed background flow shear in this paper, but we do find that locally generated zonal flows arrange themselves in regions of nearly constant shear. Structures closely resembling ferdinons are seen drifting through these sheared regions. This suggests that the formation of localised structures is a robust feature of sheared ITG turbulence as they are seen both in our simplified model (described below and in Section 2) and in more realistic 3D GK simulations.

A comprehensive treatment of the problem of transition to, and saturation of, ITG turbulence requires the gyrokinetic (GK) framework in toroidal tokamak geometry (Frieman & Chen 1982; Sugama *et al.* 1996; Sugama & Horton 1997, 1998; Abel *et al.* 2013; Catto 2019). However, its complexity makes it both analytically and numerically hard to treat. In this paper, we attempt the more modest task of tackling the problem in a simplified model for the dynamical evolution of the perturbations of electrostatic potential (or, equally well, density) and ion temperature in a tokamak plasma. The model is derived as an exact asymptotic limit of the underlying gyrokinetic equations in a physically realisable, if not necessarily most general, regime (see Section 2.4). The approximations used are chosen to ensure that our model has a number of features that we consider essential: 1) a curvature-driven ion-temperature-gradient (ITG) instability, characteristic of tokamak plasmas; 2) an appropriate modified adiabatic electron response, which has been found to be crucial for capturing essential zonal-flow properties (e.g., the correct ITG secondary instability: see Hammett *et al.* 1993; Rogers *et al.* 2000); 3) it is a two-field model linking the perturbations of the electrostatic potential and the ion temperature, rather than a one-field drift-wave model of the Hasegawa & Mima (1978) variety. A two-field model allows us to capture the important ITG linear instability, while keeping the equations simple enough to allow for an analytic treatment.

As already mentioned, fully developed ITG turbulence is critically balanced and, therefore, 3D, so we cannot hope to capture that in a 2D model. Beyond the Dimits state, we find that our model fails to reach a saturated state — perturbations grow exponentially and the box-sized perturbations eventually dominate the spectrum, regardless of the size of the integration domain. Such behaviour is typical of strongly turbulent 2D systems, e.g., 2D hydrodynamics with its inverse energy cascade (Kraichnan 1967). However, below the Dimits threshold, our model has a perfectly well-defined 2D Dimits saturated state, dominated by ZF (see Section 3). Thus, our model allows us to identify the Dimits phase transition in an unambiguous and sharp way, as a transition from a regime with a finite saturated state to one without.

There are two ways for turbulence to achieve saturation — it can either cascade injected energy down to dissipation scales or, if it is internally driven by an instability, it can assemble itself in a configuration that suppresses that instability, i.e., the initial unstable equilibrium evolves into a new equilibrium with weaker instabilities. As our model contains both ZF and zonal-temperature perturbations, in principle it can accommodate the physics of two possible instability-suppression mechanisms: shearing of the turbulence by ZFs and modifying the background temperature gradient by zonal temperature perturbations in order to cancel the ITG drive. Neither of these can be done

uniformly across the entire domain because we impose periodic boundary conditions in the radial direction. Interestingly, we find that the zonal perturbations arrange themselves in alternating wide regions of nearly-constant zonal shear, strong enough to suppress turbulence, and narrow regions of strong zonal-temperature gradient, which flattens the background temperature gradient and quenches the ITG instability. The resulting "staircase"-like overall radial temperature profiles are reminiscent of those seen in global and local flux-driven gyrokinetic simulations (Dif-Pradalier *et al.* 2010, 2017; Villard *et al.* 2013; Rath *et al.* 2016) and reported in experimental data (Dif-Pradalier *et al.* 2015). The resulting turbulent heat flux is significantly suppressed. The stability, and hence existence, of this zonal state is controlled by the background temperature gradient — a large enough gradient renders the staircase configuration unstable and the system enters a fully developed turbulent state. In Section 4, we link this behaviour to the mechanism through which the turbulence feeds the ZF, viz., the turbulent flux of poloidal momentum. By considering the linearly unstable ITG modes, we find a semi-analytical prediction of the Dimits threshold at high collisionality and high temperature gradient.

The rest of the paper is organised as follows. In Section 2, we describe our model, whose detailed derivation is given in Appendix A. Section 3 describes the nonlinear saturated state and in particular the zonally dominated state near the Dimits threshold. In Section 4, we focus on the turbulent momentum flux of ITG turbulence subject to strong zonal shear, and the physics of the Dimits regime and its breakup beyond the Dimits threshold. Our results are summarised and conclusions are drawn in Section 5.

2. ITG-Driven Dynamics in a Z-pinch

We consider the local dynamics of the perturbations of electrostatic potential and ion temperature of a 2D plasma (in the plane perpendicular to the magnetic field) in a Z-pinch magnetic geometry with an equilibrium temperature gradient. Our equations are derived in a highly collisional, cold-ion asymptotic limit of the electrostatic ion gyrokinetic equation. Their detailed derivation can be found in Appendix A. Here we present a summary of these equations, their physical motivation and key properties. If the reader wishes to skip directly to Section 3, which contains the analysis of the saturated state, she may want to glance first at the model equations — these are (2.17) and (2.18).

2.1. *Magnetic Geometry*

The magnetic geometry of constant magnetic curvature is chosen because it is the simplest one that enables an ITG instability by coupling the electrostatic potential and the temperature perturbations via the magnetic drift. The integration domain is positioned in the magnetic field of a line of current (Z-pinch†) at radial distance L_B from the current line: see Figure 1. We define the x and y axes as pointing radially outwards and parallel to the current, respectively. We assume $L_B \gg L_x, L_y$, where L_x and L_y are the "radial" (x) and "poloidal" (y) sizes of the domain, respectively. Here we use the terms "radial" and "poloidal" to reflect the intended similarity of the domain to one positioned at the outboard midplane in a tokamak geometry. In that sense, we can think of the radial x coordinate as perpendicular to flux surfaces. These surfaces are parametrised by the poloidal y and field-parallel $\hat{\boldsymbol{b}}$ coordinates. Here $\boldsymbol{B} = B\hat{\boldsymbol{b}}$ is the magnetic field and the unit vectors $\{\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{b}}\}$ form a right-handed basis. In the 2D approximation employed here, all perturbed fields depend only on x and y. The magnetic

† This simplification as a route to a minimal model of ion-scale turbulence goes back at least to Ricci *et al.* (2006).
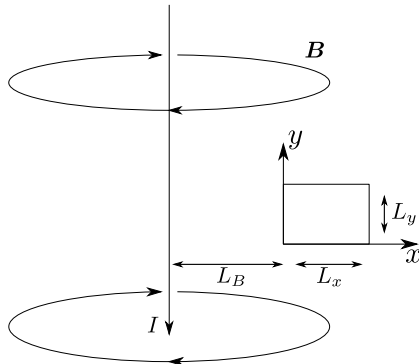
FIGURE 1. Illustration of the Z-pinch magnetic geometry.
.

field of the Z-pinch with total current I is azimuthal around the current line (as shown on Figure 1) and has magnitude $B(x) = 2I/cx$. The radial gradient of this field is then

$$\frac{1}{B}\frac{dB}{dx} = -\frac{1}{L_B}. \tag{2.1}$$

This value is constant across the domain to lowest order in $L_x/L_B \ll 1$. We define L_B to be the magnetic scale length. Similarly, we can define the ITG scale length

$$\frac{1}{L_T} \equiv -\frac{1}{T_i}\frac{dT_i}{dx}. \tag{2.2}$$

In a tokamak, L_B scales with the major radius of the device, while L_T scales with the minor radius. Here we will take the limit

$$L_B \gg L_T, \tag{2.3}$$

equivalent to a large-aspect-ratio approximation in a tokamak geometry. We do this in order to ensure that the magnetic drift in the density equation is of the appropriate order [see (2.9)]. This drift is essential for the linear curvature-driven ITG instability that we aim to capture.

2.2. *Electron Response*

The electron density is assumed to follow the modified adiabatic response

$$\frac{\delta n_e}{n_e} = \frac{e(\phi - \overline{\phi})}{T_e}, \tag{2.4}$$

taking into account the fast parallel streaming of the electrons within the flux surfaces of a tokamak, in the small-mass-ratio limit $m_e/m_i \to 0$ (Dorland & Hammett 1993; Hammett *et al.* 1993). Here δn_e and n_e are the perturbed and equilibrium electron density, respectively, ϕ is the electric potential, T_e is the electron temperature, and

$$\overline{\phi}(x) \equiv \frac{1}{L_y}\int dy\; \phi(x,y) \tag{2.5}$$

is the poloidal (zonal) spatial average of the perturbed electric potential ϕ. We refer to zonally averaged fields as being "zonal". In the 2D approximation, the turbulent fields (e.g., ϕ) are independent of z, hence we do not need to integrate over the z direction. The difference

$$\phi'(x,y) \equiv \phi(x,y) - \overline{\phi}(x) \tag{2.6}$$

is the "nonzonal" part of the field.

A cautious reader has spotted that there is no way to define flux surfaces in the magnetic geometry of a Z-pinch, as the magnetic field lines do not describe 2D surfaces, but rather close on themselves after one turn around the current axis. This problem can be fixed by demanding that the magnetic field be, in fact, sheared: $\boldsymbol{B} = B_0(\hat{\boldsymbol{z}} + x\hat{\boldsymbol{y}}/L_s)$, where L_s is the characteristic scale length of the magnetic shear. This is the magnetic field of a helimak (Gentle & He 2008). The field lines define cylindrically symmetric concentric flux surfaces and the electron parallel streaming mixes the azimuthal (z) and poloidal (y) directions. We can then take the limit $L_s \to \infty$ *after* performing the small-mass-ratio ($m_e/m_i \to 0$) expansion and eliminate magnetic shear from the ion equations, while retaining the flux-surface effect in the electron response.

2.3. *Cold-Ion Limit*

The cold-ion limit allows us to simplify the gyroaveraging operator that appears in gyrokinetics. Its corresponding Fourier-space operator is a multiplication by the Bessel function

$$J_0\left(\frac{k_\perp v_\perp}{\Omega_i}\right) = 1 - \frac{1}{4}k_\perp^2\rho_i^2\frac{v_\perp^2}{v_{ti}^2} + \mathcal{O}\left(k_\perp^4\rho_i^4\right). \tag{2.7}$$

The square of the ion gyroradius $\rho_i = v_{ti}m_i c/ZeB$ and the ion temperature $T_i = m_i v_{ti}^2/2$ are both proportional to the square of the ion thermal speed v_{ti} (here m_i and Z are the ion mass and charge in units of e, respectively). Thus, the cold-ion limit is equivalent to a long-wavelength expansion $\mathcal{O}(k_\perp\rho_i) \ll 1$ with a finite sound radius $\mathcal{O}(k_\perp\rho_s) \sim 1$, where $k_\perp$ is the perpendicular (to the magnetic field) wavenumber, $\rho_s = \rho_i/\sqrt{2\tau}$ is the sound radius, and $\tau = T_i/ZT_e$ is the temperature ratio (T_e is assumed finite). The sound radius ρ_s is the natural normalisation for the microphysical length scales in the problem: see equations (2.9) and (2.10) below and their derivation in Appendix A.

2.4. *Model Equations*

We take the density and temperature moments of the electrostatic ion gyrokinetic equation and adopt the high-collisionality, cold-ion, long-wavelength, large-aspect-ratio ordering

$$\frac{\partial_t}{\nu_i} \sim \tau \sim k_\perp^2\rho_i^2 \sim \frac{L_T}{L_B} \ll 1 \sim \frac{\varphi}{T}, \tag{2.8}$$

where $\varphi \equiv Z_i e\phi/T_i$ is the normalised electric potential, $T = \delta T/T_i$ is the normalised ion-temperature perturbation, and ν_i is the ion-ion collision frequency. The equations that we obtain in Appendix A are

$$\frac{\partial}{\partial t}\left(\tau\varphi' - \frac{1}{2}\rho_i^2\nabla_\perp^2\varphi\right) - \frac{\rho_i v_{ti}}{L_B}\frac{\partial}{\partial y}(\varphi + T) + \frac{\rho_i v_{ti}}{2L_T}\frac{\partial}{\partial y}\left(\frac{1}{2}\rho_i^2\nabla_\perp^2\varphi\right) \tag{2.9}$$

$$+ \frac{1}{2}\rho_i v_{ti}\left(\left\{\varphi, \tau\varphi' - \frac{1}{2}\rho_i^2\nabla_\perp^2\varphi\right\} + \frac{1}{2}\rho_i^2\boldsymbol{\nabla}_\perp\cdot\{\boldsymbol{\nabla}_\perp\varphi, T\}\right)$$

$$= -\frac{1}{2}\chi\rho_i^2\nabla_\perp^4(a\varphi - bT),$$

$$\frac{\partial T}{\partial t} + \frac{\rho_i v_{ti}}{2L_T}\frac{\partial\varphi}{\partial y} + \frac{1}{2}\rho_i v_{ti}\{\varphi, T\} = \chi\nabla_\perp^2 T, \tag{2.10}$$

where the Poisson bracket is defined by

$$\{f, g\} = \hat{\boldsymbol{b}}\cdot(\boldsymbol{\nabla}f \times \boldsymbol{\nabla}g) = \frac{\partial f}{\partial x}\frac{\partial g}{\partial y} - \frac{\partial f}{\partial y}\frac{\partial g}{\partial x}, \tag{2.11}$$

and

$$\chi \equiv \frac{8}{9}\sqrt{\frac{2}{\pi}}\nu_i \rho_i^2 \tag{2.12}$$

is the thermal diffusivity. The numerical factor in (2.12) and the constants $a = 9/40$, $b = 67/160$ in (2.9) are specific to the Landau collision operator (see Appendix A.5).

Let us discuss the physics content of (2.9) and (2.10). Equation (2.10) is the more obvious one — it describes the advection of the total temperature (perturbations plus equilibrium) by the $\boldsymbol{E} \times \boldsymbol{B}$ drift $\boldsymbol{V}_E = c\hat{\boldsymbol{b}} \times \boldsymbol{\nabla}\phi/B$, and the thermal diffusion perpendicular to the magnetic field. Indeed, (2.10) can be rewritten as

$$\frac{d}{dt}(\delta T + T_i) = \chi\nabla_\perp^2 \delta T, \tag{2.13}$$

where the advective time derivative is

$$\frac{d}{dt} \equiv \frac{\partial}{\partial t} + \boldsymbol{V}_E \cdot \nabla. \tag{2.14}$$

The advection of the equilbrium temperature, $\boldsymbol{V}_E \cdot \boldsymbol{\nabla} T_i$ [the second term on the left-hand side of (2.10)] is responsible for the injection of free energy (see Section 2.7), causing the ITG instability.

Equation (2.9) describes the time evolution of the sum of perturbed ion density and the vorticity of the $\boldsymbol{E} \times \boldsymbol{B}$ drift velocity:

$$\frac{\delta n_i}{n_i} + \frac{1}{n_i}\int d^3\boldsymbol{v} \, \left(\varphi - \langle\langle\varphi\rangle_R\rangle_r\right)F_i = \tau\varphi' - \frac{1}{2}\rho_i^2\nabla_\perp^2\varphi + \mathcal{O}\big(k_\perp^4\rho_i^4\varphi\big), \tag{2.15}$$

where n_i and δn_i are the equilibrium and perturbed ion densities, respectively [see (A 21)]. Thus, (2.9) can be thought of as both a perturbed-ion-density equation and as the curl of the perpendicular-momentum equation. The equality in (2.15) follows from the quasineutrality condition $Z\delta n_i = \delta n_e$, the electron response (2.4), the approximation (2.7), and the ordering (2.8).

The first term of (2.9) is the time derivative of (2.15). The second term, $(\rho_i v_{ti}/L_B)\partial_y(\varphi + T)$, is the magnetic drift (both curvature and ∇B) of pressure perturbations. This, or rather the $\partial_y T$ part of it, is essential for the curvature-driven ITG instability. It appears in the density equation because the magnetic drift creates charge separation, and hence electrostatic potential, which is then coupled to the perturbed density via quasineutrality. The third term, $(-\rho_i v_{ti}/4L_T)\partial_y\big(\rho_i^2\nabla_\perp^2\varphi\big)$, is a finite-Larmor-radius (FLR) term originating from the gyroaveraged $\boldsymbol{E} \times \boldsymbol{B}$ drift. It is the diamagnetic drift due to the equilibrium temperature gradient. The first of the nonlinear terms represents the advection of the quantity (2.15) by the $\boldsymbol{E} \times \boldsymbol{B}$ drift. The second nonlinear term $\boldsymbol{\nabla}_\perp \cdot \{\boldsymbol{\nabla}_\perp\varphi, T\}$ is another FLR effect, which describes the $\boldsymbol{E} \times \boldsymbol{B}$ advection of diamagnetic momentum. This term provides a crucial source of turbulent poloidal momentum flux that destabilises the ZF profiles, destroying the ZF-dominated Dimits regime. The nature of this term and its role in the Dimits transition are discussed in detail in Section 4.3. Note that the nonlinear terms in (2.9) and (2.10) are equivalent to the nonlinearities appearing in the model analysed by Rogers *et al.* (2000) in the limit (2.8). Finally, the collisional terms in (2.9) represent the viscous damping of the $\boldsymbol{E} \times \boldsymbol{B}$ flow and also couple the density and temperature perturbations. The latter coupling does not appear to be important for the results of this paper, but has been kept for consistency.

To prepare (2.9) and (2.10) for numerical analysis and distil important parameters, we introduce normalised variables and fields

$$\hat{t} \equiv \frac{2\rho_s \Omega_i}{L_B} t, \qquad \hat{x} \equiv \frac{x}{\rho_s}, \qquad \hat{y} \equiv \frac{y}{\rho_s},$$

$$\hat{\varphi} \equiv \frac{\tau L_B \varphi}{2\rho_s} = \frac{\tau L_B}{2\rho_s} \frac{Z_i e \phi}{T_i}, \qquad \hat{T} \equiv \frac{\tau L_B T}{2\rho_s} = \frac{\tau L_B}{2\rho_s} \frac{\delta T}{T_i}, \qquad (2.16)$$

$$\kappa_T \equiv \frac{\tau L_B}{2 L_T}, \qquad \hat{\chi} \equiv \frac{L_B}{2\rho_s} \frac{\chi}{\Omega_i \rho_s^2},$$

where $\Omega_i = v_{ti}/\rho_i$ is the ion gyrofrequency. Dropping hats and subscripts ($\boldsymbol{\nabla}_\perp \mapsto \boldsymbol{\nabla}$), we obtain from (2.9) and (2.10) the following equations in normalised units:

$$\partial_t \left(\varphi' - \nabla^2 \varphi\right) - \partial_y (\varphi + T) + \kappa_T \partial_y \nabla^2 \varphi + \{\varphi, \varphi' - \nabla^2 \varphi\} + \boldsymbol{\nabla} \cdot \{\boldsymbol{\nabla}\varphi, T\}$$
$$= -\chi \nabla^4 (a\varphi - bT), \qquad (2.17)$$
$$\partial_t T + \kappa_T \partial_y \varphi + \{\varphi, T\} = \chi \nabla^2 T. \qquad (2.18)$$

These equations have two independent parameters: the normalised equilibrium temperature gradient, κ_T, and the normalised collisionality, χ. There are two other parameters — L_x and L_y, the domain lengths in x and y — but any physically relevant results must be independent of these if our equations are indeed a valid local model of the plasma. This turns out to be true for the saturated Dimits state.

We solve (2.17) and (2.18) numerically in a doubly periodic box of size L_x and L_y using a pseudo-spectral algorithm. We integrate the linear terms implicitly in time, while the nonlinear terms are integrated explicitly using the Adams-Bashforth three-step method. This integration scheme is similar to the one implemented in the popular gyrokinetic code GS2 (Kotschenreuther *et al.* 1995*b*; Dorland *et al.* 2000).

2.5. *Relationship to Hasegawa-Mima Equation and Related Models*

It is easy to see that setting $\kappa_T = 0$ effectively decouples (2.18) from (2.17) — taking an initial condition $T(t = 0) = 0$ then leads to a trivial solution $T(t) = 0$. In that case, (2.17) reduces to

$$\partial_t \left(\varphi' - \nabla^2 \varphi\right) - \partial_y \varphi + \{\varphi, \varphi' - \nabla^2 \varphi\} = -a\chi \nabla^4 \varphi, \qquad (2.19)$$

which is the well-known (modified) Charney-Hasegawa-Mima (mCHM) equation (Hasegawa & Mima 1978) that includes the appropriate modified adiabatic electron response, with viscous damping. Even though we have considered a situation with no equilibrium density gradient, the magnetic drift provides a term identical to the one that would have arisen from the $\boldsymbol{E} \times \boldsymbol{B}$ advection of an inhomogeneous equilibrium density profile. This puts the model considered here in the same class of systems as those proposed by Hasegawa & Wakatani (1983), Terry & Horton (1983) and others — all essentially extensions of the Hasegawa-Mima equation with additional physics to account for microinstabilities in the plasma.

As (2.19) is contained within the model considered in this paper, equations (2.17) and (2.18) should, in principle, capture the behaviour of the mCHM equation as well as additional temperature and ITG effects. There has recently been a significant effort to advance the understanding of (2.19) and its relatives (Ruiz *et al.* 2016, 2019; Zhu *et al.* 2018*b*, 2019; Zhou *et al.* 2019; Plunk & Bañón Navarro 2017; St-Onge 2017; Majda *et al.* 2018; Qi *et al.* 2019). The mCHM equation does capture certain important phenomena, such as the generation of ZFs through a secondary instability (see Section 2.8); however,

its predictive capabilities for ITG turbulence are unclear. In particular, we shall find that the break up of the Dimits state of (2.17) and (2.18) is, in an essential way, governed by the behaviour of the temperature perturbations, which are absent in (2.19) (see Section 4).

2.6. *Linear Physics of ITG Instability*

Let us analyse the linear stability of (2.17) and (2.18). Dropping the nonlinear terms, we look for Fourier modes $\varphi, T \propto \exp[(\gamma_{\boldsymbol{k}} - i\omega_{\boldsymbol{k}})t + i\boldsymbol{k} \cdot \boldsymbol{r}]$, where $\gamma_{\boldsymbol{k}}$ and $\omega_{\boldsymbol{k}}$ are the real growth rate and frequency, respectively. Figure 2 shows $\gamma_{\boldsymbol{k}}$ as a function of the wavenumber $\boldsymbol{k}$. Qualitatively it resembles the growth rate of toroidal ITG modes in tokamaks (Horton *et al.* 1981). This is expected because the mechanism of the toroidal ITG instability is similar to that of the 2D curvature-driven ITG instability. The terms that give rise to the instability are the magnetic drift term $-\partial_y T$ in (2.17) and the $\boldsymbol{E} \times \boldsymbol{B}$ advection of the equilibrium temperature $\kappa_T \partial_y \varphi$ in (2.18). We find that the fastest growing linear modes are radially extended across the entire box, i.e., they have $k_x = 0$. Such modes are sometimes called "streamers".

The exact dispersion relation is

$$(\gamma_{\boldsymbol{k}} - i\omega_{\boldsymbol{k}})^2(1 + k^2) + (\gamma_{\boldsymbol{k}} - i\omega_{\boldsymbol{k}})\{-ik_y(1 + \kappa_T k^2) + \chi k^2[1 + (1 + a)k^2]\} \\ + a\chi^2 k^6 - \kappa_T k_y^2 - ik_y \chi k^2[1 + \kappa_T(1 - b)k^2] = 0. \quad (2.20)$$

We can get a good qualitative idea of the properties of the instability by setting $\chi = 0$. Then the solution of (2.20) is

$$\gamma_{\boldsymbol{k}} - i\omega_{\boldsymbol{k}} = \frac{ik_y(1 + \kappa_T k^2) \pm k_y\sqrt{4\kappa_T(1 + k^2) - (1 + \kappa_T k^2)^2}}{2(1 + k^2)}, \quad (2.21)$$

so the growth rate of the unstable mode is

$$\gamma_{\boldsymbol{k}} = \frac{k_y\sqrt{(2\sqrt{\kappa_T} - 1 + \kappa_T k^2)(2\sqrt{\kappa_T} + 1 - \kappa_T k^2)}}{2(1 + k^2)}. \quad (2.22)$$

To simplify further, consider $\kappa_T \gg 1 \gg \kappa_T^{-1/4} \gg k$. Then

$$\gamma_{\boldsymbol{k}} \approx k_y\sqrt{\kappa_T}. \quad (2.23)$$

This expression, with the normalisations (2.16) undone, is the well-known "bad-curvature-instability" growth rate (Beer 1995):

$$\gamma_{\boldsymbol{k}} = \Omega_i \frac{\rho_i^2 k_y}{\sqrt{2\tau L_B L_T}}. \quad (2.24)$$

Note that there is another, physically distinct, ITG instability usually referred to as the "slab ITG mode". This instability relies on coupling density and temperature through parallel-velocity perturbations, and so is naturally three-dimensional (Cowley *et al.* 1991). This mode is entirely absent from our 2D model.

Now let us return to the general dispersion. An important feature of the modes described by (2.20) is the boundedness of the region of unstable wavenumbers in the $\boldsymbol{k}$ plane (right panel of Figure 2). This allows us to integrate (2.17) and (2.18) without the need for artificial dissipation. There are both collisionless and collisional mechanisms that lead to the suppression of the ITG instability. Let us consider these mechanisms.
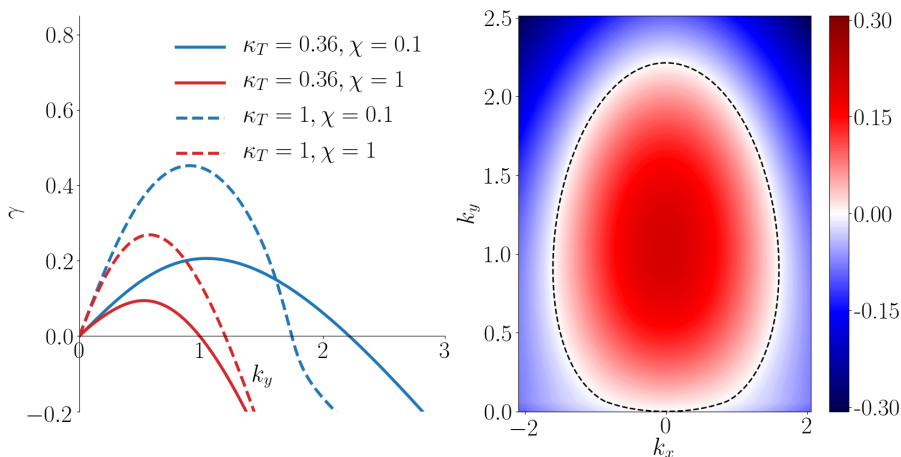
FIGURE 2. **Left:** Dependence of the growth rate $\gamma_{\boldsymbol{k}}$ on k_y for the streamer modes ($k_x = 0$). **Right:** Dependence of $\gamma_{\boldsymbol{k}}$ on k_x and k_y for $\kappa_T = 0.36, \chi = 0.1$. The dashed line is the boundary between stable and unstable modes ($\gamma_{\boldsymbol{k}} = 0$). In Section 3, we will consider nonlinear simulations with these same parameters.

2.6.1. *Collisionless Bounds on Unstable Wavenumbers*

It is easy to see that, in order to be positive, the collisionless growth rate (2.22) requires $k < k_{\mathrm{max,FLR}}$, where

$$k_{\mathrm{max,FLR}}^2 = \frac{1 + 2\sqrt{\kappa_T}}{\kappa_T}. \tag{2.25}$$

For $\kappa_T < 1/4$, (2.22) also gives a lower bound on the wavenumbers k of the unstable collisionless modes, viz., $k > k_{\mathrm{min,FLR}}$, where

$$k_{\mathrm{min,FLR}}^2 = \frac{1 - 2\sqrt{\kappa_T}}{\kappa_T}. \tag{2.26}$$

Adding collisions re-establishes the instability at low k. We deem this to be an unimportant peculiarity of our model, thus we shall only consider $\kappa_T > 1/4$.

2.6.2. *Collisional Bounds on Unstable Wavenumbers*

For nonzero ($\chi > 0$) collisionality, the term $a\chi^2 k^6$ in (2.20) dominates over the ITG term $\kappa_T k_y^2$ when k is large enough and gives strictly damped modes. To show this, let us simplify (2.20) by writing it as

$$(\gamma_{\boldsymbol{k}} - i\omega_{\boldsymbol{k}} + A)(\gamma_{\boldsymbol{k}} - i\omega_{\boldsymbol{k}} + B - iC) - fAB + igAC = 0, \tag{2.27}$$

where

$$A = \chi k^2, \quad B = \frac{a\chi k^4}{1 + k^2}, \quad C = k_y \frac{1 + \kappa_T k^2}{1 + k^2}, \quad f = \frac{\kappa_T k_y^2}{a\chi^2 k^6}, \quad g = \frac{b\kappa_T k^2}{1 + \kappa_T k^2}. \tag{2.28}$$

The instability threshold is given by $\gamma_{\boldsymbol{k}} = 0$. The real and imaginary parts of (2.27) for $\gamma_{\boldsymbol{k}} = 0$ are

$$-\omega_{\boldsymbol{k}}^2(1 + k^2) - \omega_{\boldsymbol{k}}C + (1 - f)AB = 0, \tag{2.29}$$

$$-\omega_{\boldsymbol{k}}(A + B) - (1 - g)AC = 0. \tag{2.30}$$

Substituting into (2.29) the value of $\omega_{\boldsymbol{k}}$ derived from (2.30), and using $A \neq 0$, we find

$$g(1-g)AC^2 + BC^2(1-g) + (1-f)B(A+B)^2 = 0. \tag{2.31}$$

Since $g \in (0,1)$†, a necessary condition for instability is

$$f > 1 \implies a\chi^2 k^6 < \kappa_T k_y^2. \tag{2.32}$$

Thus, the region of unstable modes is bounded by $k < k_{\mathrm{max},\chi}$, where

$$k_{\mathrm{max},\chi}^2 = \sqrt{\frac{\kappa_T}{a\chi^2}}. \tag{2.33}$$

2.7. *Conservation Laws*

Equations (2.17) and (2.18) have several conservations laws describing the time evolution of quantities that would be conserved in the absence of equilibrium gradients and dissipation:

$$\partial_t \int dxdy \, \frac{1}{2}T^2 = -\kappa_T \int dxdy \, T\partial_y\varphi - \chi \int dxdy \, (\boldsymbol{\nabla}T)^2, \tag{2.34}$$

$$\partial_t \int dxdy \, \frac{1}{2}\big[\varphi'^2 + (\boldsymbol{\nabla}\varphi)^2\big] \tag{2.35}$$

$$= -\int dxdy \, T\partial_y\varphi - \chi \int dxdy \, (\nabla^2\varphi)\big(a\nabla^2\varphi - b\nabla^2 T\big),$$

$$\partial_t \int dxdy \, \left[\frac{1}{2}\varphi'^2 + T\varphi' + \frac{1}{2}(\boldsymbol{\nabla}T + \boldsymbol{\nabla}\varphi)^2\right]$$

$$= -\chi \int dxdy \, \bigg[(\boldsymbol{\nabla}\varphi') \cdot (\boldsymbol{\nabla}T) + a\big(\nabla^2\varphi\big)^2$$

$$+ (a+1-b)\big(\nabla^2\varphi\big)\big(\nabla^2 T\big) + (1-b)\big(\nabla^2 T\big)^2\bigg]. \tag{2.36}$$

These are fairly easy to spot by inspection of the nonlinear terms in (2.17) and (2.18). They are also particular cases of the conservation laws of the gyrokinetic equation. The conservation of the variance of T, given by (2.34), is the lowest-order version of the gyrokinetic free-energy budget. The other two conservation laws, (2.35) and (2.36), can be derived from the conservation of the two-dimensional gyrokinetic invariant (see Schekochihin *et al.* 2009; Plunk *et al.* 2010). This invariant is a function of velocity in the GK formalism. The model presented here is based only on two velocity moments of the distribution function, namely density and temperature, and so the two-dimensional invariant yields two independent conservation laws. More specifically, (2.35) is a generalisation of the "electrostatic gyrokinetic invariant". The derivations of the three invariants of our system directly from the corresponding GK invariants can be found in Appendix B.

Equations (2.34) and (2.35) imply that a steady saturated state, i.e., $\partial_t = 0$ for all averaged quantities, can be achieved only if appropriate balance between injection and dissipation terms is established:

$$\frac{\chi}{\kappa_T} \int \frac{dxdy}{L_x L_y} \, (\boldsymbol{\nabla}T)^2 = \chi \int \frac{dxdy}{L_x L_y} \, (\nabla^2\varphi)\big(a\nabla^2\varphi - b\nabla^2 T\big) = Q, \tag{2.37}$$

---

† Note that for $b < 0$ or $b > 1$, there would be a collisional ($\kappa_T = 0$) instability. No such instability exists in our model because the Landau collision operator gives $g \in (0,1)$.

where the total radial heat flux Q is‡

$$Q = -\frac{1}{L_x L_y} \int dxdy \, T\partial_y\varphi. \tag{2.38}$$

Thus, a saturated state would necessarily have a net positive "turbulent" (or "anomalous") heat flux $Q > 0$. Note that the first term on the right-hand side of (2.34), which represents injection of free energy, is $\kappa_T Q$. The turbulent heat flux is enabled by the turbulence excited by the ITG instability.

Note as well that the linearly unstable modes have a positive radial heat flux. Indeed, from (2.38),

$$Q = \sum_{\boldsymbol{k}} ik_y T_{\boldsymbol{k}} \varphi_{\boldsymbol{k}}^* = \sum_{\boldsymbol{k}} ik_y |\varphi_{\boldsymbol{k}}|^2 \frac{T_{\boldsymbol{k}}}{\varphi_{\boldsymbol{k}}}. \tag{2.39}$$

The relative phase of the temperature and potential perturbations can be obtained from (2.18):

$$\frac{T_{\boldsymbol{k}}}{\varphi_{\boldsymbol{k}}} = \frac{-ik_y \kappa_T}{\gamma_{\boldsymbol{k}} - i\omega_{\boldsymbol{k}} + \chi k^2}, \tag{2.40}$$

where $\gamma_{\boldsymbol{k}} - i\omega_{\boldsymbol{k}}$ is the solution of the dispersion relation (2.20). Then

$$Q = \sum_{\boldsymbol{k}} \kappa_T k_y^2 |\varphi_{\boldsymbol{k}}|^2 \frac{\gamma_{\boldsymbol{k}} + \chi k^2}{(\gamma_{\boldsymbol{k}} + \chi k^2)^2 + \omega_{\boldsymbol{k}}^2} > 0 \tag{2.41}$$

for the unstable modes, which have $\gamma_{\boldsymbol{k}} > 0$.

Finally, the third conservation law (2.36) has some peculiar properties. First, neither the conserved quantity on the left-hand side nor the dissipation rate on the right-hand side is sign-definite. Secondly, all of the evolution is dissipative, i.e., this invariant is not injected by any equilibrium gradients and is constant in time if $\chi = 0$.

2.8. *Secondary Instability*

Before we delve into the study of nonlinear saturation, let us show how ZFs can be generated from the linearly unstable ITG modes. Consider the stability of a streamer mode with $k_x = 0, k_y = q$ (the "primary" mode) to infinitesimal "secondary" perturbations:

$$\varphi = (\varphi_q e^{iqy} + \text{c.c.}) + \delta\varphi(x, y), \tag{2.42}$$

$$T = (T_q e^{iqy} + \text{c.c.}) + \delta T(x, y). \tag{2.43}$$

A common way of analysing the secondary instability is to take a Galerkin truncation by considering only four Fourier modes $(k_x, k_y) = \{(0, q), (p, 0), (p, \pm q)\}$ and their complex conjugates: the $(0, q)$ mode is the primary streamer in (2.42) and (2.43) and the others are

$$\delta\varphi = \left(\delta\varphi_+ e^{iqy} + \delta\varphi_- e^{-iqy} + \delta\varphi_0\right)e^{ipx}e^{\gamma_2 t} + \text{c.c.}, \tag{2.44}$$

$$\delta T = \left(\delta T_+ e^{iqy} + \delta T_- e^{-iqy} + \delta T_0\right)e^{ipx}e^{\gamma_2 t} + \text{c.c.}, \tag{2.45}$$

where p is the radial wavenumber of the secondary perturbations, $\delta\varphi_0$ and δT_0 are the zonal flow and temperature, and $\delta\varphi_\pm$ and $\delta T_\pm$ are known as "sidebands". Substituting all this into (2.17) and (2.18) and linearising the nonlinear terms for $\delta\varphi \ll \varphi_q$ and $\delta T \ll T_q$, we obtain a closed set of equations. In order to keep things simple, we drop the linear

---

‡ The dimensional ion heat flux $Q_i = V^{-1} \int d^3\boldsymbol{r} \int d^3\boldsymbol{v}(\boldsymbol{V}_E \cdot \hat{\boldsymbol{x}})(m_i v^2/2)\delta f_i$ (Barnes *et al.* 2011), where V is the volume of integration and δf_i is the perturbed ion distribution function [see (A 1) in Appendix A.1], is related to Q via $Q_i/Q = 3n_i T_i v_{ti}(\rho_i/L_B)^2/\tau^{5/2}\sqrt{2}$.

terms in (2.17) and (2.18) — this is valid when the amplitude of the primary mode is large enough, so that interactions with it are more important for the evolution of $\delta\varphi$ and δT than the effects of the equilibrium gradients and collisions. Observe that, due to the structure of the Poisson bracket (2.11), all nonlinear terms are proportional to pq. Defining for convenience $\gamma_2 \equiv \sqrt{2}pq|\varphi_q|\hat{\gamma}_2$, we obtain the following equation for $\hat{\gamma}_2$:

$$\left(\hat{\gamma}_2^2 + U\right)\left(\hat{\gamma}_2^2 + V\right) = W, \tag{2.46}$$

where

$$U = 1 + \frac{q^2\mathrm{Re}(T_q/\varphi_q)}{1 + p^2 + q^2},$$

$$V = \frac{p^2\mathrm{Im}(T_q/\varphi_q)^2 + p^2[1 + \mathrm{Re}(T_q/\varphi_q)]^2 - (1 + q^2)[1 + \mathrm{Re}(T_q/\varphi_q)]}{1 + p^2 + q^2},$$

$$W = \frac{p^2 q^2}{(1 + p^2 + q^2)^2}\Big[|T_q|^2/|\varphi_q|^2 + 2\mathrm{Re}(T_q/\varphi_q)\Big]\Big[1 + \mathrm{Re}(T_q/\varphi_q)\Big]. \tag{2.47}$$

We see that the growth rate γ_2 of the secondary instability depends both on the amplitudes of the primary fields φ_q and T_q, and on their relative phase.

2.8.1. *No Temperature Perturbation*

If we set $T_q = 0$, i.e., ignore the temperature perturbation of the primary streamer, (2.46) gives the well-known dispersion relation for the secondary instability of the modified Hasegawa-Mima model (Rogers *et al.* 2000; Strintzi & Jenko 2007):

$$\gamma_2^{\mathrm{HM}} = pq|\varphi_q|\sqrt{\frac{2(1 + q^2 - p^2)}{1 + p^2 + q^2}}. \tag{2.48}$$

This form of the secondary instability has long been associated with the strong ZFs observed numerically in ITG turbulence (Hammett *et al.* 1993)†. We will show that the inclusion of the temperature perturbations can introduce qualitative and quantitative changes, and even suppress the secondary instability completely.

2.8.2. *Long-Wavelength Limit*

To simplify (2.46), we can consider the long-wavelength limit $p \ll 1$. Then (2.46) gives

$$\left[\hat{\gamma}_2^2 + 1 + \frac{q^2\mathrm{Re}(T_q/\varphi_q)}{1 + q^2}\right]\big[\hat{\gamma}_2^2 - 1 - \mathrm{Re}(T_q/\varphi_q)\big] = \mathcal{O}\big(p^2\big) \approx 0. \tag{2.49}$$

Thus, there are two independent branches of the secondary instability with instability conditions given by $\mathrm{Re}(T_q/\varphi_q) + 1/q^2 < -1$ and $\mathrm{Re}(T_q/\varphi_q) > -1$, respectively. The second branch is a modified form of the long-wavelength Hasegawa-Mima secondary instability (2.48)‡:

$$\gamma_2 = pq|\varphi_q|\sqrt{2\left(1 + \mathrm{Re}\frac{T_q}{\varphi_q}\right)}. \tag{2.50}$$

† Especially in contrast with the much weaker ZFs observed in electron-temperature-gradient–driven (ETG) turbulence on electron scales (Jenko *et al.* 2000; Strintzi & Jenko 2007). However, this distinction between ITG and ETG turbulence has recently been challenged by Colyer *et al.* (2017), who found that the long-time saturated state of ETG turbulence is also dominated by ZFs, although the system does go through a streamer-dominated quasi-saturated state at earlier times.
‡ Plunk & Bañón Navarro (2017) found the same expression in the context of the "warm-ion" approximation, i.e., dropping the FLR terms in the nonzonal part of (2.17).

We observe that (2.50) relies only on a handful of the nonlinear terms in (2.17) and (2.18). Substituting (2.42) and (2.44) into (2.17) and taking the limit $p \ll 1$ gives us the following equations for $\delta\varphi_0, \delta\varphi_+$ and $\delta\varphi_-$:

$$\hat{\gamma}_2 \delta\varphi_0 = \frac{1}{\sqrt{2}}\left[\delta\varphi_+(\varphi_q^* + T_q^*) - \delta\varphi_-(\varphi_q + T_q)\right], \tag{2.51}$$

$$\hat{\gamma}_2 \delta\varphi_+ = \frac{1}{\sqrt{2}}\varphi_q \delta\varphi_0, \tag{2.52}$$

$$\hat{\gamma}_2 \delta\varphi_- = -\frac{1}{\sqrt{2}}\varphi_q^* \delta\varphi_0. \tag{2.53}$$

Substituting (2.52) and (2.53) into (2.51) yields precisely (2.50). We do not consider the equations for the temperature perturbations because $\delta T = 0$ is a consistent solution and it corresponds to (2.50). The terms on the right-hand side of (2.52) and (2.53) arise from the zonal advection term $\{\overline{\varphi}, \varphi' - \nabla^2 \varphi'\}$ in (2.17) and represent the tilting of the primary streamer by the ZF. The terms on the right-hand side of (2.51) are the poloidal $\boldsymbol{E} \times \boldsymbol{B}$ and diamagnetic flows caused by the interaction of the primary mode and the two sidebands $(k_x, k_y) = (p, \pm q)$. The quantity $\mathrm{Re}(T_q/\varphi_q)$ controls the response of the primary mode to the zonal perturbation: $\mathrm{Re}(T_q/\varphi_q) > -1$ yields an unstable ZF, while $\mathrm{Re}(T_q/\varphi_q) < -1$ results in a stable, oscillatory perturbation.

Let us consider the collisionless case ($\chi = 0$), where analytical progress is possible, and ask for what values of κ_T the two modes described by (2.49) are unstable. Let us take $(k_x, k_y) = (0, q)$ to be the linear mode with the largest growth rate. We then define the critical gradient κ_T^{sec} for the long-wavelength secondary instability of the fastest-growing streamer as the value of κ_T at which $\mathrm{Re}(T_q/\varphi_q) = -1$. Using (2.21) and the relationship (2.40) between T_q and φ_q, we obtain

$$\mathrm{Re}\frac{T_q}{\varphi_q} = -\frac{1 + \kappa_T q^2}{2}. \tag{2.54}$$

To determine q, we seek the maximum of $\gamma_{\boldsymbol{k}}$, as given by (2.22) for $k_x = 0$ and $k_y = q$. We find

$$\frac{\partial \gamma_{\boldsymbol{k}}}{\partial q} \propto \kappa_T^2 q^6 + 3\kappa_T^2 q^4 - q^2 - 4\kappa_T + 1 = 0, \tag{2.55}$$

where the equality holds for the most unstable mode. As an equation for q^2, (2.55) is a cubic with only one positive solution for $\kappa_T > 1/4$. Substituting that solution into (2.54), we find $\mathrm{Re}(T_q/\varphi_q)$ as a function of κ_T. This relationship is given in Figure 3. In particular, we obtain that $\mathrm{Re}(T_q/\varphi_q) = -1$ at $\kappa_T = 1$, as can indeed be verified analytically from (2.55) and (2.54), and $\mathrm{Re}(T_q/\varphi_q) < -1$ for $\kappa_T > 1$. We also find that $\mathrm{Re}(T_q/\varphi_q) + 1/q^2 > -1$ always. Thus, for $\kappa_T > 1$, the most unstable collisionless ITG mode is stable to the secondary perturbations.

2.8.3. *General Case*

Let us go back to the general secondary dispersion relation (2.46). Its solution is

$$\hat{\gamma}_2^2 = \frac{-(U + V) \pm \sqrt{(U - V)^2 + 4W}}{2}, \tag{2.56}$$

where

$$U + V = \frac{p^2\left[1 + \mathrm{Im}(T_q/\varphi_q)^2\right] + p^2[1 + \mathrm{Re}(T_q/\varphi_q)]^2 - \mathrm{Re}(T_q/\varphi_q)}{1 + p^2 + q^2}. \tag{2.57}$$
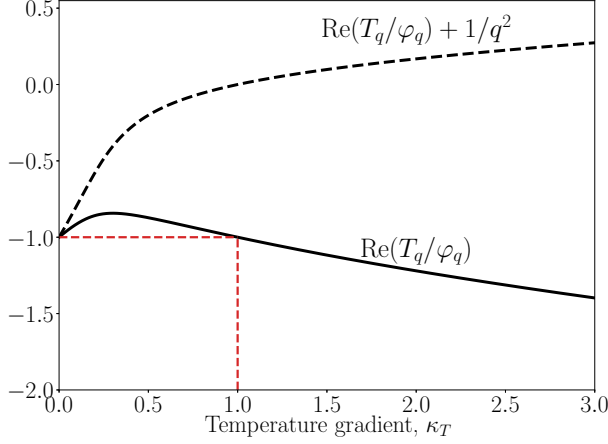
FIGURE 3. Temperature-gradient dependence of $\mathrm{Re}(T_{\boldsymbol{k}}/\varphi_{\boldsymbol{k}})$ (solid) and $\mathrm{Re}(T_{\boldsymbol{k}}/\varphi_{\boldsymbol{k}}) + 1/q^2$ (dashed) for the most unstable collisionless ($\chi = 0$) mode. We find that $\mathrm{Re}(T_{\boldsymbol{k}}/\varphi_{\boldsymbol{k}}) < -1$ for $\kappa_T > 1$ and $\mathrm{Re}(T_{\boldsymbol{k}}/\varphi_{\boldsymbol{k}}) + 1/q^2 > -1$ for all κ_T. The secondary instability is present only for $\kappa_T < 1$.

We can use the primary dispersion relation (2.20) to show that $\mathrm{Re}(T_{\boldsymbol{k}}/\varphi_{\boldsymbol{k}}) < 0$, and hence $U + V > 0$, for any unstable primary mode with wavenumber $\boldsymbol{k}$. Indeed, the real part of (2.40) is

$$\mathrm{Re}\frac{T_{\boldsymbol{k}}}{\varphi_{\boldsymbol{k}}} = \frac{k_y \kappa_T \omega_{\boldsymbol{k}}}{|\gamma_{\boldsymbol{k}} - i\omega_{\boldsymbol{k}} + \chi k^2|^2} < 0 \qquad (2.58)$$

if $k_y \omega_{\boldsymbol{k}} < 0$. Let us show that this is true. For $k \ll 1$ and $\kappa_T > 1/4$ (the reasons for the latter are discussed at the end of Section 2.6.1), the dispersion relation (2.20) gives simply $k_y \omega_{\boldsymbol{k}} = -k_y^2/2 < 0$. Since the solutions to (2.20) are continuous functions of $\boldsymbol{k}$, if $k_y \omega_{\boldsymbol{k}}$ changes sign and becomes positive, then $\omega_{\boldsymbol{k}} = 0$ somewhere. However, if we set $\omega_{\boldsymbol{k}} = 0$, the imaginary part of (2.27) gives $\gamma_{\boldsymbol{k}} = (g - 1)A < 0$. Therefore, $k_y \omega_{\boldsymbol{k}}$ cannot change sign within the region of linear instability and so $k_y \omega_{\boldsymbol{k}} < 0$ for all linearly unstable modes.

We now consider the solution (2.56) assuming that the relationship between φ_q and T_q is given by (2.40) with $k_y = q$ and $\gamma_{\boldsymbol{k}}$ and $\omega_{\boldsymbol{k}}$ corresponding to the most unstable mode. This gives us γ_2 as a function of κ_T, χ and p. Figure 4 shows the real part of γ_2 maximised over p for each pair of equilibrium parameters κ_T and χ, and the wavenumber $p_{\max}$ at which that maximum is attained. Let us discuss this figure. There are three distinct regions:

(i) $\kappa_T < \kappa_T^{\mathrm{sec}}$, where κ_T^{sec} is defined as the value of κ_T where $\mathrm{Re}(T_q/\varphi_q) = -1$; in this region, $\mathrm{Re}(T_q/\varphi_q) > -1$. Additionally, $UV - W < 0$ for $p = p_{\max}$, so $\hat{\gamma}_2^2$ given by (2.56) is real and positive. The instability exists for arbitrarily small values of p (i.e., for an arbitrarily long wavelength of the ZF). Increasing the temperature gradient κ_T towards κ_T^{sec} has a dramatic effect on the secondary instability of the most unstable mode: it diminishes both the growth rate and the region of zonal wavenumbers that go unstable. On the line $\kappa_T = \kappa_T^{\mathrm{sec}}$, $\hat{\gamma}_2$ is purely imaginary and there are no growing secondary modes, just like in the long-wavelength analysis of Section 2.8.2. Indeed, substituting $\mathrm{Re}(T_q/\varphi_q) = -1$ in (2.47), we obtain $W = 0$ and $U, V > 0$. Then, by (2.56), $\hat{\gamma}_2^2 = -U$ or $-V$. Figure 5 ($\kappa_T = 0.7, 1.1, 1.5$) shows γ_2 vs. p in region (i).
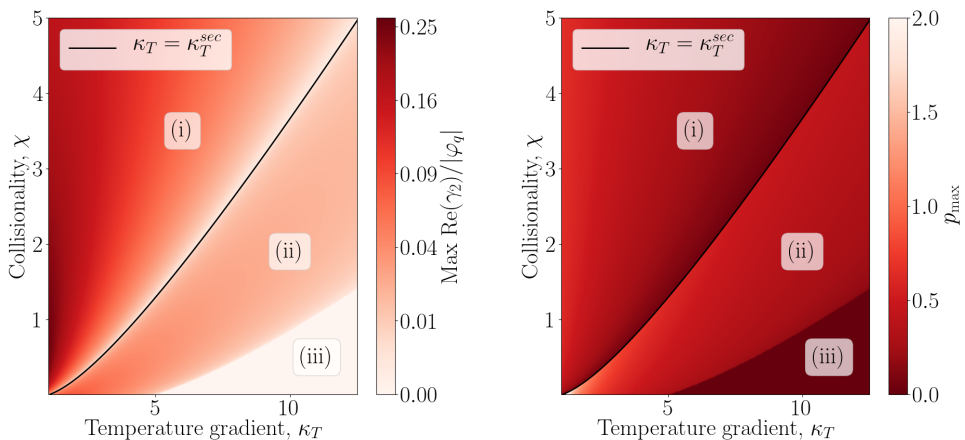
FIGURE 4. **Left:** Secondary-instability growth rate (2.56) of the most unstable streamer mode, maximised over all values of p. The growth rate vanishes on the $\mathrm{Re}(T_q/\varphi_q) = -1$ curve (shown in black). **Right:** Radial wavenumber $p_{\max}$ at which the maximum growth rate shown in the left panel is attained. A discontinuity in the most unstable wavenumber across the $\mathrm{Re}(T_q/\varphi_q) = -1$ curve is evident. The absolute-stability region (iii), visible in the bottom right of both panels, where $\gamma_2 = 0$, is the region where $(U - V)^2 + 4W$ is always positive and γ_2 is purely imaginary.

(ii) $\kappa_T > \kappa_T^{\mathrm{sec}}$. Increasing κ_T past κ_T^{sec} changes the fastest-growing secondary mode discontinuously. The fastest-growing secondary mode now has $UV - W > 0$ and $(U - V)^2 + 4W < 0$. Hence the $\hat{\gamma}_2^2$ given by (2.56) is complex. In this region, there is always $\hat{\gamma}_2$ with a positive real part. The peak-growth wavenumber $p_{\max}$ changes discontinuously across the $\mathrm{Re}(T_q/\varphi_q) = -1$ line. For $\kappa_T > \kappa_T^{\mathrm{sec}}$, the secondary instability does not extend to arbitrarily small p (Figure 5, $\kappa_T = 1.9, 2.3$), consistent with the discussion of the long-wavelength secondary instability in Section 2.8.2.

(iii) $\kappa_T > \kappa_T^{\mathrm{sec}}$, but now $(U - V)^2 + 4W > 0$ for all values of p, so $\hat{\gamma}_2^2$ given by (2.56) is real and negative. The location of this region of stability depends on the value of $\mathrm{Im}(T_q/\varphi_q)$, as well as $\mathrm{Re}(T_q/\varphi_q)$, and does not have a simple analytic form like the boundary between regions (i) and (ii).

This analysis of the secondary instability suggests that the system will fail to generate ZFs at a high enough κ_T. In what follows, we will indeed find that the zonally dominated Dimits regime ceases to exist when the temperature gradient exceeds a certain threshold, $\kappa_T > \kappa_T^c$. However, the naïve guess $\kappa_T^c \approx \kappa_T^{\mathrm{sec}}$, as given by the secondary-instability threshold of the most unstable streamer, does not yield satisfactory agreement with the observed threshold for the Dimits regime (see Section 4.4). The secondary-instability picture is incomplete because we must take into account not only whether ZFs can be generated by the ITG modes, but also whether the strong ZFs that support the Dimits regime are resilient to nonzonal perturbations. We shall pick up this topic in Section 4.

2.9. *Tertiary Instability*

To study the stability of a zonal state, we consider infinitesimal ITG perturbations over a background of strong ZF and zonal temperature:

$$\varphi = \overline{\varphi} + \varphi', \ \varphi' \ll \overline{\varphi},$$
$$T = \overline{T} + T', \ T' \ll \overline{T}, \tag{2.59}$$
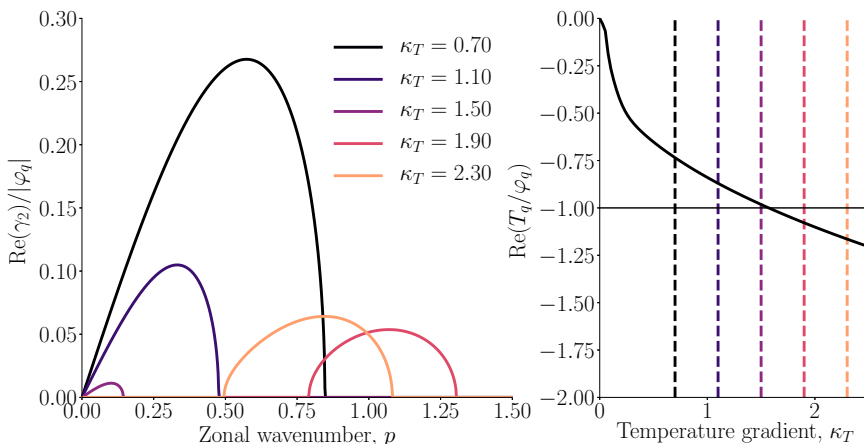
FIGURE 5. **Left:** Secondary-instability growth rate for $\chi = 0.1$ and a number of values of κ_T versus zonal (radial) wavenumber, as given by (2.46). The primary mode φ_q, T_q is taken to be the most unstable one in every case. **Right:** $\mathrm{Re}(T_k/\varphi_k)$ for the fastest-growing linear mode $\boldsymbol{k} = (0, q)$ for $\chi = 0.1$ as a function of κ_T. The dashed lines correspond to the same values of κ_T as in the left panel.

and linearise (2.17) and (2.18) to obtain evolution equations for φ' and T'. We refer to the ITG modes governed by these linearised equations as "tertiary modes", and to their linear instability as the "tertiary instability" (in truth, this is just the primary ITG instability but for an equilibrium state modified by the zonal fields). We will discover that this instability can seed turbulent perturbations in the Dimits regime, but is not solely responsible for the transition to strong turbulence (see Sections 3 and 4). Further discussion of the tertiary instability has been exiled to Appendix C.

3. Nonlinear Saturation and Zonal Staircase

We now proceed to investigate the saturated state of (2.17) and (2.18) numerically and semi-analytically. A well-defined saturated state is found only for temperature gradients below a critical gradient $\kappa_T < \kappa_T^c$, where κ_T^c is an increasing function of collisionality χ (see Figure 6). The saturated state is always dominated by strong zonal flows and exhibits levels of turbulent transport that are low compared to the equilibrium diffusive transport ($Q \lesssim \chi\kappa_T$). We will refer to this state as the Dimits state. The critical gradient κ_T^c is then the nonlinear critical gradient that marks the break up of the zonally dominated state and the onset of fully developed ITG turbulence. We will relate κ_T^c to the resilience of the zonal profiles in the face of nonzonal perturbations, which is in turn determined by the behaviour of turbulence in the presence of strong (comparable to the ITG growth rate) zonal shear. For $\kappa_T < \kappa_T^c$, zonally sheared turbulence enhances the ZFs that are doing the shearing through a negative turbulent viscosity. Beyond the Dimits threshold ($\kappa_T > \kappa_T^c$), the turbulent viscosity is positive, and strong, ITG-suppressing ZFs cannot be maintained. These results are presented in Section 4, but first, in this section, we shall describe the saturated state near the Dimits threshold.

Figure 6 shows the heat flux Q vs. κ_T and χ. We have checked that all simulations have converged by inspection of their heat flux and ZF profiles, and by ensuring that they run for several box-scale diffusion times $t_{\text{box}} = (L_x/2\pi)^2/a\chi$. The turbulent heat
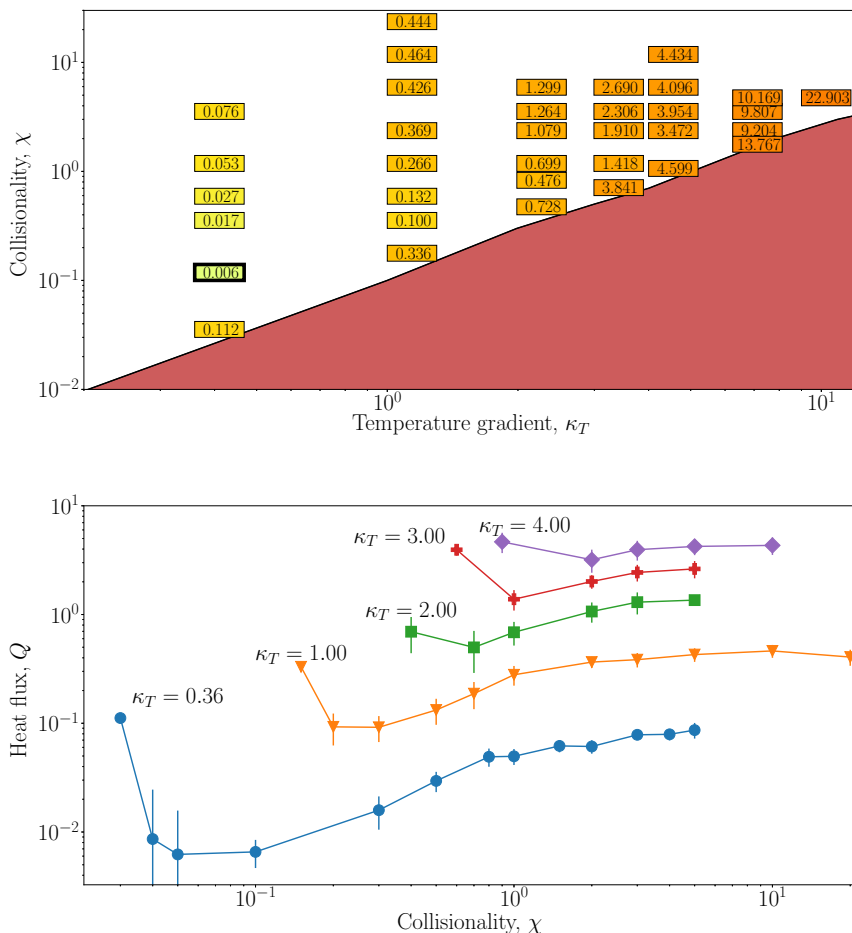
FIGURE 6. **Top:** Box-averaged heat flux Q [defined in (2.38)] as a function of κ_T and χ. The shaded (in red) region is beyond the Dimits threshold, where strong turbulence resides (see Section 4.5). Bold-framed is the parameter point corresponding to SimL and SimH, viz., $\kappa_T = 0.36, \chi = 0.1$. **Bottom:** Box-averaged heat flux Q in the saturated state versus χ for various κ_T.

flux Q depends strongly on the temperature gradient and increases monotonically with increasing κ_T. In contrast, its dependence on the collisionality is much weaker and non-monotonic (see Figure 6, bottom panel). Close to the Dimits threshold, Q decreases with increasing χ (which takes it away from the threshold), whereas farther away from the threshold, it increases and then plateaus with increasing χ.

In what follows, a significant fraction of the detailed analysis is done using two simulations of the low-collisionality near-marginal state with parameters $\kappa_T = 0.36$, $\chi = 0.1$, $L_x = 100$, $L_y = 150$, one with higher (507×337) and one with lower (167×167) number of Fourier modes (the lower-resolution simulation is used for longer runs due to its lower computational cost). They have the same initial condition, taken from an already saturated simulation. Both the low- and high-resolution simulations show good convergence of their spectra (see Figure 7). We shall refer to these two simulations as "SimL" and "SimH", respectively.
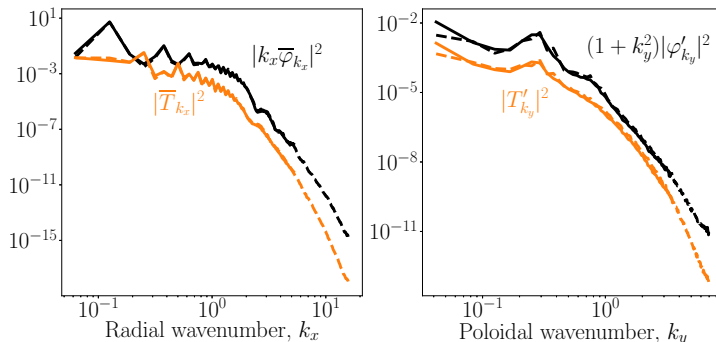
FIGURE 7. Comparison of the spectra of turbulence for SimL (solid) and SimH (dashed), described in Section 3. **Left:** ZF velocity $|k_x\overline{\varphi}_{k_x}|^2$ (black) and zonal temperature (orange) $|\overline{T}_{k_x}|^2$ spectra. **Right:** Streamer ($k_x = 0$) contribution to (2.35), viz., $(1 + k_y^2)|\varphi'_{k_y}|^2$, (black) and temperature $|T'_{k_y}|^2$ (orange). A clear peak at $k_y \approx 0.25$ is seen. This corresponds to the dominant poloidal wavenumber in the ZF minima (see also Figure 26 in Appendix C). The fastest linearly growing streamer has $k_y \approx 1$.

In the near-marginal Dimits state, turbulence is suppressed by a quasi-static "zonal staircase" arrangement of the ZFs and zonal temperature perturbations. This structure is reminiscent of the "$\boldsymbol{E} \times \boldsymbol{B}$ staircase" observed in global GK simulations (Dif-Pradalier *et al.* 2010, 2015, 2017; Villard *et al.* 2013, 2014). The zonal staircase consists of interleaved regions of strong zonal shear that suppresses the ITG turbulence in those regions, and localised turbulent patches at the turning points of the ZF velocity. We shall refer to the former as the "shear zones" (Section 3.1) and to the latter as the "convection zones" (Section 3.2). A typical near-marginal ZF configuration can be seen in Figure 8 and corresponding snapshots of the perturbed temperature in Figure 9. Turbulence is always present, in a highly localised form, in the convection zones, but not in the shear zones.

The ZF in the staircase is not steady, but subject to viscous decay. In the low-collisionality ($\chi \lesssim 1$), near-marginal regime, this decay is slow and turbulent bursts are triggered periodically when the zonal shear in the shear zones has decayed to a level that is insufficient for the suppression of turbulence. These bursts lead to a significant (order-of-magnitude) increase in the radial heat flux. They are seeded by highly localised, coherent, turbulent structures, reminiscent of those reported by van Wyk *et al.* (2016) in gyrokinetic turbulence with an imposed equilibrium flow shear (see Section 3.3). A typical turbulent burst is illustrated in Figures 8, 9 and 10 where we see the evolution of the quiescent state into a turbulent one and then back. Figure 11 shows a longer time evolution for the same parameters, illustrating the (quasi)periodic nature of the bursts. At higher collisionality, the ZFs decay faster, the turbulent bursts start to overlap, and it becomes difficult to isolate quiescent periods from turbulent ones. However, we find that the mechanism that governs the stability of the Dimits state and the transition to strong turbulence is very similar for all values of collisionality that we have explored (see Section 4.2).

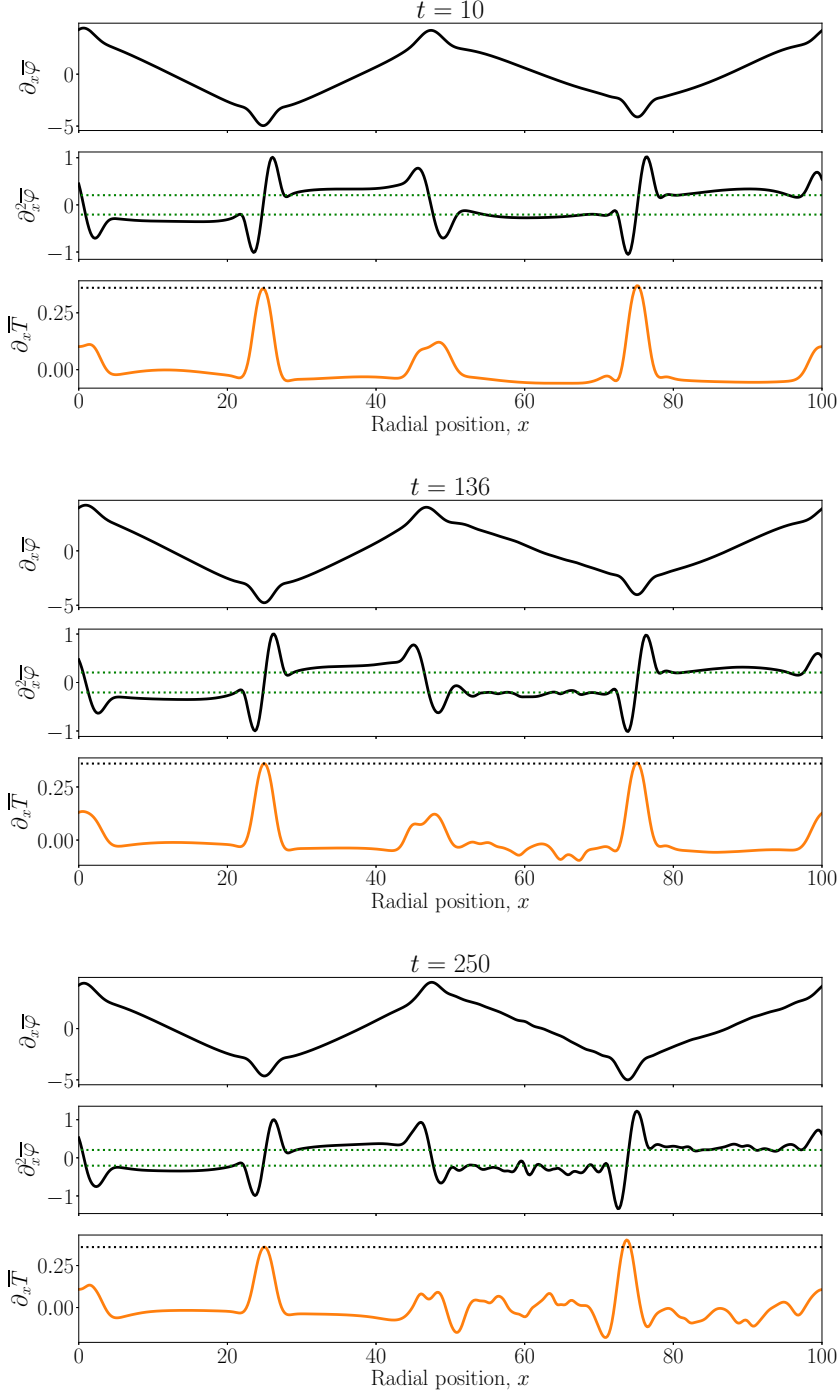We now proceed to describe the features of the zonal staircase in more detail.

FIGURE 8. Radial profiles of ZF ($\partial_x \overline{\varphi}$), zonal shear ($\partial_x^2 \overline{\varphi}$) and zonal-temperature gradient ($\partial_x \overline{T}$). The dotted green line corresponds to the largest linear ITG growth rate $\pm\gamma_{\mathrm{max}}$. The dotted black line shows the value of the equilibrium temperature gradient κ_T. Note that the turbulence that develops in the shear zones does not disturb the ZF and zonal shear significantly. See Figure 9 for 2D snapshots at these same times. The data is from SimH.
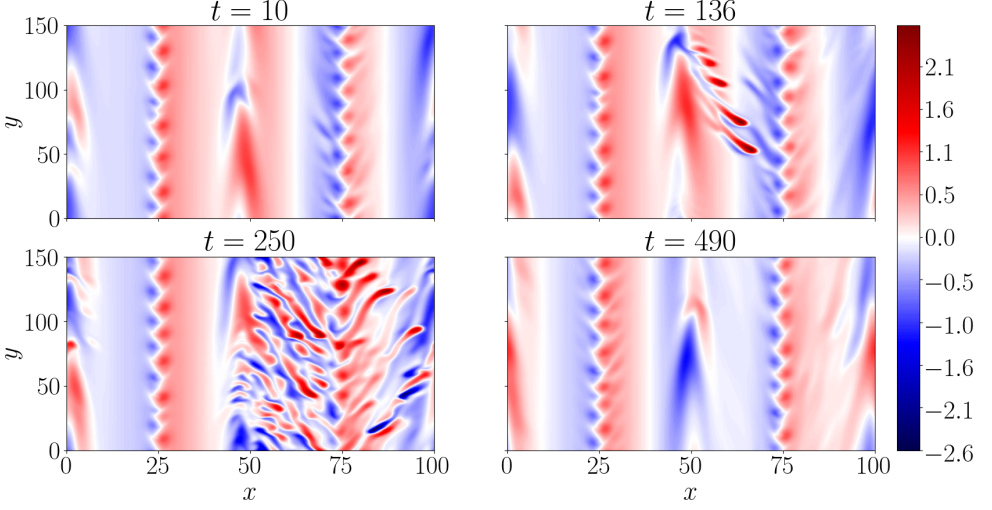
FIGURE 9. Snapshots of temperature perturbations in the Dimits state (a movie is available in the supplementary material). The data is from SimH. **Top left:** Quiescent, $t = 10$; there is a ZF minimum at $x = 24.5$ and a ZF maximum at $x = 47$. **Top right:** Ferdinons visible around $x = 65$, $t = 136$. A zoomed-in version can be found in Figure 18. **Bottom left:** Turbulent burst, $t = 250$. **Bottom right:** Relaxation back to the zonal staircase after the burst, $t = 490$. The full time history of SimH can be found in Figure 10 and the radial profiles of the zonal fields and heat flux are shown in Figure 8.
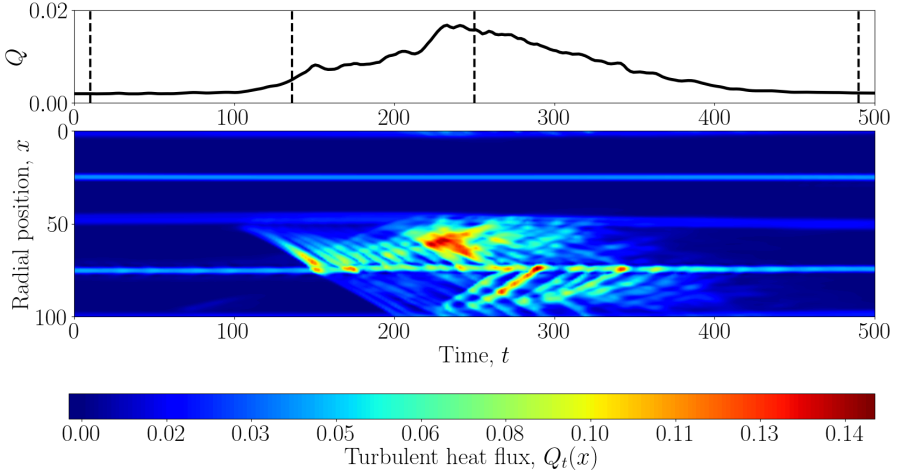

FIGURE 10. **Top:** Time evolution of the total heat flux Q during a turbulent burst for SimH. The dashed lines correspond to the times used for Figures 8 and 9. **Bottom:** Time trace of the local (integrated only over y) radial turbulent heat flux $Q_t(x) = -\overline{T \partial_y \varphi}$ as a function of radial position. A turbulent burst in the right half of the domain is clearly visible for $t \in [100, 400]$. The linear streaks correspond to radially drifting ferdinons (see Section 3.3).
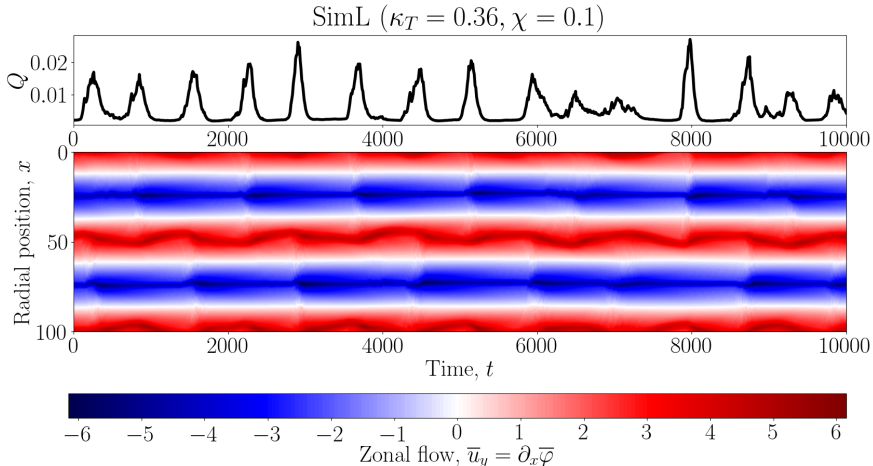
FIGURE 11.   Box-integrated radial heat flux Q and ZF velocity $(\partial_x \overline{\varphi})$ vs. time for SimL ($\kappa_T = 0.36$, $\chi = 0.1$). Each turbulent burst is accompanied by an order-of-magnitude increase in Q and a radial oscillation of the locations of the ZF maxima.
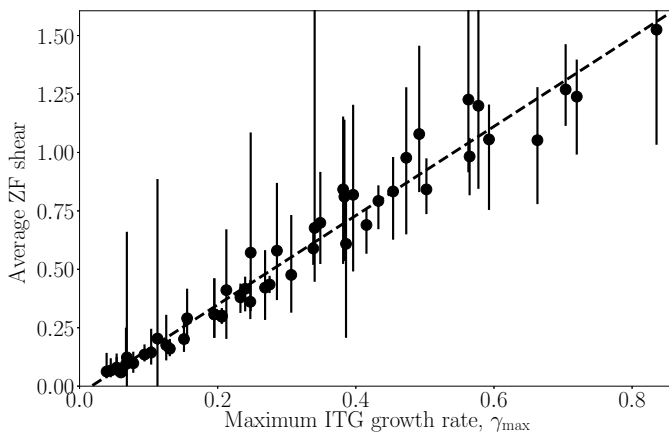

FIGURE 12. ZF shear S (time- and space-averaged over the shear zones) vs. maximum linear ITG growth rate γ_{max} [as given by (2.20)]. The data is taken from a number of simulations over a range of parameters: $\kappa_T \in [0.16, 7.29]$ and $\chi \in [0.1, 10]$. The error bars represent the smallest and the largest values of the spatially averaged ZF shear for each simulation. The best-fit line (dashed) is $S \approx 2\gamma_{\mathrm{max}}$.

3.1. *Shear Zones*

3.1.1. *Suppression of Turbulence*

The zonal staircase is arranged in such a way that it efficiently suppresses turbulence in the shear zones via strong ZF shear. We find that this shear, $S \equiv \partial_x^2 \overline{\varphi}$, satisfies $S \gtrsim \gamma_{\mathrm{max}}$, where γ_{max} is the largest linear ITG growth rate determined from the dispersion relation (2.20). The notion that ITG turbulence requires comparable S and γ_{max} to be suppressed by shear is known as the "quench rule" (Waltz *et al.* 1994, 1998; Kinsey *et al.* 2005; Kobayashi & Rogers 2012). Quantitatively, this is supported by Figure 12, which shows that the time- and space-averaged (over the shear zones) shear satisfies $S \approx 2\gamma_{\mathrm{max}}$ over

a range of simulation parameters†. Note that the *particular snapshots* of zonal profiles seen in Figure 8 suggest $S \approx \gamma_{\max}$. However, the *time-averaged S* is larger due to the variation of shear over time (see also Figure 13).

3.1.2. *Decay of Zonal Flows*

Let us study the viscous decay of the ZFs. The equation for the evolution of the ZFs is given by the zonal part of (2.17):

$$\partial_t \partial_x^2 \overline{\varphi} = \partial_x^2 \overline{\partial_x \varphi \partial_y (\varphi + T)} + \chi \partial_x^4 (a\overline{\varphi} - b\overline{T}). \tag{3.1}$$

Integrating (3.1) once yields

$$\partial_t \overline{u}_y = \partial_x \left[\overline{(\partial_x \varphi) \partial_y (\varphi + T)} + \chi \partial_x^2 (a\overline{\varphi} - b\overline{T}) \right] = -\partial_x (\Pi_t + \Pi_d), \tag{3.2}$$

where the zonal flow velocity is $\overline{u}_y \equiv \partial_x \overline{\varphi}$ and we have identified the turbulent, $\Pi_t \equiv -\overline{(\partial_x \varphi) \partial_y (\varphi + T)}$, and diffusive, $\Pi_d \equiv -\chi \partial_x^2 (a\overline{\varphi} - b\overline{T})$, radial fluxes of poloidal momentum. The integration constant in (3.2) is zero because both sides of the equation are exact derivatives with respect to x and our domain is periodic. Integrating (3.2) once more yields a term that is not necessarily an exact derivative — the turbulent momentum flux Π_t:

$$\partial_t \overline{\varphi} + \Pi_t + \Pi_d = \Pi, \tag{3.3}$$

where the integration constant $\Pi = (1/L_x) \int_0^{L_x} dx \, \Pi_t$ is the total box-averaged poloidal momentum flux. However, (2.17) and (2.18) are invariant under the symmetry

$$x \mapsto -x, \; y \mapsto y, \; \varphi \mapsto -\varphi, \; T \mapsto -T. \tag{3.4}$$

Under this symmetry, $\Pi \mapsto -\Pi$, a property of our model inherited from gyrokinetics (Parra *et al.* 2011). Therefore, assuming that the volume-averaged solutions to (2.17) and (2.18) respect (3.4), we conclude that $\Pi = 0$. This is confirmed by our numerical solutions. Thus, the right-hand side of (3.3) vanishes.

During the quiescent periods of the Dimits-state evolution (i.e., between turbulent bursts), the turbulent momentum flux in the shear zones is negligible compared to the diffusive momentum flux, $\Pi_t \ll \Pi_d$. This is a consequence of the suppression of the $k_y \neq 0$ ITG modes by the zonal shear‡. We also find that the zonal temperature gradient $\partial_x \overline{T}$ is approximately constant in the quiescent shear zones (see Figure 8), so $\partial_x^2 \overline{T} = 0$. Therefore, (3.3) becomes

$$\partial_t \overline{\varphi} = a\chi \partial_x^2 \overline{\varphi}. \tag{3.5}$$

This is a diffusion equation governing the viscous decay of the ZFs with a collisional viscosity $a\chi$. As Figure 13 shows, quiescent periods of low heat flux and, thus, low levels of nonzonal perturbations, are correlated with the periods of decay of the zonal shear. We find that, despite the ever-present turbulence in the convection zones, where (3.5) does not hold, the decay rate of the zonal shear is closely approximated by the viscous

---

† The averaging is performed numerically over regions of near-uniform zonal shear, where, at every time step, we identify the radial locations of the uniform shear zones by applying the following conditions: $\partial_x^3 \overline{\varphi} < 0.1 \max\{\partial_x^3 \overline{\varphi}\}$ to isolate regions of near-uniform shear, and $\partial_x^2 \overline{\varphi} < 0.5 \max\{\partial_x^2 \overline{\varphi}\}$ to exclude the large variations of shear around the ZF extrema (see Figure 8).

‡ Note that Π_t is not small if there is turbulence present in the shear zones (which happens in the run up to and during turbulent bursts) — we shall investigate Π_t in Section 4.
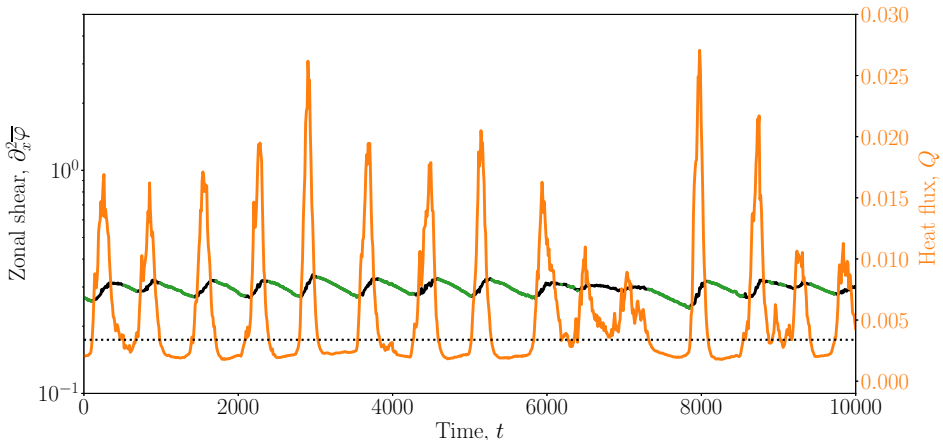
FIGURE 13. Heat flux Q (orange) and zonal shear $\partial_x^2\overline{\varphi}$ (black) vs. time for SimL. The highlighted (in green) sections of the zonal shear correspond to the quiescent periods. They are identified as those in which Q is smaller than a threshold value (dashed black line), defined as 60% of the time-averaged Q. The average decay rate of the zonal shear in the thus-identified quiescent periods is $\gamma \approx -3.5 \times 10^{-4}$ and the decay rate given by (3.6) is $\gamma_{\mathrm{s}} \approx -3.6 \times 10^{-4}$.

decay rate of the longest-wavelength ZF that comprises the zonal staircase, viz.,

$$\gamma_{\mathrm{s}} = -a\chi\left(\frac{2\pi n}{L_x}\right)^2, \tag{3.6}$$

where n is the number of periods of the zonal staircase in the domain of radial size L_x.

Let us now discuss what the ZF periodicity is.

3.1.3. *Scale of Zonal Flows*

In general, increasing/decreasing the radial extent of the integration domain by a factor increases/decreases the number n of shear zones by the same factor. This suggests that the characteristic length scale of the staircase, viz., the time-averaged radial separation of ZF extrema, is determined by a box-size-independent mechanism (see further discussion in Section 5). Ascertaining definitively whether this is the case is made difficult by the numerically observed time scales of convergence, which are at least an order of magnitude larger than the longest linear time scales, i.e., than the box-scale diffusion time $t_{\mathrm{box}} = (L_x/2\pi)^2/a\chi$ (see Figure 14). Note that the long-time evolution of the zonal profile and its length scale is not accompanied by a significant change in the average heat flux. In fact, the latter appears to reach saturation on time scales comparable to the box-scale diffusion time. Therefore, it is reasonable to trust the numerical values of the box-averaged heat flux (e.g., those shown in Figure 6), even though we could not be certain that the zonal profiles have reached ultimate saturation.

As we increase collisionality and thus move away from the near-marginal regime and into the collisionality-independent regime (the plateau seen in the bottom panel of Figure 6), the ZFs become more dynamic — they can merge, split and drift, as shown in Figure 15 (bottom panel). Here we focus on the near-marginal regime and the transition to strong turbulence, so this higher-collisionality regime will not be studied.

Even though the zonal staircase arises naturally from white-noise initial conditions for both the zonal and the nonzonal fields, its shape suggests initialising the ZFs with a "triangular" pattern. We find that this helps achieve more quickly a "less noisy" and
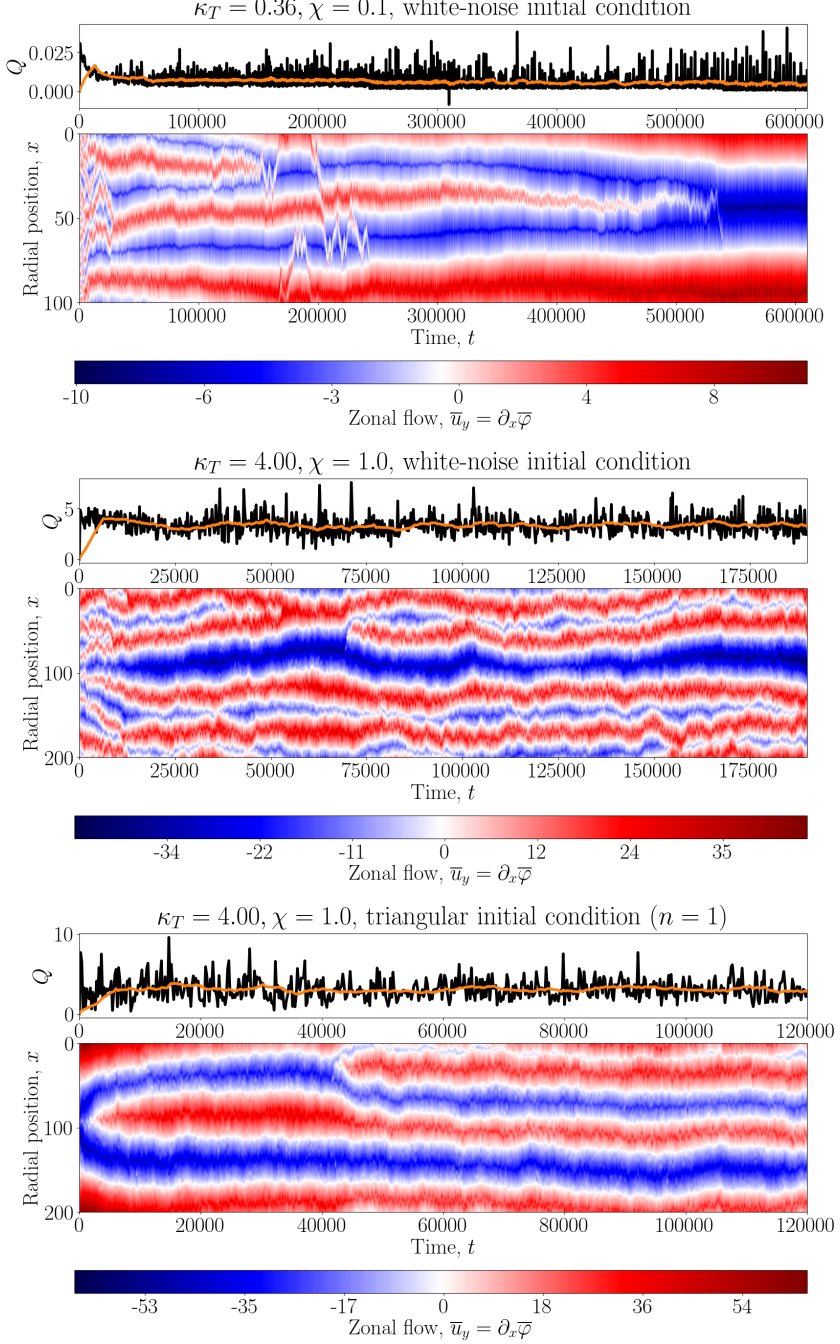
FIGURE 14. **Top:** Heat flux, raw (black) and rolling-averaged with a window of 6000 time units (orange) for $\kappa_T = 0.36$, $\chi = 0.1$, $L_x = 100$, $L_y = 100$ and 125×125 Fourier modes. This is the same as SimL, but we have used small-amplitude white noise as an initial condition. The diffusion time for the box-scale ZF is $t_{\text{box}} = (L_x/2\pi)^2/a\chi \approx 11200$. Convergence to a box-sized ZF occurs on a very long timescale ($> 50t_{\text{box}}$). **Middle:** Same as the top panel, but for $\kappa_T = 4$, $\chi = 1$, $L_x = 200$, and $L_y = 200$. Here $t_{\text{box}} = (L_x/2\pi)^2/a\chi \approx 4500$. **Bottom:** Same as the middle panel, but with a single-peak triangular ZF as an initial condition. It is evident that the ZF profile does not converge to a single peak.
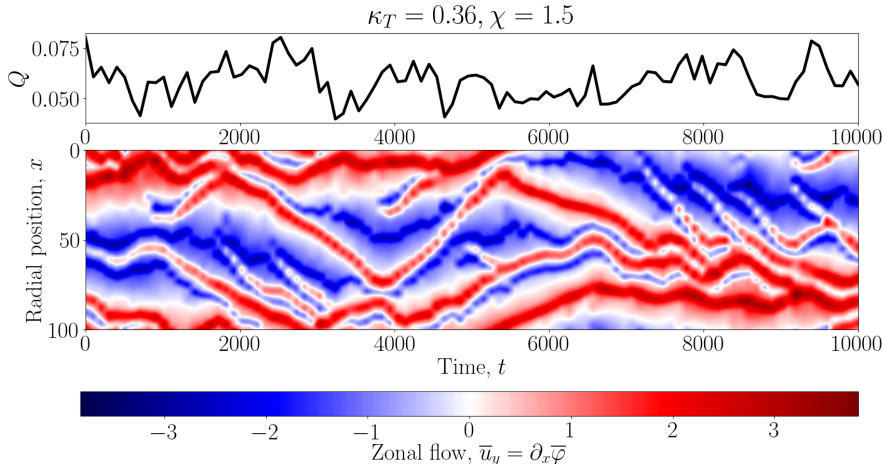
$\kappa_T = 0.36, \chi = 1.5$

FIGURE 15. Same as Figure 11, but for $\chi = 1.5$. The diffusion time for the box-scale ZF is $t_{\text{box}} = (L_x/2\pi)^2/a\chi \approx 750$. The locations of the ZF extrema of the staircase drift significantly over times comparable to t_{box}.

more symmetric final state, which is easier to handle both numerically and analytically. Of course, we do not know in advance how many "steps" the staircase will "choose" to have in the saturated state, so their number for the "triangular" initial condition is just an informed guess. Most results in this paper are from simulations that used such a triangular ZF initial condition, including SimL and SimH. Notable exceptions are Figures 6, 12, and 23, where we used data from many simulations, some with white-noise initial conditions and others with "triangular" ones.

3.2. *Convection Zones*

The convection zones located at the extrema of the ZFs contain localised patches of ITG turbulence and have a high radial turbulent heat conductivity (see Figures 8 and 9). The imposed equilibrium temperature gradient is flattened in the convection zones and slightly steepened in the shear zones. This results in a staircase-like radial temperature profile, shown in Figure 16. The turbulence in the convection zones is driven by a tertiary instability, localised by the zonal shear. In the low-collisionality, near-marginal regime, which we consider to be the most important, there is a qualitative difference between the way in which the tertiary instability operates at the ZF maxima and minima. A similar difference exists in both the Hasegawa-Mima equation (Zhu *et al.* 2018*b*) and gyrokinetics (McMillan *et al.* 2011).

3.2.1. *Turbulence at ZF Minima*

At the ZF minima, we find both ITG and Kelvin-Helmholtz tertiary instabilities. The former is dominant (faster) and saturates by producing a zonal-temperature gradient that cancels the background temperature gradient. This effectively decouples the evolution of the temperature perturbations from that of the electrostatic potential and leaves a KH mode that seems to determine the poloidal wavenumber at the ZF minima (the peak at $k_y \approx 0.26$ in Figure 7 is precisely the wavenumber of the fastest-growing KH mode at the ZF minima). Further details on the tertiary instability at the ZF minima and its saturation can be found in Appendix C.3.1.
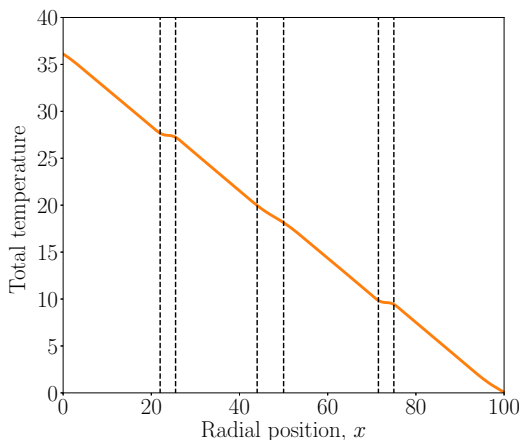
FIGURE 16. Time-averaged normalised total temperature (SimL), relative to the absolute temperature T_R at the right edge of the domain: $(\tau L_B/2\rho_s)\big(T_i - T_R + \overline{\delta T}\big) = (L_x - x)\kappa_T + \overline{T}$. A strong flattening of the gradient is visible around the ZF minima at $x \approx 25$ and 75, a weaker one around the ZF maximum at $x \approx 50$.

3.2.2. *Turbulence at ZF Maxima*

In contrast to the ZF minima, the regions around the ZF maxima cannot support a Kelvin-Helmholtz instability because the Rayleigh-Kuo criterion for instability is not satisfied there (Kuo 1949; Zhu *et al.* 2018*a*): see Appendix C.2. The ITG instability in these regions is significantly weaker than that at the ZF minima and does not appear to saturate in a similar fashion (by cancelling the equilibrium temperature gradient). The profile of the zonal-temperature gradient shown in Figure 8 suggests that the instability might not even be localised to the ZF maximum itself: there are two peaks of the zonal-temperature gradient visible on either side of the ZF maximum at $x \approx 47$ at $t = 10$. The poloidal scale of the modes at the ZF maxima is significantly longer than that at the ZF minima, see Appendix C.3.2.

Additionally, an asymmetric flattening of the zonal shear develops on one side of the ZF maximum, accompanied by a drift of the location of this maximum in the opposite direction (such a flattening is seen to the right of the central ZF maximum at $x \approx 47$ in Figure 8). Eventually, ferdinons are launched in the direction of the flattening (see also Section 3.3). This is likely due to the inability of the diminished zonal shear there to suppress the nonzonal perturbations. The burst of ferdinons causes the ZF maxima to change the direction of its drift and a flattening of the zonal shear develops on the opposite side. This causes an oscillation of the position of the ZF maximum, as seen in the bottom panel of Figure 11.

Thus, while turbulence is suppressed by zonal shear in the shear zones and by the cancellation of the equilibrum temperature gradient by the zonal temperature around the ZF minima, the regions around the ZF maxima remain locally unstable. As long as the zonal shear in the shear zones is strong enough to suppress turbulence, this instability is tamed, with any perturbations launched from the unstable regions into the shear zones unable to survive. Once the zonal shear decays below a certain level, it is no longer able to suppress these perturbations ("ferdinons", see Section 3.3) and a turbulent burst is initiated. Thus, the quasi-stationary zonal staircase contains the seeds of its own
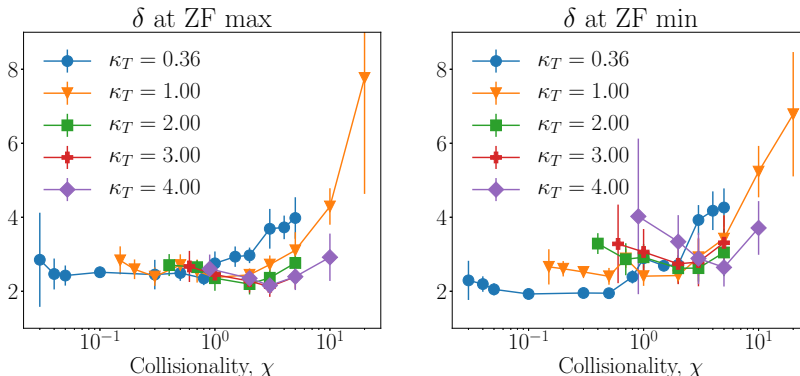
FIGURE 17. Numerically determined time-averaged values of δ [see (3.7)] at the global maximum (left) and global minimum (right) of the ZFs in the simulations whose heat flux is shown in Figure 6. We find no significant variation of δ with χ or κ_T, except at large χ.

destruction: the perilous combination of decaying ZFs and unstable convections zones around the ZF maxima.

3.2.3. *Scale of Convection Zones*

The width of the convection zones can be characterised by the quantity

$$\delta \equiv \sqrt{\frac{\partial_x \overline{\varphi}}{\partial_x^3 \overline{\varphi}}}. \tag{3.7}$$

Figure 17 shows that δ does not depend very strongly on either κ_T or χ, except far from the marginal state, where collisionality appears to smooth out the gradients in the convection zones and thus increase δ. This suggests that near the Dimits threshold, δ is an $\mathcal{O}(1)$ quantity in the normalised units of (2.17) and (2.18), i.e., it is equal to a few times the sound radius ρ_s.

3.3. *Ferdinons*

After the ZF has decayed sufficiently to weaken its ability to suppress perturbations, vortex-like propagating structures are spawned from the ZF maxima and drift radially through the shear zones. Strikingly similar structures — "ferdinons" — have been observed in GK simulations with imposed background flow shear (van Wyk *et al.* 2016, 2017). Figures 8 and 9 ($t = 136$) show a particular instance of the launching of ferdinons (see also Figure 10).

As the ferdinons smash into the turbulent modes in the convection zones at the ZF minima, more structures are produced and a burst of turbulence ensues (see Figure 8 and 9, $t = 250$). These travelling structures, as well as the resulting turbulence, cause a significant spike in the box-averaged heat flux (see Figure 10). As Figure 8 shows, they do not carry a significant ZF perturbation. They are created and propagate even if the ZF is held artificially constant in the numerical simulations. In other words, a localised ZF perturbation is not an essential part of these structures.

Figure 18 shows that ferdinons consist of a vortex dipole and a strong temperature perturbation trapped in one of the vortices of this dipole. There are ferdinons carrying both positive ("hot") and negative ("cold") temperature perturbations. Hot ferdinons drift towards the cooler (right) side of the domain, while the cold ones drift in the opposite direction, towards the hotter (left) side (see also Figure 10). The top and middle panels

of Figure 18 demonstrate that the direction of the drift does not depend on the sign of the zonal shear. Net flow circulation around the ferdinons is also independent of the sign of the shear — it is always anticlockwise for hot and clockwise for cold ones (see bottom panels of Figure 18).

Note that the ferdinons that emerge in our simple ITG model bear a striking qualitative resemblance to the avalanches reported by Villard *et al.* (2013) in global GK simulations, namely, they propagate both inwards and outwards, but always with a positive heat flux, and originate from the local maxima of the ZF. Simple soliton solutions have already been proposed as a model for GK avalanches (McMillan *et al.* 2009, 2018). Vortex-dipole solitons called "modons" have been investigated in Hasegawa-Mima-like models of turbulence (Horton & Hasegawa 1994). We do not yet know how and whether any of these are related to the ferdinons that we observe.

Let us discuss what we expect the ferdinon solution to be. Numerically, we find that the existence and propagation of these structures depend crucially on the two ITG-drive terms in (2.17) and (2.18), as well as on the nonlinear terms. In particular, the poloidal localisation of these structures is due to the nonzonal-nonzonal interactions. Indeed, we have found that (2.17) and (2.18) with the nonzonal-nonzonal nonlinear terms taken out (in what is sometimes referred to as the "quasilinear approximation"; see Srinivasan & Young 2012) has soliton solutions that are not localised poloidally, but rather appear to be of definite poloidal wavenumber k_y. Models have been proposed for structure formation in a sheared flow that rely on the tilting of turbulence by shear and a nonzero group velocity to produce moving structures (McMillan *et al.* 2009, 2018; Zhou *et al.* 2020). The radial group velocity is (at least in the Hasegawa-Mima-related models) proportional to the product $k_x k_y$†, which acquires a definite sign in the presence of flow shear (Section 4.3). However, we observe ferdinons moving in both radial directions in regions of definite zonal shear and, thus, definite radial group velocity. Therefore, at the moment, we consider it unlikely that the propagation of ferdinons can be explained using such group-velocity arguments. We leave the detailed investigation of ferdinon generation and propagation for future work.

Understanding ferdinons and their properties can also put an upper bound on the radial scale of the ZF. Indeed, our numerical simulations show that the ZFs can have a well-defined radial scale smaller than the box size (Section 3.1.3). This scale could perhaps be estimated via a causality argument — assuming that ferdinons, and, thus, turbulence, can only propagate a finite radial distance in a region of self-consistently evolving zonal shear, then an infinitely wide shear zone cannot be sustained for long. Note that finite-lifetime ferdinons over a dynamic ZF background, with which they can interact and gain or lose energy, are not in contradiction with the infinite-lifetime ferdinons seen by van Wyk *et al.* (2016), where a constant flow shear was imposed, and thus the shear profile was unable to react to the presence of ferdinons.

Once ferdinons are generated and turbulence develops in the shear zones, our analysis of the viscous decay of the zonal staircase in Section 3.1.2, which ignored the turbulent momentum flux, is no longer valid. Instead, we must focus on the effect of the turbulence on the ZFs. We find that the turbulence in the shear zones has a restoring effect on the zonal staircase in the Dimits regime, whereas beyond the Dimits threshold, it inhibits staircase formation.

---

† The radial group velocity $\partial\omega_{\boldsymbol{k}}/\partial k_x$ is proportional to $k_x k_y$ because $\omega_{\boldsymbol{k}} \propto k_y$ and $\omega_{\boldsymbol{k}}$ depends on k_x only through k^2.
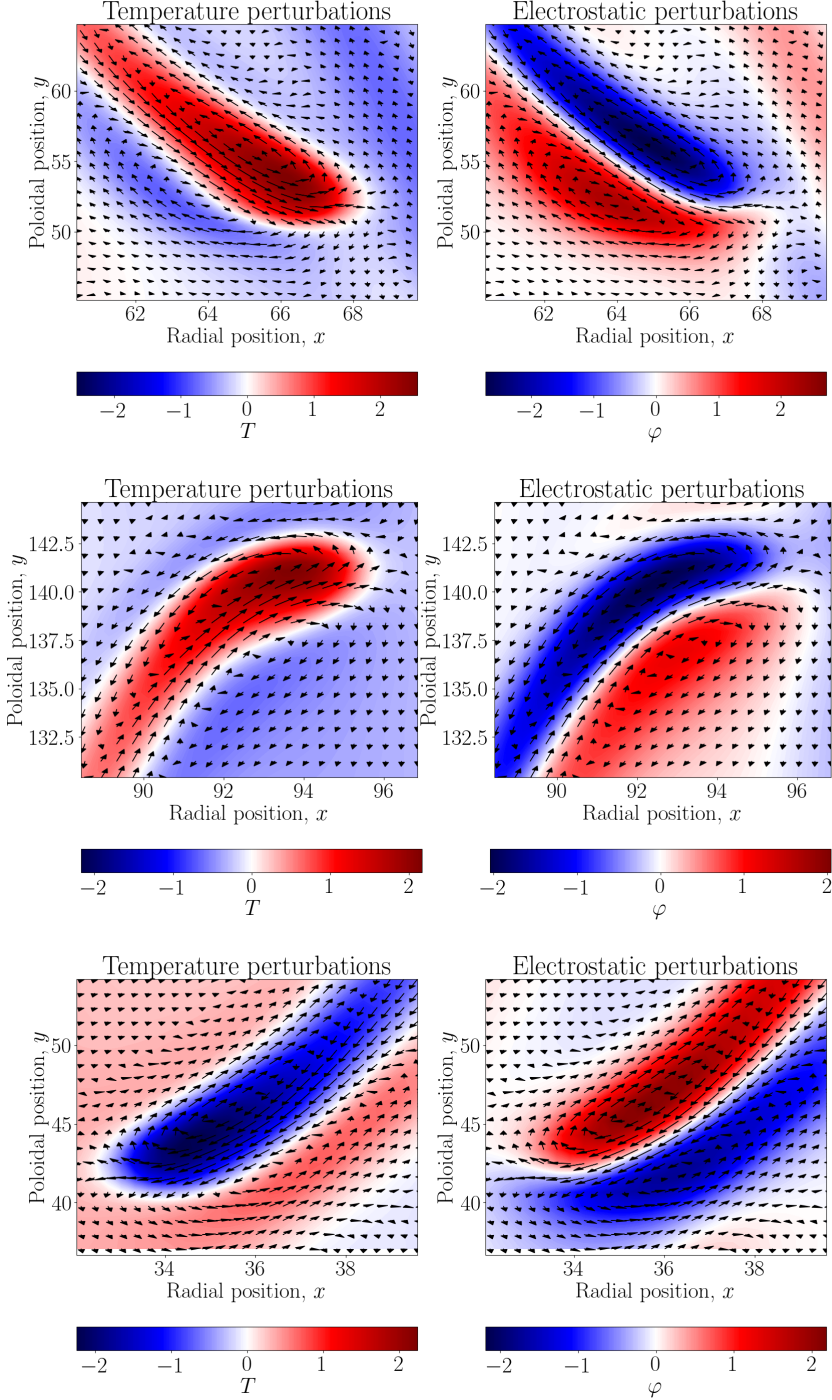
FIGURE 18. **Top:** A close-up view (from the top-right plot of Figure 9) of the temperature and electrostatic-potential perturbations of a hot ferdinon in a region of negative zonal shear, $S = \partial_x^2 \varphi < 0$. The arrows represent the local nonzonal $\boldsymbol{E} \times \boldsymbol{B}$ velocity $\boldsymbol{u}' = \hat{\boldsymbol{z}} \times \boldsymbol{\nabla} \varphi'$ (in arbitrary units). **Middle:** A hot ferdinon in a region of positive zonal shear $S = \partial_x^2 \varphi > 0$. **Bottom:** A cold ferdinon in a region of positive zonal shear $S = \partial_x^2 \varphi > 0$.

4. Resilience of the Zonal State and the Dimits Threshold

4.1. *Turbulent Momentum Flux*

In order to investigate the way in which ZF profiles are formed and maintained during turbulent periods, let us ask the following question: does ITG turbulence in the shear zones (produced in bursts) have a definite effect on the ZFs, and is it to oppose or feed them?

The ZF evolution equation (3.3) is

$$\partial_t \overline{\varphi} + \Pi_t + \Pi_d = 0. \tag{4.1}$$

In Section 3.1.2, we discussed the effect of the diffusive momentum flux Π_d, viz., the viscous decay of the ZFs. For the rest of this section, we focus on the effects of turbulence by examining the turbulent momentum flux Π_t.

It is evident from (4.1) that ZF saturation requires

$$\langle \Pi_t(t, x) + \Pi_d(t, x) \rangle_{\Delta t} \approx 0 \tag{4.2}$$

to be satisfied at every radial location x, where $\langle f(t) \rangle_{\Delta t} \equiv (1/\Delta t) \int_{\Delta t} dt f(t)$ is a time average in the saturated state over a time Δt longer than the typical evolution time of the ZF (e.g., longer than the duration of turbulent bursts if the saturated state is bursty). Recall that $\Pi_d \approx -a\chi \partial_x^2 \overline{\varphi} = -a\chi S$ in the shear zones (see Section 3.1.2). Therefore, within a shear zone with nearly constant (in time *and* in space) zonal shear, (4.2) tells us that the time-averaged turbulent momentum flux in that shear zone must have a definite value†, determined by the local zonal shear. Thus, in the saturated state, Π_t will be correlated with S.

To quantify this correlation, let us multiply both sides of (4.1) by $S = \partial_x^2 \overline{\varphi}$ and integrate across the radial extent of the domain. We find

$$\int_0^{L_x} dx \ (\partial_x^2 \overline{\varphi}) \partial_t \overline{\varphi} + \int_0^{L_x} dx \ \Pi_t S + \int_0^{L_x} dx \ \Pi_d S = 0. \tag{4.3}$$

Since $\Pi_d = -\chi \partial_x^2 (a\overline{\varphi} - b\overline{T}) = -a\chi S + b\chi \partial_x^2 \overline{T}$, we have

$$\int_0^{L_x} dx \ \Pi_d S = -a\chi \int_0^{L_x} dx \ S^2 + b\chi \int_0^{L_x} dx \ S \partial_x^2 \overline{T} \approx -a\chi \int_0^{L_x} dx \ S^2, \tag{4.4}$$

where we have assumed that the second term is negligible because the main contribution to S comes from the shear zones, where $\partial_x^2 \overline{T} \approx 0$ [see also the discussion leading to (3.5)]. Therefore, after integrating by parts the first term in (4.3) and time averaging the resulting equation, we find

$$\left\langle \int_0^{L_x} dx \ (\Pi_t S - a\chi S^2) \right\rangle_{\Delta t} = -\left\langle \frac{1}{2} \partial_t \int_0^{L_x} dx \ (\partial_x \overline{\varphi})^2 \right\rangle_{\Delta t} \approx 0. \tag{4.5}$$

This gives a prediction for the effective "turbulent viscosity" in the shear zones:

$$\nu_t \equiv -\frac{\left\langle \int_0^{L_x} dx \ \Pi_t S \right\rangle_{\Delta t}}{\left\langle \int_0^{L_x} dx \ S^2 \right\rangle_{\Delta t}} \approx -a\chi. \tag{4.6}$$

Relation (4.6) is, of course, corroborated by numerical simulations: see Figure 19.

---

† Also, the spatial average over a shear zone of the turbulent momentum flux must be nonzero. This is not in contradiction with the argument in Section 3.1.2 that the spatial average over the entire box is zero, viz., $\Pi = 0$, because a definite uniform zonal shear breaks the symmetry (3.4) locally within each shear zone.
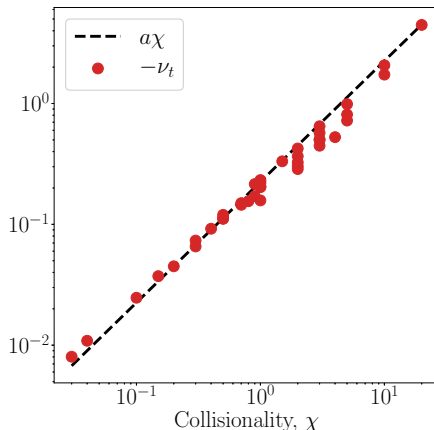
FIGURE 19. Comparison of estimated turbulent viscosity ν_t and the collisional viscosity $a\chi$ for the simulations from Figure 6. We see that $\nu_t \approx -a\chi$.

4.2. *Sign Reversal of the Turbulent Momentum Flux at the Dimits Threshold*

An important consequence of (4.6) is that, in a shear zone, the sign of the turbulent momentum flux must coincide with the sign of the zonal shear. Therefore, if, for certain parameters, sheared turbulence has a momentum flux with a sign *opposing* that of the local shear, saturation cannot be achieved. We shall see that this is exactly what happens beyond the Dimits threshold.

Let us investigate how turbulence responds to an imposed static zonal profile. We solve (2.17) and (2.18) numerically with an imposed static triangular ZF pattern (i.e., we do not evolve the ZFs at all), in a box of size $L_x = L_y = 100$ and 169×169 Fourier modes, for a range of parameters around the Dimits transition. The chosen ZF pattern in shown in Figure 20 (top panel) and is adjusted for every simulation so that the value of the zonal shear in the shear zones matches the largest ITG growth rate for that simulation. The chosen radial scale of the ZF ($= 100$) is inspired by the typical ZF scale that we observe in the low-collisionality regime, and is held fixed as we vary χ and κ_T. Then we calculate the effective turbulent viscosity ν_t associated with the turbulent momentum flux.

As the bottom panel of Figure 20 shows, we find a negative turbulent viscosity ν_t (and, thus, a positive correlation between local zonal shear and turbulent momentum flux) in the Dimits regime and a positive ν_t beyond it (thus, a negative correlation). Let us denote by κ_T^{static} the temperature gradient at which ν_t reverses its sign. The designation "*static*" reflects the fact that this is a numerical result for ITG turbulence with an artificially imposed *static* ZF profile. We find that the value of κ_T^{static} is insensitive to the exact shape of the ZF profile, and, most importantly, it nearly perfectly coincides with the Dimits threshold, i.e., $\kappa_T^{\text{static}} \approx \kappa_T^c$.

Thus, in the Dimits regime, shear zones are resilient because, when the zonal shear there decays due to viscosity and turbulence is thus unleashed, this turbulence acts to reinforce the ZFs and the zonal shear in the shear zones is restored to its turbulence-suppressing level. Beyond the Dimits regime, the zonal staircase cannot be sustained because both turbulence and collisional viscosity act to flatten out the ZFs.
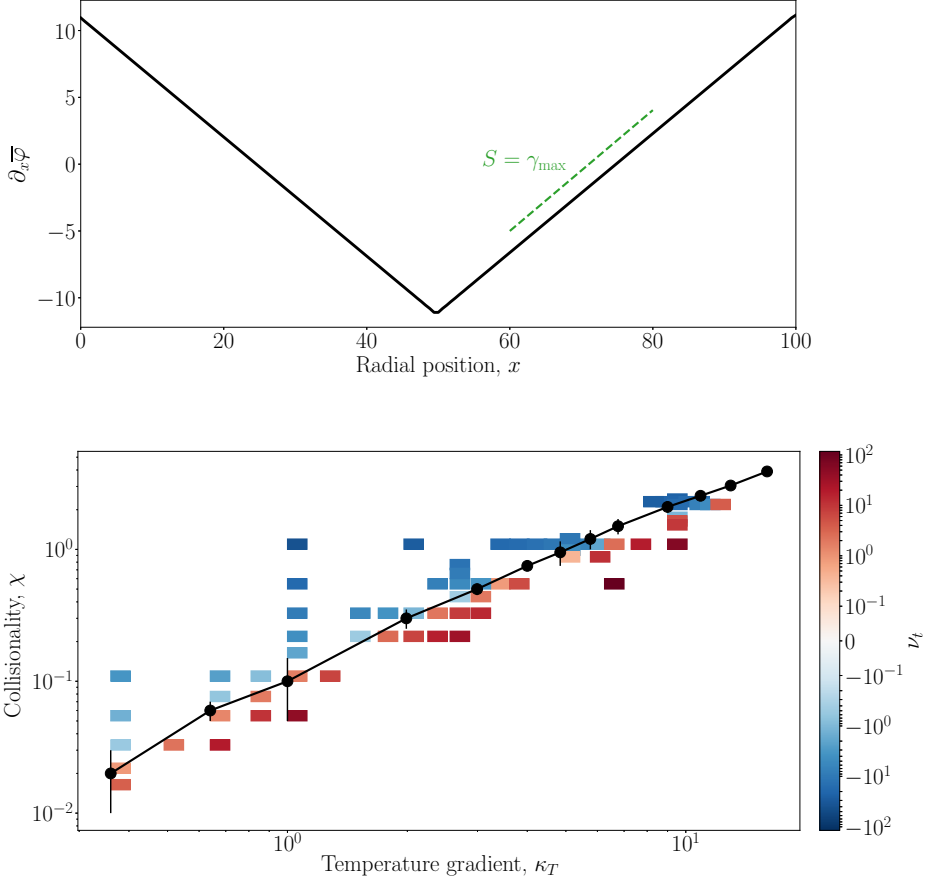
FIGURE 20. **Top:** The artificial triangular zonal pattern used to generate the bottom panel of this figure. The zonal shear in the shear zones is chosen to be equal in absolute value to the largest ITG growth rate (represented by the dashed green line). **Bottom:** The effective turbulent viscosity ν_t, as defined by (4.6), for the static triangular ZF profile given in the top panel of this figure (coloured data points). The black line represents the numerically established Dimits threshold.

4.3. *Reynolds Stress and Diamagnetic Stress*

Let us analyse what causes the turbulent momentum flux Π_t to reverse its sign at the Dimits transition. We split $\Pi_t = \Pi_\varphi + \Pi_T$ and define

$$\Pi_\varphi \equiv -\overline{(\partial_x\varphi)(\partial_y\varphi)} = \overline{u_y u_x}, \qquad \Pi_T \equiv -\overline{(\partial_x\varphi)(\partial_y T)} = \overline{u_y w_x}, \qquad (4.7)$$

where $\boldsymbol{u} = (-\partial_y\varphi, \partial_x\varphi)$ is the $\boldsymbol{E} \times \boldsymbol{B}$ flow and $\boldsymbol{w} = (-\partial_y T, \partial_x T)$ is the diamagnetic flow. Here Π_φ is the radial flux of the poloidal momentum due to the Reynolds stress of the $\boldsymbol{E} \times \boldsymbol{B}$ flow and the "diamagnetic stress" Π_T is a contribution to the momentum flux that physically arises due to the advection of the poloidal diamagnetic flow $w_y = \partial_x T$ by the radial $\boldsymbol{E} \times \boldsymbol{B}$ flow $u_x = -\partial_y\varphi$. To see this, let us take the zonal average of (2.18) and differentiate once with respect to x, to obtain an equation for the zonal diamagnetic flow:

$$\partial_t \overline{w}_y + \partial_x \overline{u_x w_y} - \partial_x \Pi_T = \chi \partial_x^2 \overline{w}_y. \qquad (4.8)$$
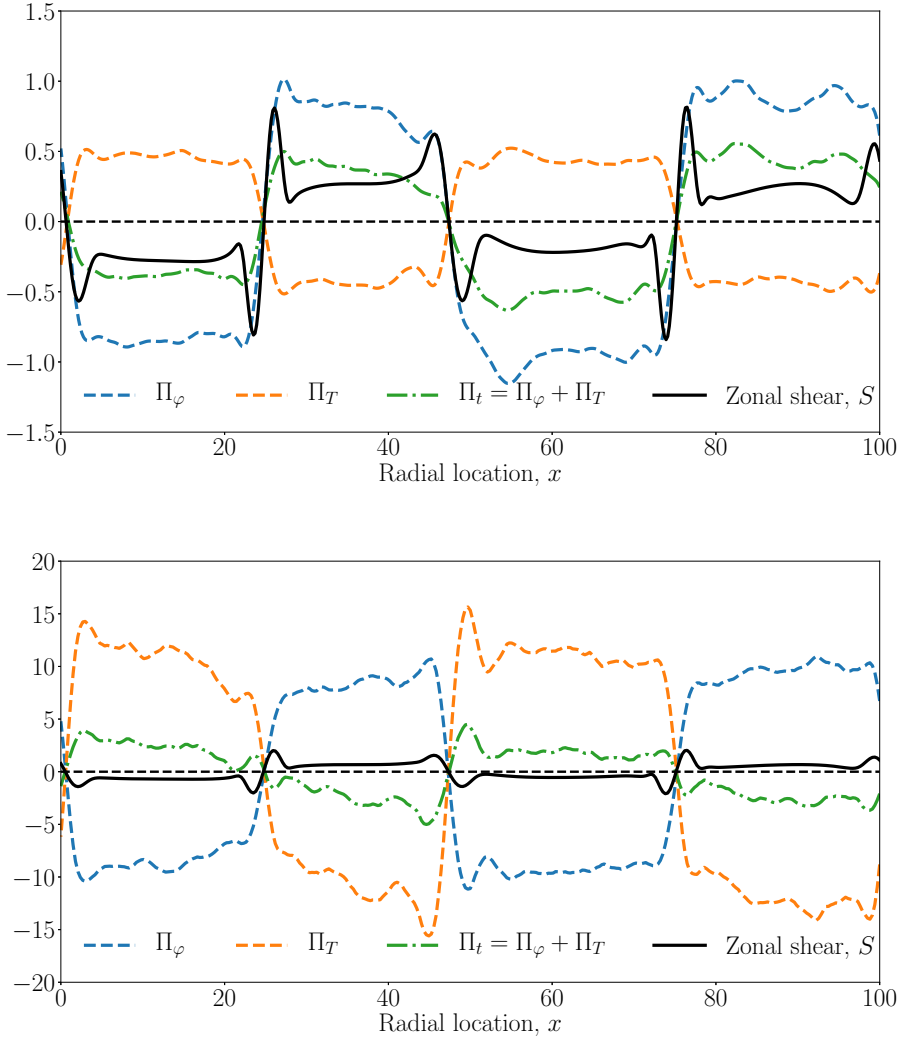
FIGURE 21. **Top:** Time-averaged momentum fluxes Π_φ, Π_T and $\Pi_t = \Pi_\varphi + \Pi_T$ for saturated ITG turbulence over a fixed zonal background. Note the correlation between the signs of the various fluxes and the zonal shear S: the sign of Π_t coincides with that of Π_φ and opposes the sign of S. This reflects that the temperature gradient is lower than the Dimits threshold, $\kappa_T = 0.36 < \kappa_T^c \approx 1$. The ZF profile used here was extracted from SimH at $t = 10$, but reduced by a factor of 0.8 in order to allow ITG turbulence to develop in the shear zones. **Bottom:** Same as the top panel, but with $\kappa_T = 1.21$. The sign of Π_t now opposes the sign of Π_φ and coincides with the sign of S. This reflects that $\kappa_T = 1.21 > \kappa_T^c \approx 1$. The extracted ZF is augmented by a factor of 2 to account for the increased ITG growth rate (due to the larger κ_T). This is necessary for the turbulence to saturate at numerically feasible amplitudes. Note that saturation is possible only because we have fixed the ZF profile. If the ZF is left to evolve according to (3.2), the poloidal momentum generated by the nonzonal perturbations flattens it and the system fails to reach a finite-amplitude saturated state (see Section 4.5).
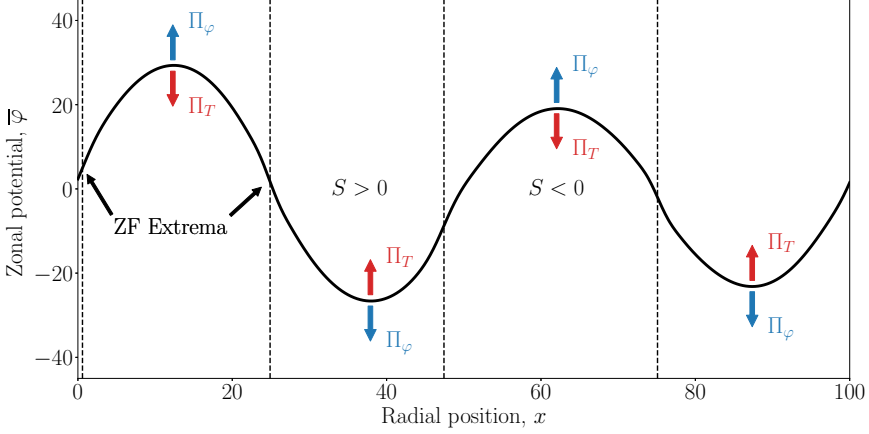
FIGURE 22. An illustration of the enhancing and suppressing effects of Π_φ and Π_T on the ZF. The black curve shows the $\overline{\varphi}(x)$ profile taken from SimH at $t = 10$. The ZF extrema are the locations where $S = \partial_x^2 \overline{\varphi} = 0$ (marked by dashed lines).

The evolution of the zonal poloidal $\boldsymbol{E} \times \boldsymbol{B}$ flow is described by (3.2), which can be recast as

$$\partial_t \overline{u}_y + \partial_x \overline{u_x u_y} + \partial_x \Pi_T = -\partial_x \Pi_d. \tag{4.9}$$

Added together, (4.8) and (4.9) describe the advection of the total poloidal flow ($\boldsymbol{E} \times \boldsymbol{B}$ + diamagnetic), $v_y \equiv u_y + w_y$, by the radial $\boldsymbol{E} \times \boldsymbol{B}$ flow:

$$\partial_t \overline{v}_y + \partial_x \overline{u_x v_y} = \text{dissipative terms}. \tag{4.10}$$

This makes physical sense because the diamagnetic flow is not a real flow and thus cannot advect anything.

The numerical solutions of (2.17) and (2.18) reveal that Π_φ and Π_T are in competition: on average, Π_φ has the same sign as the zonal shear $S = \partial_x^2 \overline{\varphi}$, while Π_T has the opposite sign. This is evident in Figure 21. Equation (4.1) then tells us that Π_φ feeds the ZFs by increasing the zonal potential $\overline{\varphi}$ in the shear zones of negative zonal shear (where $\overline{\varphi}$ is concave) and decreasing it in the shear zones of positive zonal shear (where $\overline{\varphi}$ is convex), whereas Π_T relaxes the ZFs by opposing Π_φ. Their combined effect either steepens or relaxes the ZF velocity $\overline{u}_y = \partial_x \overline{\varphi}$ at the turning points $\partial_x^2 \overline{\varphi} = 0$, depending on which stress is larger. Figure 22 is an illustration of this. This competition is crucial for the ability of the ZF to reconstitute itself after a turbulent burst and thus sets the threshold for the Dimits regime.

In order to assess what decides the outcome of this competition (i.e., the relative size of Π_φ and Π_T), let us consider how zonal shear affects ITG turbulence. For this purpose, consider a shear zone of radial extent d with a constant zonal shear $S = \partial_x^2 \overline{\varphi}$ throughout it. We can then perform the usual shearing-box change of variables $(t, x, y) \mapsto (\tilde{t}, \tilde{x}, \tilde{y})$, where

$$\tilde{t} = t, \; \tilde{x} = x, \; \tilde{y} = y - Stx. \tag{4.11}$$

This coordinate transformation eliminates the spatially inhomogeneous zonal-advection terms $(\partial_x \overline{\varphi})\partial_y = Sx\partial_y$ in (2.17) and (2.18). Consider a Fourier mode $\varphi, T \propto \exp(i\tilde{k}_x \tilde{x} + i\tilde{k}_y \tilde{y})$ in this shearing frame. In the laboratory frame (t, x, y), this mode has the form

$\varphi, T \propto \exp(ik_x x + ik_y y)$, where

$$k_x = \tilde{k}_x - St\tilde{k}_y, \ k_y = \tilde{k}_y. \tag{4.12}$$

Thus, the ZF shear introduces an effective drift of the laboratory-frame radial wavenumber†. The direction of this drift is given by the sign of S, viz., $S > 0$ gives rise to an anticorrelation of k_x and k_y, i.e., $k_x k_y < 0$, whereas for $S < 0$, $k_x k_y > 0$. Integrating the effect of Π_φ over the sheared region, we obtain

$$\frac{1}{d} \int dx \ \Pi_\varphi = -\frac{1}{dL_y} \int dxdy \ (\partial_x \varphi)(\partial_y \varphi) = -\sum_{\boldsymbol{k}} k_x k_y |\varphi_{\boldsymbol{k}}|^2. \tag{4.13}$$

Therefore, on average, Π_φ has the same sign as S, and, thus, feeds the ZFs that generate the shear zones‡.

We can write a similar expression for the diamagnetic stress:

$$\frac{1}{d} \int dx \ \Pi_T = -\sum_{\boldsymbol{k}} k_x k_y |\varphi_{\boldsymbol{k}}|^2 \mathrm{Re} \frac{T_{\boldsymbol{k}}}{\varphi_{\boldsymbol{k}}}. \tag{4.14}$$

Then the total turbulent momentum flux integrated over a shear region is

$$\frac{1}{d} \int dx \ \Pi_t = -\sum_{\boldsymbol{k}} k_x k_y |\varphi_{\boldsymbol{k}}|^2 \left(1 + \mathrm{Re} \frac{T_{\boldsymbol{k}}}{\varphi_{\boldsymbol{k}}}\right). \tag{4.15}$$

Recall that we already encountered the quantity $\mathrm{Re}(T_{\boldsymbol{k}}/\varphi_{\boldsymbol{k}})$ when dealing with the secondary instability in Section 2.8. There we found that $\mathrm{Re}(T_{\boldsymbol{k}}/\varphi_{\boldsymbol{k}}) < 0$ for all linearly unstable modes. Thus, linear theory predicts that Π_T and Π_φ are anti-correlated due to the negative sign of $\mathrm{Re}(T_{\boldsymbol{k}}/\varphi_{\boldsymbol{k}})$. Now let us perform a more detailed analysis of the linear modes and attempt to construct a model for the Dimits threshold based on it.

4.4. *Dimits Threshold from Linear Physics*

Using our knowledge of ITG perturbations in a region of uniform ZF shear, and of the turbulent momentum flux produced by them, we can make a heuristic linear-physics-based estimate for the Dimits threshold κ_T^c. In view of (4.15), it is given by the temperature gradient at which the *relevant* ITG modes have $\mathrm{Re}(T_{\boldsymbol{k}}/\varphi_{\boldsymbol{k}}) = -1$. By *relevant* we mean those ITG modes that dominate the turbulence in the shear zones. It is tempting to assume that these modes would be the most unstable modes in the system. This, however, cannot be the case because the most unstable modes are the radial streamers with $k_x = 0$, but the zonal shear that we find is comparable in magnitude to the largest ITG growth rate ($S \sim \gamma_{\mathrm{max}}$), and, therefore, is bound to break these streamers. Following the discussion in Section 4.3, we may assume that the typical ITG modes in sheared turbulence satisfy $k_x \sim \alpha k_y$, where $\alpha \sim S\tau_{\mathrm{nl}}$ characterises how tilted the mode is, τ_{nl} being the nonlinear correlation time of the turbulence in the shear zones. If $\tau_{\mathrm{nl}}^{-1} \sim \gamma_{\mathrm{max}} \sim S$, then $\alpha \sim 1$.

Thus, we assume that the relevant modes are tilted with $k_x = \alpha k_y$, where $\alpha \sim 1$ is an unknown tilt parameter that depends on the structure of the turbulence. We then look for the temperature gradient κ_T at which the fastest-growing ITG mode with $k_x = \alpha k_y$

---

† It is certainly true that an equilibrium shear would have such an effect on the turbulence. However, this is not guaranteed for ZFs. Their influence on the turbulence depends crucially on the modified electron response (2.4). This is a distinguishing feature of ion-scale physics that does not exist in, e.g., the electron-scale version of the model presented here.

‡ This is a well-known result in the context of Rossby-wave turbulence (see Vallis 2017, chapter 15.1.2)
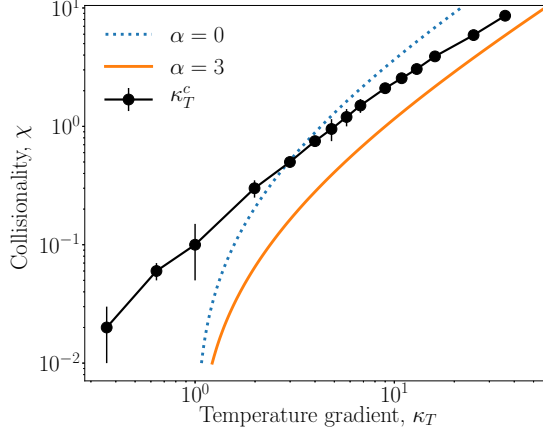
FIGURE 23. Comparison of numerical data to the analytical estimate for the threshold of the Dimits regime (Section 4.4). The black points represent the numerically observed κ_T^c. The other two curves correspond to the parameters for which the fastest-growing mode with $k_x = \alpha k_y$ satisfies $\mathrm{Re}(T_{\boldsymbol{k}}/\varphi_{\boldsymbol{k}}) = -1$. The case $\alpha = 0$ corresponds to the threshold for the suppression of the secondary instability of the fastest streamer ($k_x = 0$) mode (see Section 2.8), and the curve with $\alpha = 3$ asymptotes to the numerically determined slope of the Dimits threshold as $\kappa_T \to \infty$.

satisfies $\mathrm{Re}(T_{\boldsymbol{k}}/\varphi_{\boldsymbol{k}}) = -1$. This yields a prediction for the Dimits threshold in the (κ_T, χ) plane that we refer to as the "fastest-mode approximation". Note that there is no *a priori* reason to assume that α is itself not a function of κ_T.

4.4.1. *High-Collisionality Limit*

We can take the $\chi \to \infty$ limit of the fastest-mode approximation analytically. We formally order $\chi \sim \kappa_T$, as suggested by Figure 23. We then use the dispersion relation (2.20) to find the growth rate $\gamma_{\boldsymbol{k}}$ and real frequency $\omega_{\boldsymbol{k}}$ of the fastest mode with $k_x = \alpha k_y$. In Section 2.6.2, we showed that a mode of wavenumber $\boldsymbol{k}$ is unstable if and only if

$$\kappa_T k_y^2 > a\chi^2 k^6 = a\chi^2(1 + \alpha^2)^3 k_y^6, \tag{4.16}$$

so all unstable modes with $k_x = \alpha k_y$, where $\alpha \sim \mathcal{O}(1)$, satisfy

$$k_y^4 < \frac{\kappa_T}{a\chi^2(1 + \alpha^2)^3} \sim \mathcal{O}(\kappa_T^{-1}). \tag{4.17}$$

Similarly, using the results in Section 2.6.1 for the FLR bounds on the region of unstable wavenumbers, we find $k_y^4 \sim \mathcal{O}(\kappa_T^{-1})$ in the limit $\kappa_T \to \infty$. Thus, both mechanisms that bound the region of instability (and hence restrict the largest ITG growth rate), lead to the same scaling for the unstable wavenumbers. Therefore, the wavenumber of the most unstable mode must also satisfy $k_y \sim \kappa_T^{-1/4}$. Applying the ordering $\chi \sim \kappa_T \gg 1$ and $k_x = \alpha k_y \sim \kappa_T^{-1/4}$ to (2.28), we find

$$A \sim \mathcal{O}\left(\kappa_T^{1/2}\right), \qquad B \sim \mathcal{O}(1), \qquad C \sim \mathcal{O}\left(\kappa_T^{1/4}\right), \qquad f \sim \mathcal{O}(1), \qquad g \sim \mathcal{O}(1). \tag{4.18}$$

The unstable solution of (2.27) is

$$\gamma_{\boldsymbol{k}} - i\omega_{\boldsymbol{k}} = \frac{-A - B + iC + \sqrt{(A - B + iC)^2 + 4fAB - 4igAC}}{2}. \tag{4.19}$$

After expanding it using (4.18), we find

$$\gamma_{\boldsymbol{k}} \sim \mathcal{O}(1) \ll \omega_{\boldsymbol{k}} \sim \mathcal{O}\left(\kappa_T^{1/4}\right) \ll \chi k^2 \sim \mathcal{O}(\sqrt{\kappa_T}), \tag{4.20}$$

$$\omega_{\boldsymbol{k}} = -(1-g)C + \mathcal{O}\left(\kappa_T^{-1/4}\right) = -\kappa_T(1-b)(1+\alpha^2)k_y^3 + \mathcal{O}\left(\kappa_T^{-1/4}\right). \tag{4.21}$$

Therefore, (2.58) gives

$$\mathrm{Re}\frac{T_{\boldsymbol{k}}}{\varphi_{\boldsymbol{k}}} = \frac{k_y \kappa_T \omega_{\boldsymbol{k}}}{|\gamma_{\boldsymbol{k}} - i\omega_{\boldsymbol{k}} + \chi k^2|^2} \approx \frac{k_y \kappa_T \omega_{\boldsymbol{k}}}{\chi^2 k^4} \approx -\frac{1-b}{1+\alpha^2}\frac{\kappa_T^2}{\chi^2}. \tag{4.22}$$

Thus, the large-temperature-gradient fastest-mode approximation of the Dimits threshold is a straight line in the (κ_T, χ) plane, given by

$$\kappa_T^c \approx \chi\sqrt{\frac{1+\alpha^2}{1-b}}, \tag{4.23}$$

a posteriori confirming the ordering $\chi \sim \kappa_T$. The numerically determined Dimits threshold is indeed close to a straight line. Fitting the slope of that line to (4.23) yields $\alpha \approx 3$. Comparison of the prediction for the Dimits threshold for this value of α, as well as $\alpha = 0$, which corresponds to the threshold for the secondary instability of a primary streamer (as discussed in Section 2.8), can be found in Figure 23. The convergence is slow ($\propto \kappa_T^{-1/4}$), hence the sizeable discrepancy for the values of κ_T shown there, but we consider the asymptotic result to be sound.

4.4.2. *Low-Collisionality Limit*

Using a calculation that is nearly identical to the one in Section 2.8.2, we can analytically take the limit $\chi \rightarrow 0$ of the fastest-mode approximation using the collisionless dispersion relation (2.21) and inserting it into (2.54). We obtain that $\kappa_T^c \rightarrow 1$ as $\chi \rightarrow 0$ for the fastest mode with $k_x = \alpha k_y$, regardless of the value of α. This is a weakness of our "fastest-mode approximation" because the numerical data suggests instead that $\kappa_T^c \rightarrow 0$ as $\chi \rightarrow 0$. Thus, the assumptions that we made above about the relevance of the fastest-growing modes appear to be inadequate at low collisionality.

To summarise, the assumption that the momentum flux and $\mathrm{Re}(T_{\boldsymbol{k}}/\varphi_{\boldsymbol{k}})$ are dominated by the most unstable mode with some tilt given by $\alpha = k_x/k_y$ allows us to predict the Dimits threshold at high collisionality, but fails at low collisionality. This partial success is likely due to the fact that $\mathrm{Re}(T_{\boldsymbol{k}}/\varphi_{\boldsymbol{k}})$ for the most unstable mode is independent of $\boldsymbol{k}$ for $\kappa_T \sim \chi \gg 1$ [see (4.22)]. So, not only the most unstable, but in fact all modes in its vicinity will have the same value of $\mathrm{Re}(T_{\boldsymbol{k}}/\varphi_{\boldsymbol{k}})$. On the other hand, the failure of these assumptions at low collisionality suggests that we cannot use linear theory to predict the threshold there, but must rather focus on the nonlinear structure of the ITG turbulence seeded by ferdinons during bursts. This will be further discussed in Section 5.

4.5. *Beyond the Dimits Regime*

Beyond the Dimits threshold ($\kappa_T > \kappa_T^c$), our 2D system fails to reach saturation on a scale smaller than the domain size — perturbations grow exponentially and the box-sized streamer ($k_x = 0, k_y = 2\pi/L_y$) eventually dominates the spectrum. Figures 24 and 25 show that the large-scale, coherent ZFs that comprise the zonal staircase are quickly destroyed and never reappear. This is consistent with the illustration in Figure 22 and the discussion in Section 4.3. For $\kappa_T > \kappa_T^c$, if a shear zone of coherent zonal shear (like the ones we observe in the Dimits regime) were formed, the turbulent stress Π_t would flatten
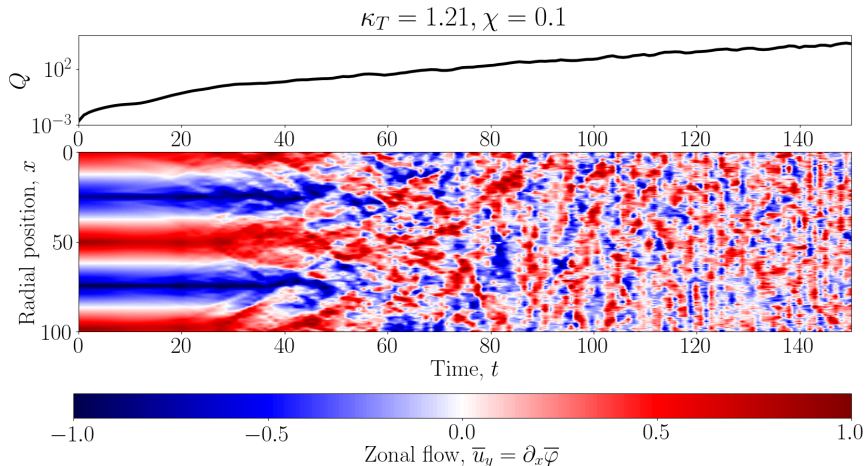
FIGURE 24. Time evolution of the heat flux (top) and ZF velocity $\overline{u}_y$ (bottom) beyond the Dimits regime. The ZF amplitude in the lower panel is normalised to a maximum of 1 at each time. The initial conditions are the same as those for SimH, but with an augmented temperature gradient $\kappa_T = 1.21 > \kappa_T^c \approx 1$ and a lower resolution of 337×167 Fourier modes due to the numerical cost of simulating the blow-up regime. The ZFs are quickly destroyed and large-scale ZFs never reappear, while the box-averaged heat flux Q grows exponentially. This state is eventually dominated by a streamer with a poloidal scale equal to that of the integration domain (see Figure 25).
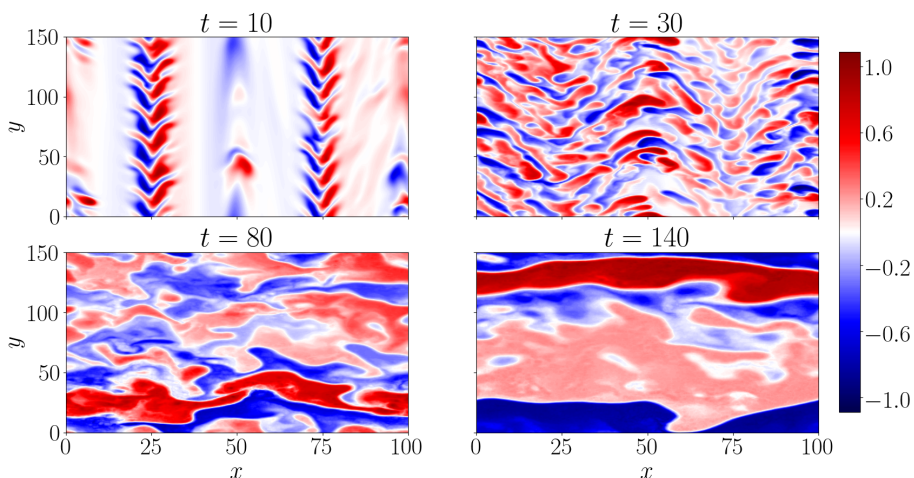

FIGURE 25. Snapshots of temperature perturbations in the blow-up state beyond the Dimits threshold (a movie is available in the supplementary material). The amplitudes are normalised to a maximum of 1 at each time. The initial conditions are the same as those for SimH, but with an augmented temperature gradient $\kappa_T = 1.21 > \kappa_T^c \approx 1$ and a lower resolution of 337×167 Fourier modes due to the numerical cost of simulating the blow-up regime.

out the ZF profile. Any coherent zonal shear is thus the harbinger of its own demise due to the momentum flux of the tilted turbulent eddies. The nonzonal perturbations grow exponentially, and so do the ZFs, but the latter are now dominated by small-scale time-incoherent zonal modes that are unable to quench the instability.

The lack of saturation beyond the Dimits regime in 2D is not surprising. GK simulations have shown that, beyond the Dimits regime in saturated 3D ITG turbulence, the ITG frequency at the injection ("outer") scale of the perpedicular plane is balanced by the parallel propagation time — the turbulence is in "critical balance" (Barnes *et al.* 2011):

$$\omega^{\mathrm{o}} \sim v_{\parallel} k_{\parallel}, \tag{4.24}$$

where $v_{\parallel}$ is some appropriate speed of parallel propagation (e.g., the ion thermal speed v_{ti}), and $\omega^{\mathrm{o}} \propto k_y$ is the ITG frequency at the outer scale, proportional to k_y by (2.20). In a tokamak, the smallest allowed value of $k_{\parallel}$ is $k_{\parallel} \sim L_{\parallel}^{-1}$, where $L_{\parallel}$ is the parallel connection length of the device. Thus, a parallel length scale is enforced by the magnetic geometry. The poloidal outer scale (k_y) then follows by (4.24) and the radial outer scale is enforced by zonal shearing ($k_x \sim k_y$). The 2D approximation can be obtained as the $k_{\parallel} \to 0$ limit of the 3D system. In this case, (4.24) implies that $k_y \to 0$, in agreement with the blow up dominated by the box-sized streamer that we observe beyond the Dimits threshold. Thus, the 2D approximation is fundamentally inadequate as a description of fully developed ITG turbulence†. However, we have shown that ZF-mediated saturation and the Dimits transition are captured by a 2D model. Of course, it is an outstanding task (left for future work) to confirm that the physics of the 2D Dimits transition remains (qualitatively) valid in 3D.

5. Discussion

We have found that the saturation of 2D ITG turbulence in Z-pinch geometry is mediated by strong quasi-static ZFs with patchwise-constant zonal shear (Section 3). There is a clear transition between a ZF-dominated Dimits regime and a strongly turbulent state, which in 2D fails to saturate at a finite amplitude (Section 4.5). The mechanism that sustains the ZFs in the Dimits regime ($\kappa_T < \kappa_T^c$) and undermines them beyond it ($\kappa_T > \kappa_T^c$) is linked to the turbulent momentum flux of ITG modes in the presence of a coherent zonal shear. Namely, in the Dimits regime, the response of ITG turbulence to strong (comparable to the ITG-instability growth rate), coherent zonal shear can be described in terms of a negative turbulent viscosity that reinforces the ZFs. This turbulent viscosity vanishes at the Dimits threshold and becomes positive beyond it, thus impeding any strong zonal shear that could suppress turbulence (Section 4.2). Viewed this way, the Dimits transition is caused by a change in the properties of sheared ITG turbulence. In the model considered here, the turbulent momentum flux consists of the usual Reynolds stress, familiar from hydrodynamics, and a diamagnetic contribution. We find that the former acts to reinforce the ZFs, while the latter opposes the ZFs (Section 4.3).

In general, therefore, determining whether a set of equilibrium parameters lies within the Dimits regime, requires one to make a statement about the combined momentum flux of all turbulent modes. In Section 4.4, we employed the heuristic assumption that the momentum flux is determined predominantly by the most unstable modes with a finite tilt ($k_x = \alpha k_y$). We found that $\alpha \approx 3$ models reasonably well the Dimits threshold

---

† Also because of the presence of 2D invariants (see Section 2.7), which can lead to an inverse cascade and energy pile-up at the largest available (box) scale, as they do in 2D hydrodynamic turbulence (Frisch 1995).

for large temperature gradients and collisionalities‡. At low collisionalities, such simple considerations do not produce quantitatively satisfactory results.

The mechanism for the Dimits transition described above is not directly tied to the onset of the radially localised tertiary instability found at the extrema of the ZF (Section 2.9 and Appendix C). It is more appropriate to view it as the quenching of some nonlinear version of the secondary instability (Section 2.8). This appears to be in contrast with the transition seen in the Hasegawa-Wakatani equations, where the tertiary instability was argued to determine the threshold for the strong-turbulence regime (Zhu *et al.* 2020)†.

As concluded in Section 4.4, an accurate prediction of the Dimits threshold requires a detailed understanding of the properties of sheared ITG turbulence. The nature of this turbulence is likely to be closely related to the properties of the localised structures ("ferdinons") that are seeded by the tertiary-unstable regions around the ZF maxima (see Section 3.2.2) and then drift through the shear zones (see Section 3.3). If the averaged properties of sheared turbulence correspond to those of a collection of (independent) ferdinons, we might be able to use the momentum flux of a ferdinon in order to make predictions about sheared turbulence. Developing an analytical approximation of a simple ferdinon is therefore an important outstanding task for future research.

The insight into the significance of the turbulent momentum flux of sheared turbulence provides us with a natural starting point for the investigation of the Dimits threshold in 3D gyrokinetics. To reiterate, we have showed that, within the Dimits regime, the time-averaged poloidal momentum flux of turbulence sheared by a region of constant zonal shear acts to reinforce the ZFs and, thus, zonal shear. On the other hand, beyond the Dimits regime, the overall sign of the momentum flux reverses and the ZFs are destroyed by the turbulence. The validity this statement is a well-posed question that can be answered by 3D gyrokinetic numerical simulations, regardless of the number of parameters (which, in a realistic scenario, greatly outnumbers two).

Acknowledgements

‡ Note that $\alpha \sim 1$ is consistent with a balance of zonal shear and turbulent turnover time ($S \sim \tau_{\mathrm{nl}}^{-1}$).

† The Hasegawa-Wakatani system does not contain the diamagnetic stress discussed in Section 4.3 or, indeed, any other form of poloidal momentum flux apart from the Reynolds stress. Thus, by construction, it cannot describe the effects discussed in this paper.

Appendix A. Derivation of the Model Equations

A.1. *Gyrokinetic Equation*

The derivation of our fluid model is similar to the one by Newton *et al.* (2010). We start from the 2D ion gyrokinetic (GK) equation in a Z-pinch-like equilibrium, as discussed in Section 2.1, omitting the parallel streaming term (for a review of gyrokinetics, see Abel *et al.* 2013). The ion distribution function is $f_i = F_i + \delta f_i$, where $F_i = n_i/(\pi^{3/2} v_{ti}^3) \exp(-v^2/v_{ti}^2)$ is the equilibrium ion distribution function and

$$\delta f_i = -\frac{Ze\phi}{T_i} F_i + h(t, \boldsymbol{R}, v_\parallel, v_\perp). \tag{A1}$$

The gyrocentre distribution h satisfies

$$\frac{\partial}{\partial t}\left(h - \frac{ZeF_i}{T_i}\langle\phi\rangle_R\right) + \boldsymbol{V}_D \cdot \frac{\partial h}{\partial \boldsymbol{R}} + \langle\boldsymbol{V}_E\rangle_R \cdot \left[\frac{\partial h}{\partial \boldsymbol{R}} - \hat{\boldsymbol{x}}\left(\frac{v^2}{v_{ti}^2} - \frac{3}{2}\right)\frac{F_i}{L_T}\right] = \langle C_l[h]\rangle_R. \tag{A2}$$

Here $T_i = m_i v_{ti}^2/2$ and $L_T^{-1} = -\partial_x \ln T_i$ are the ion equilibrium temperature and temperature gradient, respectively. The ion mass is m_i, the ion charge is Ze, and $\Omega_i = ZeB/m_ic$ is the ion gyrofrequency. $\boldsymbol{B}$ is the equilibrium magnetic field and $\langle.\rangle_r$ and $\langle.\rangle_R$ denote, respectively, the gyroaverages at fixed position and fixed ion guiding centre $\boldsymbol{R} = (X, Y, Z) = \boldsymbol{r} - \hat{\boldsymbol{b}} \times \boldsymbol{v}/\Omega_i$, where $\hat{\boldsymbol{b}} = \boldsymbol{B}/B$ is the unit vector parallel to the magnetic field. These gyroaverages are defined as

$$\langle f(\boldsymbol{R})\rangle_r \equiv \langle f(\boldsymbol{r} - \boldsymbol{\rho}(\theta))\rangle = \int_0^{2\pi} \frac{d\theta}{2\pi} \, f(\boldsymbol{r} - \boldsymbol{\rho}(\theta)), \tag{A3}$$

$$\langle f(\boldsymbol{r})\rangle_R \equiv \langle f(\boldsymbol{R} + \boldsymbol{\rho}(\theta))\rangle = \int_0^{2\pi} \frac{d\theta}{2\pi} \, f(\boldsymbol{R} + \boldsymbol{\rho}(\theta)), \tag{A4}$$

where $\langle.\rangle$ denotes the average with respect to the gyroangle θ, $\boldsymbol{\rho}(\theta) \equiv \boldsymbol{b} \times \boldsymbol{v}/\Omega_i$, $\boldsymbol{v} = v_\parallel \hat{\boldsymbol{b}} + v_\perp(\cos\theta \hat{\boldsymbol{y}} - \sin\theta \hat{\boldsymbol{x}})$, and the unit vectors $\{\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{b}}\}$ form a right-handed orthonormal basis, as shown in Figure 1. We will require the following properties of the gyroaverage:

$$\langle\boldsymbol{\rho}\rangle = 0, \qquad \langle\boldsymbol{\rho\rho}\rangle = \frac{1}{2}\frac{v_\perp^2}{v_{ti}^2}\rho_i^2 \mathbb{1}_\perp, \tag{A5}$$

where $\mathbb{1}_\perp$ is the identity matrix in the (x, y) plane. Using (A5) and Taylor expanding (A3) and (A4), we obtain

$$\langle h\rangle_r = \left[1 + \frac{1}{4}\frac{v_\perp^2}{v_{ti}^2}\rho_i^2\nabla_\perp^2 + \mathcal{O}(k_\perp^4\rho_i^4)\right]h(\boldsymbol{r}), \tag{A6}$$

$$\langle\varphi\rangle_R = \left[1 + \frac{1}{4}\frac{v_\perp^2}{v_{ti}^2}\rho_i^2\nabla_\perp^2 + \mathcal{O}(k_\perp^4\rho_i^4)\right]\varphi(\boldsymbol{R}). \tag{A7}$$

The velocities $\boldsymbol{V}_D$ and $\langle\boldsymbol{V}_E\rangle_R$ in (A2) are the magnetic and (gyroaveraged) $\boldsymbol{E} \times \boldsymbol{B}$ drifts, respectively, given by

$$\boldsymbol{V}_D = \frac{1}{\Omega}\left[v_\parallel^2\hat{\boldsymbol{b}} \times (\boldsymbol{b}.\nabla\boldsymbol{b}) + \frac{v_\perp^2}{2}\boldsymbol{b} \times \nabla\ln B\right] = \frac{1}{\Omega_i B}\left(v_\parallel^2 + \frac{v_\perp^2}{2}\right)\hat{\boldsymbol{b}} \times \boldsymbol{\nabla}B, \tag{A8}$$

$$\langle\boldsymbol{V}_E\rangle_R = \frac{c}{B}\hat{\boldsymbol{b}} \times \frac{\partial\langle\phi\rangle_R}{\partial\boldsymbol{R}}. \tag{A9}$$

The second equality in (A8) is obtained by assuming that the magnetic field is created by currents external to the spatial domain, so $\boldsymbol{\nabla} \times \boldsymbol{B} = 0$. For the Z-pinch geometry and

coordinates discussed in Section 2 and shown in Figure 1, we obtain

$$\boldsymbol{V}_D = -\frac{\rho_i v_{ti}}{L_B}\left(\frac{v_\parallel^2}{v_{ti}^2} + \frac{v_\perp^2}{2v_{ti}^2}\right)\hat{\boldsymbol{y}}. \qquad (A\,10)$$

The term on the right-hand side of (A 2) is the gyroaveraged linearised Landau collision operator. To lowest order in the mass-ratio expansion ($m_e/m_i \ll 1$), only ion-ion collisions contribute and the linearised operator is given by

$$C_l[h] = \frac{\nu_i v_{ti}^3}{n_i}\frac{\partial}{\partial \boldsymbol{v}}\bigg|_{\boldsymbol{r}} \cdot \left\{ F_i(\boldsymbol{v})\int d^3\,\boldsymbol{v}'F_i(\boldsymbol{v}')\boldsymbol{U}\cdot\left[\frac{\partial}{\partial\boldsymbol{v}}\bigg|_{\boldsymbol{r}}\frac{h(\boldsymbol{v})}{F_i(\boldsymbol{v})} - \frac{\partial}{\partial\boldsymbol{v}'}\bigg|_{\boldsymbol{r}}\frac{h(\boldsymbol{v}')}{F_i(\boldsymbol{v}')}\right]\right\}, \quad (A\,11)$$

where $\boldsymbol{U} = \left(u^2\boldsymbol{I} - \boldsymbol{u}\boldsymbol{u}\right)/u^3$, $\boldsymbol{u} = \boldsymbol{v} - \boldsymbol{v}'$, and

$$\nu_i = \frac{2\pi Z^4 e^4 n_i \ln \varLambda_i}{m_i^2 v_{ti}^3} \qquad (A\,12)$$

is the ion-ion collision frequency, where $\ln \varLambda_i$ is the Coulomb logarithm (Helander & Sigmar 2002).

The electrostatic gyrokinetic equation (A 2) is closed by the quasineutrality condition $\delta n_e = Z\delta n_i$ and the modified adiabatic electron response (Abel & Cowley 2013)

$$\delta f_e = \frac{e\phi'}{T_e}F_e. \qquad (A\,13)$$

Therefore,

$$\frac{1}{n_i}\int d^3\boldsymbol{v}\,\langle h\rangle_r = \frac{Ze}{T_i}\phi + \frac{e}{T_e}\phi'. \qquad (A\,14)$$

Putting all of this together, we arrive at the 2D electrostatic gyrokinetic system in our Z-pinch equilibrium:

$$\frac{\partial}{\partial t}(h - \langle\varphi\rangle_R F_i) + \frac{\rho_i v_{ti}}{2L_T}\left(\frac{v^2}{v_{ti}^2} - \frac{3}{2}\right)F_i\frac{\partial\langle\varphi\rangle_R}{\partial Y} - \frac{\rho_i v_{ti}}{L_B}\left(\frac{v_\parallel^2}{v_{ti}^2} + \frac{v_\perp^2}{2v_{ti}^2}\right)\frac{\partial h}{\partial Y}$$

$$+ \frac{1}{2}\rho_i v_{ti}\{\langle\varphi\rangle_R, h\} = \langle C_l[h]\rangle_R, \qquad (A\,15)$$

$$\frac{1}{n_i}\int d^3\boldsymbol{v}\,\langle h\rangle_r = \varphi + \tau\varphi', \qquad (A\,16)$$

where the normalised (to ion units) electric potential is $\varphi = Ze\phi/T_i$ and the temperature ratio is $\tau = T_i/ZT_e$.

A.2. *Lowest-Order Solution*

We now apply the high-collisionality, long-wavelength, cold-ion ordering (2.8). In this expansion, we write $h = h^{(0)} + h^{(1)}$, where $\mathcal{O}\big(h^{(1)}\big) \sim \mathcal{O}\big(k_\perp^2\rho_i^2 h^{(0)}\big)$. Then, to lowest order, equation (A 15) gives $C_l[h^{(0)}] = 0$, whose solution is a perturbed Maxwellian

$$h^{(0)}(\boldsymbol{R}) = \left[\frac{\delta N(\boldsymbol{R})}{n_i} + \frac{\delta T(\boldsymbol{R})}{T_i}\left(\frac{v^2}{v_{ti}^2} - \frac{3}{2}\right)\right]F_i. \qquad (A\,17)$$

Here the abstract quantities $\delta N(\boldsymbol{R})$ and $\delta T(\boldsymbol{R})$ are taken to be functions of the guiding centre $\boldsymbol{R}$. Substituting the expansion for h into the quasineutrality equation (A 16) and

using (A 6), we obtain

$$\frac{1}{n_i} \int d^3\boldsymbol{v} \left(1 + \frac{1}{4}\frac{v_\perp^2}{v_{ti}^2}\rho_i^2\nabla_\perp^2\right)\left[\frac{\delta N}{n_i} + \frac{\delta T}{T_i}\left(\frac{v^2}{v_{ti}^2} - \frac{3}{2}\right)\right]F_i = \varphi + \tau\varphi', \tag{A 18}$$

where we have absorbed the density and temperature moments of h into $h^{(0)}$ by imposing

$$\int d^3\boldsymbol{v}\ h^{(1)} = \int d^3\boldsymbol{v}\ v^2 h^{(1)} = 0. \tag{A 19}$$

Formally, in writing down (A 18), we have only assumed that the density moment of $h^{(1)}$ vanishes. We will use the condition that $h^{(1)}$ has a zero temperature moment in Appendix B.

Under the ordering $\tau \sim k_\perp^2 \rho_i^2 \ll 1$, (A 18) yields

$$\frac{\delta N}{n_i} = \varphi + \tau\varphi' - \frac{1}{4}\rho_i^2\nabla_\perp^2\left(\varphi + \frac{\delta T}{T_i}\right) + \mathcal{O}(k_\perp^4\rho_i^4\varphi). \tag{A 20}$$

We now proceed to take density and temperature moments of the GK equation (A 15) at fixed particle position $\boldsymbol{r}$ and retain only the lowest-order terms in the ordering (2.8). We will find that the density moment of (A 15) vanishes to lowest order, hence we are required to expand that moment to order $\mathcal{O}(k_\perp^2\rho_i^2 h)$. We shall only require terms up to $\mathcal{O}(h)$ for the temperature moment.

A.3. *Density Moment*

Let us consider one-by-one the density moments at fixed particle position, $(1/n_i) \int d^3\boldsymbol{v}\ \langle.\rangle_r$, of the terms in (A 15). The first term is

$$\frac{1}{n_i} \int d^3\boldsymbol{v}\ \langle h - \langle\varphi\rangle_R F_i\rangle_r = \varphi + \tau\varphi' - \left(1 + \frac{1}{2}\rho_i^2\nabla_\perp^2\right)\varphi = \tau\varphi' - \frac{1}{2}\rho_i^2\nabla_\perp^2\varphi, \tag{A 21}$$

where we have used quasineutrality (A 16) and the lowest-nontrivial-order expressions for the gyroaverages (A 6) and (A 7).

The next term is

$$\begin{aligned}
&\frac{1}{n_i} \int d^3\boldsymbol{v}\ \frac{\rho_i v_{ti}}{2L_T}\left(\frac{v^2}{v_{ti}^2} - \frac{3}{2}\right)F_i\left\langle\frac{\partial\langle\varphi\rangle_R}{\partial Y}\right\rangle_r \\
&\approx \frac{1}{n_i} \int d^3\boldsymbol{v}\ \frac{\rho_i v_{ti}}{2L_T}\left(\frac{v^2}{v_{ti}^2} - \frac{3}{2}\right)F_i\left(1 + \frac{1}{2}\frac{v_\perp^2}{v_{ti}^2}\rho_i^2\nabla_\perp^2\right)\frac{\partial\varphi}{\partial y} \\
&= \frac{\rho_i v_{ti}}{2L_T}\frac{1}{2}\rho_i^2\nabla_\perp^2\frac{\partial\varphi}{\partial y}.
\end{aligned} \tag{A 22}$$

To lowest order in $k_\perp^2 \rho_i^2 \ll 1$, the magnetic-drift term gives

$$\begin{aligned}
&\frac{1}{n_i} \int d^3\boldsymbol{v}\ \frac{\rho_i v_{ti}}{L_B}\left(\frac{v_\parallel^2}{v_{ti}^2} + \frac{v_\perp^2}{2v_{ti}^2}\right)\left\langle\frac{\partial h(\boldsymbol{R})}{\partial Y}\right\rangle_r \\
&\approx \frac{1}{n_i} \int d^3\boldsymbol{v}\ \frac{\rho_i v_{ti}}{L_B}\left(\frac{v_\parallel^2}{v_{ti}^2} + \frac{v_\perp^2}{2v_{ti}^2}\right)\frac{\partial h^{(0)}(\boldsymbol{r})}{\partial y} \\
&= \frac{1}{n_i} \int d^3\boldsymbol{v}\ \frac{\rho_i v_{ti}}{L_B}\frac{2}{3}\frac{v^2}{v_{ti}^2}\frac{\partial h^{(0)}(\boldsymbol{r})}{\partial y} = \frac{\rho_i v_{ti}}{L_B}\frac{\partial}{\partial y}\left(\varphi + \frac{\delta T}{T_i}\right),
\end{aligned} \tag{A 23}$$

where we have used

$$\int dv \, v_\parallel^2 h^{(0)} = \int dv \, \frac{v_\perp^2}{2} h^{(0)} = \int dv \, \frac{v^2}{3} h^{(0)}, \tag{A 24}$$

which is a consequence of the isotropic form (A 17) of $h^{(0)}$.

The density moment of the nonlinear term in (A 15) is

$$\frac{1}{n_i} \int d^3\boldsymbol{v} \, \langle \{ \langle \varphi \rangle_R(\boldsymbol{R}), h(\boldsymbol{R}) \} \rangle_r$$

$$\approx \frac{1}{n_i} \int d^3\boldsymbol{v} \, \left\langle \left\{ \varphi(\boldsymbol{R}) + \frac{1}{4} \frac{v_\perp^2}{v_{ti}^2} \rho_i^2 \nabla_\perp^2 \varphi(\boldsymbol{R}), h(\boldsymbol{R}) \right\} \right\rangle_r$$

$$\approx \frac{1}{n_i} \int d^3\boldsymbol{v} \, \left\langle \left\{ \varphi(\boldsymbol{r}) - \boldsymbol{\rho} \cdot \boldsymbol{\nabla}\varphi(\boldsymbol{r}) + \frac{1}{2}\boldsymbol{\rho}\boldsymbol{\rho} : \boldsymbol{\nabla}\boldsymbol{\nabla}\varphi(\boldsymbol{r}) + \frac{1}{4} \frac{v_\perp^2}{v_{ti}^2} \rho_i^2 \nabla_\perp^2 \varphi(\boldsymbol{r}), h(\boldsymbol{r}-\boldsymbol{\rho}) \right\} \right\rangle$$

$$= \frac{1}{n_i} \int d^3\boldsymbol{v} \, \left\langle \{ \varphi(\boldsymbol{r}), h(\boldsymbol{r}-\boldsymbol{\rho}) \} + \boldsymbol{\rho}\boldsymbol{\rho} : \{ \boldsymbol{\nabla}\varphi(\boldsymbol{r}), \boldsymbol{\nabla}h(\boldsymbol{r}) \} \right.$$

$$\left. + \frac{1}{2}\boldsymbol{\rho}\boldsymbol{\rho} : \{ \boldsymbol{\nabla}\boldsymbol{\nabla}\varphi(\boldsymbol{r}), h(\boldsymbol{r}) \} + \frac{1}{4} \frac{v_\perp^2}{v_{ti}^2} \{ \rho_i^2 \nabla_\perp^2 \varphi(\boldsymbol{r}), h(\boldsymbol{r}) \} \right\rangle$$

$$= \frac{1}{n_i} \int d^3\boldsymbol{v} \, \left[\{ \varphi(\boldsymbol{r}), \langle h \rangle_r(\boldsymbol{r}) \} + \langle \boldsymbol{\rho}\boldsymbol{\rho} \rangle_r : \{ \boldsymbol{\nabla}\varphi(\boldsymbol{r}), \boldsymbol{\nabla}h(\boldsymbol{r}) \} \right.$$

$$\left. + \frac{1}{2}\langle \boldsymbol{\rho}\boldsymbol{\rho} \rangle_r : \{ \boldsymbol{\nabla}\boldsymbol{\nabla}\varphi(\boldsymbol{r}), h(\boldsymbol{r}) \} + \frac{1}{4} \frac{v_\perp^2}{v_{ti}^2} \{ \rho_i^2 \nabla_\perp^2 \varphi(\boldsymbol{r}), h(\boldsymbol{r}) \} \right]$$

$$= \frac{1}{n_i} \int d^3\boldsymbol{v} \, \left[\{ \varphi(\boldsymbol{r}), \langle h \rangle_r(\boldsymbol{r}) \} + \frac{1}{2} \frac{v_\perp^2}{v_{ti}^2} \rho_i^2 \mathbb{1}_\perp : \{ \boldsymbol{\nabla}\varphi(\boldsymbol{r}), \boldsymbol{\nabla}h(\boldsymbol{r}) \} \right.$$

$$\left. + \frac{1}{2} \frac{v_\perp^2}{v_{ti}^2} \{ \rho_i^2 \nabla_\perp^2 \varphi(\boldsymbol{r}), h(\boldsymbol{r}) \} \right], \tag{A 25}$$

where we have used (A 5). Using the lowest-order contribution (A 17) to h and the fact that $\{g, g\} = 0$ for any g, we find that (A 25) becomes

$$\{ \varphi, \tau\varphi' \} + \frac{1}{2}\rho_i^2 \mathbb{1}_\perp : \left\{ \boldsymbol{\nabla}\varphi, \boldsymbol{\nabla}\frac{\delta T}{T_i} \right\} + \frac{1}{2}\rho_i^2 \left\{ \nabla_\perp^2 \varphi, \varphi + \frac{\delta T}{T_i} \right\}$$

$$= \left\{ \varphi, \tau\varphi' - \frac{1}{2}\rho_i^2 \nabla_\perp^2 \varphi \right\} + \frac{1}{2}\rho_i^2 \boldsymbol{\nabla}_\perp \cdot \left\{ \boldsymbol{\nabla}_\perp \varphi, \frac{\delta T}{T_i} \right\}. \tag{A 26}$$

Finally, collecting terms, dividing by τ and introducing the ion sound radius $\rho_s = \rho_i/\sqrt{2\tau}$, we obtain

$$\frac{\partial}{\partial t}(\varphi' - \rho_s^2 \nabla_\perp^2 \varphi) - \frac{\rho_i v_{ti}}{\tau L_B} \frac{\partial}{\partial y}\left(\varphi + \frac{\delta T}{T_i} \right) + \frac{\rho_i v_{ti}}{2L_T} \frac{\partial}{\partial y}(\rho_s^2 \nabla_\perp^2 \varphi)$$

$$+ \frac{1}{2}\rho_i v_{ti}\left(\{ \varphi, \varphi' - \rho_s^2 \nabla_\perp^2 \varphi \} + \rho_s^2 \boldsymbol{\nabla}_\perp \cdot \left\{ \boldsymbol{\nabla}_\perp \varphi, \frac{\delta T}{T_i} \right\} \right) = \frac{1}{\tau n_i} \int d^3\boldsymbol{v} \, \langle \langle C_l[h] \rangle_R \rangle_r. \tag{A 27}$$

This will become (2.9) after we calculate the collisional term in Appendix A.5.3.

A.4. *Temperature Moment*

In a similar way, let us consider the temperature moments, $(1/n_i) \int d^3\boldsymbol{v} \, v^2/v_{ti}^2 \langle . \rangle_r$, of the terms in (A 15) to lowest order in $k_\perp^2 \rho_i^2 \ll 1$. The first term is

$$\frac{1}{n_i} \int d^3\boldsymbol{v} \, \frac{v^2}{v_{ti}^2} \langle h - \langle \varphi \rangle_R F_i \rangle_r \approx \frac{1}{n_i} \int d^3\boldsymbol{v} \, \frac{v^2}{v_{ti}^2} \left(h^{(0)} - \varphi F_i \right) = \frac{3}{2} \frac{\delta T}{T_i}, \qquad (\text{A } 28)$$

where we have used (A 17).

The temperature-gradient term is

$$\frac{1}{n_i} \int d^3\boldsymbol{v} \, \frac{\rho_i v_{ti}}{2L_T} \frac{v^2}{v_{ti}^2} \left(\frac{v^2}{v_{ti}^2} - \frac{3}{2} \right) F_i \left\langle \frac{\partial \langle \varphi \rangle_R}{\partial Y} \right\rangle_r$$

$$\approx \frac{1}{n_i} \int d^3\boldsymbol{v} \, \frac{\rho_i v_{ti}}{2L_T} \frac{v^2}{v_{ti}^2} \left(\frac{v^2}{v_{ti}^2} - \frac{3}{2} \right) F_i \frac{\partial \varphi}{\partial y} = \frac{3}{2} \frac{\rho_i v_{ti}}{2L_T} \frac{\partial \varphi}{\partial y}. \qquad (\text{A } 29)$$

The magnetic-drift term is

$$\frac{1}{n_i} \int d^3\boldsymbol{v} \, \frac{\rho_i v_{ti}}{L_B} \frac{v^2}{v_{ti}^2} \left(\frac{v_\parallel^2}{v_{ti}^2} + \frac{v_\perp^2}{2v_{ti}^2} \right) \left\langle \frac{\partial h(\boldsymbol{R})}{\partial Y} \right\rangle_r$$

$$\approx \frac{1}{n_i} \int d^3\boldsymbol{v} \, \frac{\rho_i v_{ti}}{L_B} \frac{v^2}{v_{ti}^2} \left(\frac{v_\parallel^2}{v_{ti}^2} + \frac{v_\perp^2}{2v_{ti}^2} \right) \frac{\partial h^{(0)}(\boldsymbol{r})}{\partial y}$$

$$= \frac{1}{n_i} \int d^3\boldsymbol{v} \, \frac{\rho_i v_{ti}}{L_B} \frac{2}{3} \frac{v^4}{v_{ti}^4} \frac{\partial h^{(0)}(\boldsymbol{r})}{\partial y} = \frac{5}{2} \frac{\rho_i v_{ti}}{L_B} \frac{\partial}{\partial y} \left(\varphi + 2\frac{\delta T}{T_i} \right), \qquad (\text{A } 30)$$

where we have used the isotropy of $h^{(0)}$ again. By the ordering (2.8), this term is an order $L_T/L_B \sim \mathcal{O}\left(k_\perp^2 \rho_i^2\right) \ll 1$ smaller than the temperature-gradient term (A 29). Hence it will not contribute to the final expression for the temperature moment of (A 15).

The nonlinear term is

$$\frac{1}{n_i} \int d^3\boldsymbol{v} \, \frac{v^2}{v_{ti}^2} \langle \{\langle \varphi \rangle_R, h\} \rangle_r \approx \frac{1}{n_i} \int d^3\boldsymbol{v} \, \frac{v^2}{v_{ti}^2} \left\{ \varphi, h^{(0)} \right\} = \left\{ \varphi, \frac{\delta T}{T_i} \right\}. \qquad (\text{A } 31)$$

Collecting terms, we find that the temperature moment of (A 15) is

$$\frac{\partial}{\partial t} \frac{\delta T}{T_i} + \frac{\rho_i v_{ti}}{2L_T} \frac{\partial \varphi}{\partial y} + \frac{1}{2} \rho_i v_{ti} \left\{ \varphi, \frac{\delta T}{T_i} \right\} = \frac{2}{3n_i} \int d^3\boldsymbol{v} \, \frac{v^2}{v_{ti}^2} \langle \langle C_l[h] \rangle_R \rangle_r. \qquad (\text{A } 32)$$

This will become (2.10) after we calculate the collisional term in Appendix A.5.2.

A.5. *Moments of the Collision Operator*

A.5.1. *Gyroaveraged Collision Operator*

The gyroaveraged collision term in (A 2), expanded in a Fourier basis, $h(\boldsymbol{R}) = \sum_{\boldsymbol{k}} h_{\boldsymbol{k}} e^{i\boldsymbol{k} \cdot \boldsymbol{R}}$, is

$$\langle C_l[h] \rangle_R = \sum_{\boldsymbol{k}} \langle C_l[h_{\boldsymbol{k}} e^{i\boldsymbol{k} \cdot \boldsymbol{R}}] \rangle_R = \sum_{\boldsymbol{k}} \langle C_l[h_{\boldsymbol{k}} e^{-i\boldsymbol{k} \cdot \boldsymbol{\rho}}] e^{i\boldsymbol{k} \cdot \boldsymbol{r}} \rangle_R$$

$$= \sum_{\boldsymbol{k}} \langle C_l[h_{\boldsymbol{k}} e^{-i\boldsymbol{k} \cdot \boldsymbol{\rho}}] e^{i\boldsymbol{k} \cdot \boldsymbol{\rho}} \rangle e^{i\boldsymbol{k} \cdot \boldsymbol{R}}, \qquad (\text{A } 33)$$

where all derivatives and integrals with respect to $\boldsymbol{v}$ in the collision operator are taken at fixed $\boldsymbol{r}$, and the $\langle . \rangle$ operation is the gyroangle average [see (A 3) and (A 4)]. In order to obtain collisional terms in (A 27) and (A 32), we expand the exponential factors $e^{\pm i\boldsymbol{k} \cdot \boldsymbol{\rho}}$ in

the small quantity $\boldsymbol{k} \cdot \boldsymbol{\rho}$. Recall that (A 27) and (A 32) are contained at different orders in the GK equation. We need (A 33) to order $\mathcal{O}(k_\perp^2 \rho_i^2 \nu_i h)$ for the temperature moment and to order $\mathcal{O}(k_\perp^4 \rho_i^4 \nu_i h)$ for the density moment. We first consider the collisional term in the temperature equation (A 32), which represents thermal diffusion, before turning to the more involved calculation for the viscosity in (A 27).

A.5.2. *Collisional Thermal Diffusion*

The collisional term in the temperature equation (A 32) is

$$\frac{2}{3n_i}\left\langle \sum_{\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \int d^3\boldsymbol{v} \, \frac{v^2}{v_{ti}^2} \langle C_l[h_{\boldsymbol{k}} e^{-i\boldsymbol{k}\cdot\boldsymbol{\rho}}] e^{i\boldsymbol{k}\cdot\boldsymbol{\rho}} \rangle \right\rangle_r$$

$$\approx \frac{2}{3n_i} \sum_{\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \int d^3\boldsymbol{v} \, \frac{v^2}{v_{ti}^2} \langle C_l[h_{\boldsymbol{k}} e^{-i\boldsymbol{k}\cdot\boldsymbol{\rho}}] e^{i\boldsymbol{k}\cdot\boldsymbol{\rho}} \rangle, \tag{A 34}$$

where, to order $\mathcal{O}(k_\perp^2 \rho_i^2 \nu_i h)$, we can expand

$$\langle C_l[h_{\boldsymbol{k}} e^{-i\boldsymbol{k}\cdot\boldsymbol{\rho}}] e^{i\boldsymbol{k}\cdot\boldsymbol{\rho}} \rangle = C_l[h_{\boldsymbol{k}}^{(1)}] + \left\langle i\boldsymbol{k}\cdot\boldsymbol{\rho} C_l[(-i\boldsymbol{k}\cdot\boldsymbol{\rho})h_{\boldsymbol{k}}^{(0)}] \right\rangle$$

$$- \frac{1}{2}\left\langle (\boldsymbol{k}\cdot\boldsymbol{\rho})^2 \right\rangle C_l[h_{\boldsymbol{k}}^{(0)}] + \left\langle C_l\left[-\frac{1}{2}(\boldsymbol{k}\cdot\boldsymbol{\rho})^2 h_{\boldsymbol{k}}^{(0)}\right] \right\rangle + \mathcal{O}(k_\perp^4 \rho_i^4 \nu_i h). \tag{A 35}$$

Taking a temperature moment of (A 35) annihilates the first and fourth terms due to the conservation-of-energy property of the collision operator ($\int d^3\boldsymbol{v} \, v^2 \, C_l[f] = 0$ for any f) and the third term vanishes because $C_l[h^{(0)}] = 0$. Just as in Newton *et al.* (2010), after performing the integration in (A 34), we find that the second term of (A 35) gives

$$\int d^3\boldsymbol{v} \, \frac{v^2}{v_{ti}^2}(i\boldsymbol{k}\cdot\boldsymbol{\rho})C_l[(-i\boldsymbol{k}\cdot\boldsymbol{\rho})h_{\boldsymbol{k}}^{(0)}] = -\frac{3}{2}n_i\chi k_\perp^2 \frac{\delta T_{\boldsymbol{k}}}{T_i}, \tag{A 36}$$

where

$$\chi \equiv \frac{8}{9}\sqrt{\frac{2}{\pi}}\nu_i\rho_i^2 \tag{A 37}$$

and the ion-ion collision frequency ν_i is defined in (A 12). Thus, (A 34) is

$$-\chi \sum_{\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} k_\perp^2 \frac{\delta T_{\boldsymbol{k}}}{T_i} = \chi \nabla_\perp^2 \frac{\delta T(\boldsymbol{r})}{T_i}. \tag{A 38}$$

The right-hand side of (A 32) is (A 38), hence we arrive at the temperature equation (2.10).

A.5.3. *Collisional Viscous Damping*

As we discussed before, under the ordering (2.8), the density moment of (A 15) is obtained at an order higher in $k_\perp^2 \rho_i^2 \ll 1$ than the temperature moment. As a sanity check, note that taking a density moment ($\int d^3\boldsymbol{v} \, \langle . \rangle_r$) of (A 35) annihilates all terms: terms one and four vanish because the collision operator conserves particle number ($\int d^3\boldsymbol{v} \, C_l[f] = 0$ for any f), term two vanishes because of conservation of momentum ($\int d^3\boldsymbol{v} \, \boldsymbol{v} C_l[f] = 0$ for any f), and, as before, term three is identically zero due to the form of $h^{(0)}$. We thus need to expand (A 33) to order $\mathcal{O}(k_\perp^4 \rho_i^4 \nu_i h)$.

The collisional term of the density moment in (A 27) is

$$
\sum_{\boldsymbol{k}} \int d^3\boldsymbol{v} \; \langle\langle C_l\big[h_{\boldsymbol{k}}e^{-i\boldsymbol{k}\cdot\boldsymbol{\rho}}\big]e^{i\boldsymbol{k}\cdot\boldsymbol{\rho}}\rangle e^{i\boldsymbol{k}\cdot\boldsymbol{R}}\rangle_r = \sum_{\boldsymbol{k}} \int d^3\boldsymbol{v} \; J_0\!\left(\frac{k_\perp v_\perp}{\Omega}\right)\langle C_l\big[h_{\boldsymbol{k}}e^{-i\boldsymbol{k}\cdot\boldsymbol{\rho}}\big]e^{i\boldsymbol{k}\cdot\boldsymbol{\rho}}\rangle e^{i\boldsymbol{k}\cdot\boldsymbol{r}}
$$

$$
= \sum_{\boldsymbol{k}} \int d^3\boldsymbol{v} \; J_0\!\left(\frac{k_\perp v_\perp}{\Omega}\right) C_l\big[h_{\boldsymbol{k}}e^{-i\boldsymbol{k}\cdot\boldsymbol{\rho}}\big]e^{i\boldsymbol{k}\cdot\boldsymbol{\rho}}e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \equiv \sum_{\boldsymbol{k}} C_{D\boldsymbol{k}}e^{i\boldsymbol{k}\cdot\boldsymbol{r}}. \tag{A 39}
$$

Note that any term with an odd power of $\boldsymbol{k}\cdot\boldsymbol{\rho}$ is annihilated by gyroaveraging. Expanding (A 39) to order $\mathcal{O}(k_\perp^4\rho_i^4\nu_i h)$, we find

$$
C_{D\boldsymbol{k}} = \int d^3\boldsymbol{v} \left\{ -\frac{1}{2}(\boldsymbol{k}\cdot\boldsymbol{\rho})^2 C_l\!\left[-\frac{1}{2}(\boldsymbol{k}\cdot\boldsymbol{\rho})^2 h_{\boldsymbol{k}}^{(0)}\right] \right.
$$

$$
-\frac{1}{2}(\boldsymbol{k}\cdot\boldsymbol{\rho})^2 C_l\!\left[h_{\boldsymbol{k}}^{(1)}\right] - \frac{1}{6}i(\boldsymbol{k}\cdot\boldsymbol{\rho})^3 C_l\!\left[(-i\boldsymbol{k}\cdot\boldsymbol{\rho})h_{\boldsymbol{k}}^{(0)}\right]
$$

$$
\left. -\frac{k_\perp^2 v_\perp^2}{4\Omega^2}\left(\boldsymbol{k}\cdot\boldsymbol{\rho}\,C_l\!\left[\boldsymbol{k}\cdot\boldsymbol{\rho}h_{\boldsymbol{k}}^{(0)}\right] - \frac{1}{2}C_l\!\left[(\boldsymbol{k}\cdot\boldsymbol{\rho})^2 h_{\boldsymbol{k}}^{(0)}\right] + C_l\!\left[h_{\boldsymbol{k}}^{(1)}\right]\right) \right\}. \tag{A 40}
$$

Since the linearised collision operator is isotropic in velocity space and the gyrokinetic distribution h is gyroangle independent, it follows that $C[h^{(1)}]$ is also gyroangle independent. Therefore, we can write

$$
\int d^3\boldsymbol{v} \; \frac{1}{2}(\boldsymbol{k}\cdot\boldsymbol{\rho})^2 C_l\!\left[h_{\boldsymbol{k}}^{(1)}\right] = \int d^3\boldsymbol{v} \; \frac{1}{2}\langle(\boldsymbol{k}\cdot\boldsymbol{\rho})^2\rangle C_l\!\left[h_{\boldsymbol{k}}^{(1)}\right] = \int d^3\boldsymbol{v} \; \frac{k_\perp^2 v_\perp^2}{4\Omega^2}C_l\!\left[h_{\boldsymbol{k}}^{(1)}\right]. \tag{A 41}
$$

Substituting (A 41) into (A 40), we obtain

$$
C_{D\boldsymbol{k}} = \int d^3\boldsymbol{v} \left\{ \frac{1}{4}(\boldsymbol{k}\cdot\boldsymbol{\rho})^2 C_l\!\left[(\boldsymbol{k}\cdot\boldsymbol{\rho})^2 h_{\boldsymbol{k}}^{(0)}\right] - \frac{k_\perp^2 v_\perp^2}{2\Omega^2}C_l\!\left[h_{\boldsymbol{k}}^{(1)}\right] - \frac{1}{6}(\boldsymbol{k}\cdot\boldsymbol{\rho})^3 C_l\!\left[\boldsymbol{k}\cdot\boldsymbol{\rho}h_{\boldsymbol{k}}^{(0)}\right] \right.
$$

$$
\left. -\frac{k_\perp^2 v_\perp^2}{4\Omega^2}\boldsymbol{k}\cdot\boldsymbol{\rho}\,C_l\!\left[\boldsymbol{k}\cdot\boldsymbol{\rho}h_{\boldsymbol{k}}^{(0)}\right] + \frac{k_\perp^2 v_\perp^2}{8\Omega^2}C_l\!\left[(\boldsymbol{k}\cdot\boldsymbol{\rho})^2 h_{\boldsymbol{k}}^{(0)}\right] \right\}. \tag{A 42}
$$

The high-collisionality limit allowed us to obtain the form of $h^{(0)}$, but $h^{(1)}$ is unknown without inverting the collision operator. We can take advantage of the form of the GK equation (A 15) to lowest non-trivial order and express directly the required moment of $C[h^{(1)}]$ using moments of $C[h^{(0)}]$. Consider the $\int d^3\boldsymbol{v}(v_\perp^2 - 2v^2/3)$ moment of the GK equation up to order $\mathcal{O}(k_\perp^2\rho_i^2\nu_i h)$. In Appendix A.4, where we took a v^2 moment, we found that all non-collisional terms in the temperature moment of (A 15) are isotropic in velocity space to order $\mathcal{O}(k_\perp^2\rho_i^2\nu_i h)$, hence they are annihilated by the operator $\int d^3\boldsymbol{v}(v_\perp^2 - 2v^2/3)$. Using the expansion (A 35) of the collision operator, we find

$$
0 = \int d^3\boldsymbol{v} \left(v_\perp^2 - \frac{2}{3}v^2\right)\left\{ \boldsymbol{k}\cdot\boldsymbol{\rho}\,C_l\!\left[\boldsymbol{k}\cdot\boldsymbol{\rho}h_{\boldsymbol{k}}^{(0)}\right] - \frac{1}{2}C_l\!\left[(\boldsymbol{k}\cdot\boldsymbol{\rho})^2 h_{\boldsymbol{k}}^{(0)}\right] + C_l\!\left[h_{\boldsymbol{k}}^{(1)}\right] \right\}
$$

$$
= \int d^3\boldsymbol{v} \left\{ v_\perp^2\boldsymbol{k}\cdot\boldsymbol{\rho}\,C_l\!\left[\boldsymbol{k}\cdot\boldsymbol{\rho}h_{\boldsymbol{k}}^{(0)}\right] - \frac{1}{2}v_\perp^2 C_l\!\left[(\boldsymbol{k}\cdot\boldsymbol{\rho})^2 h_{\boldsymbol{k}}^{(0)}\right] \right.
$$

$$
\left. + v_\perp^2 C_l\!\left[h_{\boldsymbol{k}}^{(1)}\right] - \frac{2}{3}v^2\boldsymbol{k}\cdot\boldsymbol{\rho}\,C_l\!\left[\boldsymbol{k}\cdot\boldsymbol{\rho}h_{\boldsymbol{k}}^{(0)}\right] \right\}. \tag{A 43}
$$

Here a few terms have dropped out due to the energy-conservation properties of the

collision operator. Extracting from (A 43) an expression for $v_\perp^2 C_l\left[h_{\bm{k}}^{(1)}\right]$ and substituting this expression into (A 42), we obtain an expression for $C_{D\bm{k}}$ involving only $h^{(0)}$:

$$
\begin{aligned}
C_{D\bm{k}} = \int d^3\bm{v} \Bigg\{ &\frac{1}{4}(\bm{k}\cdot\bm{\rho})^2 C_l\left[(\bm{k}\cdot\bm{\rho})^2 h_{\bm{k}}^{(0)}\right] - \frac{1}{6}(\bm{k}\cdot\bm{\rho})^3 C_l\left[\bm{k}\cdot\bm{\rho} h_{\bm{k}}^{(0)}\right] \\
&+ \frac{k_\perp^2 v_\perp^2}{4\Omega^2}\bm{k}\cdot\bm{\rho}\, C_l\left[\bm{k}\cdot\bm{\rho} h_{\bm{k}}^{(0)}\right] - \frac{k_\perp^2 v^2}{3\Omega^2}\bm{k}\cdot\bm{\rho}\, C_l\left[\bm{k}\cdot\bm{\rho} h_{\bm{k}}^{(0)}\right] \\
&- \frac{k_\perp^2 v_\perp^2}{8\Omega^2} C_l\left[(\bm{k}\cdot\bm{\rho})^2 h_{\bm{k}}^{(0)}\right] \Bigg\}.
\end{aligned}
\tag{A 44}
$$

We proceed to evaluate these integrals:

$$
\int d^3\bm{v}\, \frac{1}{4}(\bm{k}\cdot\bm{\rho})^2 C_l\left[(\bm{k}\cdot\bm{\rho})^2 h_{\bm{k}}^{(0)}\right] = -n_i\chi\rho_i^2 k_\perp^4\left(\frac{3}{20}\varphi_{\bm{k}} + \frac{3}{16}\frac{\delta T_{\bm{k}}}{T_i}\right),
\tag{A 45}
$$

$$
\int d^3\bm{v}\, \frac{1}{6}(\bm{k}\cdot\bm{\rho})^3 C_l\left[\bm{k}\cdot\bm{\rho} h_{\bm{k}}^{(0)}\right] = -n_i\chi\rho_i^2 k_\perp^4\, \frac{3}{20}\frac{\delta T_{\bm{k}}}{T_i},
\tag{A 46}
$$

$$
\int d^3\bm{v}\, \frac{k_\perp^2 v_\perp^2}{4\Omega^2}(\bm{k}\cdot\bm{\rho}) C_l\left[\bm{k}\cdot\bm{\rho} h_{\bm{k}}^{(0)}\right] = -n_i\chi\rho_i^2 k_\perp^4\, \frac{3}{10}\frac{\delta T_{\bm{k}}}{T_i},
\tag{A 47}
$$

$$
\int d^3\bm{v}\, \frac{k_\perp^2 v^2}{3\Omega^2}(\bm{k}\cdot\bm{\rho}) C_l\left[\bm{k}\cdot\bm{\rho} h_{\bm{k}}^{(0)}\right] = -n_i\chi\rho_i^2 k_\perp^4\, \frac{1}{2}\frac{\delta T_{\bm{k}}}{T_i},
\tag{A 48}
$$

$$
\int d^3\bm{v}\, \frac{k_\perp^2 v_\perp^2}{8\Omega^2} C_l\left[(\bm{k}\cdot\bm{\rho})^2 h_{\bm{k}}^{(0)}\right] = -n_i\chi\rho_i^2 k_\perp^4\left(\frac{3}{80}\varphi_{\bm{k}} + \frac{3}{64}\frac{\delta T_{\bm{k}}}{T_i}\right).
\tag{A 49}
$$

Combining all of these, we get the following expression for the collision term on the right-hand side of (A 27):

$$
\sum_{\bm{k}} C_{D\bm{k}} e^{i\bm{k}\cdot\bm{r}} = -\sum_{\bm{k}} n_i\chi\frac{\rho_i^2}{2}k_\perp^4\left(\frac{9}{40}\varphi_{\bm{k}} - \frac{67}{160}\frac{\delta T_{\bm{k}}}{T_i}\right) e^{i\bm{k}\cdot\bm{r}} = -\frac{1}{2}n_i\chi\rho_i^2 \nabla_\perp^4\left(a\varphi - b\frac{\delta T}{T_i}\right),
\tag{A 50}
$$

where $a = 9/40$ and $b = 67/160$. The right-hand side of (A 27) is (A 50)$/\tau n_i$, hence follows the ion-density equation (2.9).

Appendix B. Conservation Laws

All calculations in this section are done to order $\mathcal{O}\left(k_\perp^4 \rho_i^4\right)$.

The nonlinearly conserved free energy in electrostatic GK is given by

$$
W = \sum_s \int d^3\bm{r} \int d^3\bm{v}\, \frac{T_s \delta f_s^2}{2F_s},
\tag{B 1}
$$

where s labels the particle species and F_s is the corresponding Maxwellian equilibrium distribution.

Using the modified adiabatic electron response (A 13), the electron contribution to (B 1) is

$$
\int d^3\bm{r} \int d^3\bm{v}\, \frac{T_e \delta f_e}{2F_e} = \frac{1}{2}T_i n_i \int d^3\bm{r}\, \tau\varphi'^2.
\tag{B 2}
$$

The ion contribution to W requires some work. Using (A 1), viz., $\delta f_i = h - \varphi F_i$, we

obtain

$$\int d^3\boldsymbol{r} \int d^3\boldsymbol{v} \, \frac{T_i \delta f_i^2}{2F_i} = \int d^3\boldsymbol{r} \int d^3\boldsymbol{v} \, \frac{1}{2}T_i \left(\frac{\langle h^2 \rangle_r}{F_i} + \varphi^2 F_i - 2\varphi \langle h \rangle_r \right). \qquad \text{(B\,3)}$$

Using the quasineutrality condition (A 16) and the gyroaverage expansion (A 6), we find

$$\int d^3\boldsymbol{r} \int d^3\boldsymbol{v} \, \frac{T_i \langle h^2 \rangle_r}{2F_i} = \frac{1}{2}T_i n_i \int d^3\boldsymbol{r} \left[\frac{3}{2}\frac{\delta T}{T_i} + \varphi^2 + 2\tau \varphi'^2 - \frac{1}{2}\varphi \rho_i^2 \nabla_\perp^2 \left(\varphi + \frac{\delta T}{T_i} \right) \right], \tag{B\,4}$$

$$\int d^3\boldsymbol{r} \int d^3\boldsymbol{v} \, \varphi \langle h \rangle_r = \varphi^2 + \tau \varphi'^2. \tag{B\,5}$$

Finally, substituting these expressions into (B 3) and combining with (B 2), we get

$$W = \frac{1}{2}T_i n_i \int d^3\boldsymbol{r} \left[\frac{3}{2}\left(\frac{\delta T}{T_i} \right)^2 + \tau \varphi'^2 - \frac{1}{2}\varphi \rho_i^2 \nabla_\perp^2 \left(\varphi + \frac{\delta T}{T_i} \right) \right]. \tag{B\,6}$$

To lowest order in $k_\perp^2 \rho_i^2$, the free energy is, therefore,

$$W = \frac{3}{2}T_i n_i \int d^3\boldsymbol{r} \, \frac{1}{2}\left(\frac{\delta T}{T_i} \right)^2, \tag{B\,7}$$

as promised in Section 2.7. Using (2.17) and (2.18), it is straightforward to show that, up to multiplicative constants related to the normalisations (2.16), (B 7) satisfies (2.34).

The 2D ($k_\parallel = 0$) GK equation for species s has an additional conserved quantity (Schekochihin *et al.* 2009), given by

$$I_s = \frac{T_s}{2F_s} \int d^3\boldsymbol{R} \, \langle \delta f_s \rangle_R^2 = \frac{T_s}{2F_s} \int d^3\boldsymbol{R} \left(h_s - \frac{Z_s e}{T_s} \langle \phi \rangle_R F_s \right)^2. \tag{B\,8}$$

In the model that we consider in this paper, only the ions are assumed to have 2D dynamics — indeed, the modified adiabatic electron response (2.4) arises as a consequence precisely of the fast parallel streaming of the electrons. Thus, for the ions,

$$\begin{aligned}
I \equiv I_i &= \frac{T_i}{2F_i} \int d^3\boldsymbol{R} \, (h - \langle \varphi \rangle_R F_i)^2 \\
&= \frac{1}{2}T_i \int d^3\boldsymbol{R} \left\{ \left(\frac{\delta T}{T_i} \right)^2 \left(\frac{v^2}{v_{ti}^2} - \frac{3}{2} \right)^2 F_i + \frac{\delta T}{T_i}\left(\frac{v^2}{v_{ti}^2} - \frac{3}{2} \right) h^{(1)} \right. \\
&\quad \left. + 2\frac{\delta T}{T_i}\left(\frac{v^2}{v_{ti}^2} - \frac{3}{2} \right)\left[\tau \varphi' - \frac{1}{4}\rho_i^2 \nabla_\perp^2 \left(\varphi + \frac{\delta T}{T_i} \right) - \frac{1}{4}\frac{v_\perp^2}{v_{ti}^2}\rho_i^2 \nabla_\perp^2 \varphi \right] F_i \right\}.
\end{aligned} \tag{B\,9}$$

Note that I is a function of velocity $\boldsymbol{v}$. In order to eliminate the unknown $h^{(1)}$, we can integrate I:

$$\int d^3\boldsymbol{v} \, I = \frac{1}{2}T_i n_i \int d^3\boldsymbol{r} \left[\frac{3}{2}\left(\frac{\delta T}{T_i} \right)^2 - \frac{1}{2}\frac{\delta T}{T_i}\rho_i^2 \nabla_\perp^2 \varphi \right]. \tag{B\,10}$$

Subtracting this from the free energy (B 6), we obtain

$$W - \int d^3\boldsymbol{v} \, I = \frac{1}{2}T_i n_i \int d^3\boldsymbol{r} \left(\tau \varphi'^2 - \frac{1}{2}\varphi \rho_i^2 \nabla_\perp^2 \varphi \right) = \frac{1}{2}T_i n_i \tau \int d^3\boldsymbol{r} \left[\varphi'^2 + \rho_s^2 (\boldsymbol{\nabla}_\perp \varphi)^2 \right], \tag{B\,11}$$

which is the conserved quantity in (2.35). This can be viewed as a version of the electrostatic GK invariant (Schekochihin *et al.* 2009)

$$Y \equiv W - \sum_s \int d^3\boldsymbol{v} \ I_s, \tag{B 12}$$

but without the electron contribution I_e, which is not conserved because the electrons do not obey $k_\parallel = 0$.

In order to obtain the third conserved quantity, we go back to (B 9) and consider

$$\int d^3\boldsymbol{v} \left(\frac{v^2}{v_{ti}^2} - \frac{3}{2}\right)^{-1} I = T_i n_i \tau \int d^3\boldsymbol{r} \left(\varphi' \frac{\delta T}{T_i} - \frac{1}{2}\frac{\delta T}{T_i}\rho_s^2 \nabla_\perp^2 \frac{\delta T}{T_i} - \frac{\delta T}{T_i}\rho_s^2\nabla_\perp^2 \varphi\right). \tag{B 13}$$

Adding this to (B 11), we obtain

$$W - \int d^3\boldsymbol{v} \left[1 - \left(\frac{v^2}{v_{ti}^2} - \frac{3}{2}\right)^{-1}\right] I = T_i n_i \tau \int d^3\boldsymbol{r} \left[\frac{1}{2}\varphi'^2 + \frac{\delta T}{T_i}\varphi' + \frac{1}{2}\rho_s^2\left(\boldsymbol{\nabla}\varphi + \boldsymbol{\nabla}\frac{\delta T}{T_i}\right)^2\right]. \tag{B 14}$$

This is the quantity that satisfies (2.36).

Appendix C. Tertiary Instability

Using the decomposition (2.59) and dropping the nonzonal-nonzonal interaction terms in (2.17) and (2.18), we obtain the linearised tertiary-mode equations:

$$(\partial_t + \underbrace{\partial_x\overline{\varphi}\partial_y}_{\textcircled{1}})(1 - \nabla^2)\varphi' - (1 \underbrace{-\partial_x^3\overline{\varphi}}_{\textcircled{2}})\partial_y(\varphi' + T') + (\kappa_T \underbrace{-\partial_x\overline{T}}_{\textcircled{3}})\partial_y\nabla^2\varphi'$$
$$+ \underbrace{(\partial_x^2\overline{\varphi})\partial_x\partial_y T' - (\partial_x^2\overline{T})\partial_x\partial_y\varphi'}_{\textcircled{4}} = -\chi\nabla^4(a\varphi' - bT'), \tag{C 1}$$

$$(\partial_t + \underbrace{\partial_x\overline{\varphi}\partial_y}_{\textcircled{5}})T' + (\kappa_T \underbrace{-\partial_x\overline{T}}_{\textcircled{6}})\partial_y\varphi' = \chi\nabla^2 T'. \tag{C 2}$$

Let us examine (C 1) and (C 2) to gain some insight into the way in which zonal fields might affect the ITG instability. Terms "1" and "5" represent the advection of density and temperature perturbations by the ZF. In the locations of nonzero ZF shear ($\partial_x^2\overline{\varphi} \neq 0$), the zonal advection is responsible for shearing the turbulent eddies and thus suppressing turbulence. We do indeed find that the growing tertiary modes are localised where the zonal shear vanishes, $\partial_x^2\overline{\varphi} = 0$. Terms "3" and "6" reflect the modification of the background temperature gradient by the zonal temperature gradient $\partial_x\overline{T}$. Their presence suggests that a possible mechanism for ITG saturation (or mitigation) is to excite a zonal temperature gradient that cancels the background gradient, thus effectively eliminating the turbulent drive (see Appendix C.4). Term "2" shows that the derivative of the zonal shear, $\partial_x^3\overline{\varphi}$, modifies the effective background magnetic field gradient. Terms "4" do not have an obvious simple interpretation that we know of.

Even though we can always embark on a 4-mode calculation similar to the one done in Section 2.8, it is, in fact, not useful in this case. In contrast to the linear regime, where there is always a well-defined fastest-growing primary mode, neither a monochromatic zonal profile nor a monochromatic tertiary mode are ever observed in simulations. Instead, we find ITG modes that are localised around the points of vanishing zonal shear.

C.1. *ITG Tertiary Instability*

Let us derive a heuristic approximation for the growth rate of the ITG instability localised around a local extremum of the ZF and of the zonal temperature gradient, i.e., around x_0 such that $\partial_x^2 \overline{\varphi}(x_0) = \partial_x^2 \overline{T}(x_0) = 0$. Note that terms "4" in (C 1) then vanish at x_0. Assuming that the tertiary modes are sufficiently localised, we can take $\partial_x^3 \overline{\varphi}$ and $\partial_x \overline{T}$ to be constant (Rogers & Dorland 2005). We can then repeat the linear calculation of Section 2.6, but now including the zonally modified gradients of the equilibrium magnetic field and background temperature, and obtain a dispersion relation for the tertiary modes. For simplicity, let us consider the collisionless ($\chi = 0$) modes. The resulting growth rate for a mode with poloidal wavenumber k_y is

$$\gamma_3 = \frac{k_y \sqrt{4(1+k^2)(1-\partial_x^3\overline{\varphi})(\kappa_T - \partial_x\overline{T}) - \left[1 - \partial_x^3\overline{\varphi} + (\kappa_T - \partial_x\overline{T})k^2\right]^2}}{2(1+k^2)}, \tag{C 3}$$

which is a modified version of (2.22).

Under the same assumptions that we used before to obtain the simplified collisionless ITG growth rate (2.23), viz., $\kappa_T - \partial_x\overline{T} \gg 1$ and $k \ll (\kappa_T - \partial_x\overline{T})^{-1/4}$, we find that the tertiary growth rate is

$$\gamma_3 \approx k_y \sqrt{(1 - \partial_x^3\overline{\varphi})(\kappa_T - \partial_x\overline{T})}. \tag{C 4}$$

If there are no background gradients, then $\gamma_3 \approx k_y \sqrt{\partial_x^3\overline{\varphi}(x_0)\partial_x\overline{T}(x_0)}$, as found by Rogers *et al.* (2000). To work out how wide this mode can be, note that we have ignored the effects of the nonzero zonal shear away from the point x_0. Taylor-expanding terms "1" and "5" in (C 1) and (C 2) around x_0, we have

$$ik_y\partial_x\overline{\varphi}(x) = ik_y\partial_x\overline{\varphi}(x_0) + \frac{1}{2}ik_y\partial_x^3\overline{\varphi}(x_0)(x - x_0)^2 + \mathcal{O}\big((x - x_0)^3\big), \tag{C 5}$$

where the first term represents a Doppler shift in the frequency and the second captures the effect of the nonzero zonal shear. Dropping the latter is, therefore, valid only in an interval around x_0 such that

$$\gamma_3 \gg \frac{1}{2}k_y\partial_x^3\overline{\varphi}(x_0)(x - x_0)^2$$

$$\implies (x - x_0)^2 \ll \Delta^2 \equiv \frac{2\gamma_3}{k_y\partial_x^3\overline{\varphi}(x_0)} \approx 2\sqrt{\frac{\left[1 - \partial_x^3\overline{\varphi}(x_0)\right]\left[\kappa_T - \partial_x\overline{T}(x_0)\right]}{\left[\partial_x^3\overline{\varphi}(x_0)\right]^2}}, \tag{C 6}$$

where we used (C 4) for the final approximation. Rogers *et al.* (2000) and Rogers & Dorland (2005) found that the scale Δ is a good approximation for the radial width of the ITG tertiary mode in the case of no equilibrium gradients.

We find that the standard simplified picture of the ITG tertiary instability outlined above does not quite describe the observed tertiary modes. Namely, (i) we find a strong temperature-gradient-driven instability at ZF minima where $\partial_x^3\overline{\varphi} > 1$ and (C 3) predicts no instability; and (ii) the instability at the ZF maxima is significantly (an order-of-magnitude) slower than predicted by (C 3). This is detailed in Appendix C.3.

C.2. *KH Tertiary Instability*

Recall that, as we showed in Section 2.5, (2.17) reduces to the Hasegawa-Mima equation in the case of $\kappa_T = 0$. This reduction is naturally achieved in the vicinity of a point x_0 where $\partial_x^2\overline{\varphi}(x_0) = \partial_x^2\overline{T}(x_0) = 0$ and $\partial_x\overline{T}(x_0) = \kappa_T$. Then the temperature equation (C 2) decouples from (C 1) due to the cancellation (or "flattening") of the equilibrium

temperature gradient by the zonal temperature. In that case, $T' = 0$ is a solution to (C 2) and (C 1) reduces to

$$(\partial_t + \overline{u}_y \partial_y)(1 - \nabla^2)\varphi' - (1 - \partial_x^2 \overline{u}_y)\partial_y \varphi' = -a\chi \nabla^4 \varphi', \tag{C 7}$$

which is the linearised Hasegawa-Mima equation for tertiary modes. This equation has a Kelvin-Helmholtz-like (KH) tertiary instability (Kim & Diamond 2002; Numata *et al.* 2007; St-Onge 2017; Zhu *et al.* 2018*a*). This KH instability is also localised around the radial locations of zero zonal shear and has a threshold roughly given by the necessary (but, in general, not sufficient) condition that $\partial_x^2 \overline{u}_y - 1$ must change sign in the region of instability, known as the Rayleigh-Kuo criterion (Kuo 1949; Zhu *et al.* 2018*a*).

C.3. *Tertiary Instabilities of the Zonal Staircase*

We now turn to a numerical investigation of the tertiary instability in the low-collisionality regime. We will consider the parameters and staircase profile of SimH ($\kappa_T = 0.36, \chi = 0.1$). The ITG turbulence trapped in the convection zones at the ZF extrema can differ substantially between ZF maxima and minima. For example, it is evident from Figures 8 and 9 that the ZF minima harbour turbulence with a larger poloidal wavenumber compared to the turbulence at the ZF maxima.

Let us describe this difference quantitatively for SimL. Consider the poloidal Fourier transforms of our fields at a fixed radial location:

$$\varphi(x, y) = \sum_{k_y} \hat{\varphi}(x, k_y)e^{ik_y y}, \qquad T(x, y) = \sum_{k_y} \hat{T}(x, k_y)e^{ik_y y}. \tag{C 8}$$

Figure 26 shows that the perturbations located at the ZF minima have a k_y spectrum peaked at a finite wavenumber, whereas the perturbations around the ZF maxima saturate at the largest available poloidal scale in the box.

An asymmetry between the ZF extrema was noticed already by McMillan *et al.* (2011), who, in their GK simulations, found that the ZF minima are less effective than the ZF maxima at stabilising turbulence. This difference has also been extensively studied using the Hasegawa-Mima equation (Zhu *et al.* 2018*b*). We find that these differences are a consequence of the mechanisms the drive the tertiary modes. By inspecting equations (C 1) and (C 2), one can instantly identify the culprit of the asymmetry: the sign of $\partial_x^3 \overline{\varphi}$ is different in the maxima (negative) and minima (positive) of the ZF. According to our analysis of the tertiary instabilities in Appendices C.1 and C.2, this has two distinct effects on the tertiary instability at the ZF extrema: one is related to the modification of the magnetic drift term and the second one to the existence of the KH tertiary at the ZF minima.

In order to investigate the tertiary instabilities that operate in the saturated state that we observe, we solve (2.17) and (2.18) numerically with an imposed static ZF profile extracted from the quasi-static zonal staircase. We do not impose a static zonal temperature profile because the observed flattening of the equilibrium temperature gradient suggests that the zonal temperature is an agent of saturation, rather than an instability mechanism.

C.3.1. *Tertiary Instability at ZF Minima*

Figure 27 shows the behaviour of the tertiary instability at the ZF maxima and minima. Let us first discuss the ZF minima. There we observe a very fast ITG-like initial instability that features both φ' and T' growing at the rate $\gamma_3^{\text{fast}} \approx 0.1$. This is of the same order as the largest linear ITG growth rate $\gamma_{\text{max}} \approx 0.2$ for the parameters of this simulation. The strong radial transport caused by the ITG eddies drives a zonal temperature perturbation

that opposes the equilibrium temperature gradient, thus flattening it. This causes a "quasilinear" saturation of the ITG instability around $t = 50$ when the zonal temperature gradient reaches the level of the equilibrium gradient and the instability is largely quenched (see also Appendix C.4). Afterwards, φ' continues to grow exponentially, albeit at a much lower growth rate of $\gamma_3^{\text{quenched}} \approx 0.003$. The nonzonal temperature T' does not grow during this quenched-growth phase and remains about an order of magnitude lower than the electrostatic potential. As we discussed in Appendix C.2, the flattening of the temperature gradient ($\kappa_T \approx \partial_x \overline{T}$) and the comparatively low levels of temperature perturbations ($T' \ll \varphi'$) imply that φ' satisfies the Hasegawa-Mima tertiary equation (C 7). We find that the Rayleigh-Kuo criterion $\partial_x^3 \overline{\varphi} \gtrsim 1$ is satisfied at the ZF minima and so there is an unstable KH mode. Its growth rate ($\gamma_3^{\text{HM}} \approx 0.008$) and poloidal spectrum are found to be similar to those of the quenched tertiary following the temperature-profile flattening (see Figure 26). We have found that if we do not evolve the zonal temperature and, thus, do not allow a flattening of the temperature gradient, the initial fast growth is never quenched.

In the vicinity of the ZF minima, we find $\partial_x^3 \overline{\varphi} \gtrsim 1$ and (C 3) predicts no ITG mode there. In fact, we do find an ITG instability (the fast initial one in Figure 27) peaked exactly at x_0 where $\partial_x^2 \overline{\varphi}(x_0) = 0$, even though $\partial_x^3 \overline{\varphi}(x_0) > 1$. We have verified that this is an ITG instability and not a KH one by setting $\kappa_T = 0$ or by freezing the nonzonal temperature perturbations and solving the corresponding Hasegawa-Mima equation numerically. Either of these reduces the tertiary growth rate by an order of magnitude, down to the level of the KH instability, indicating that the fast mode is indeed an ITG instability. Even artificially scaling the ZFs in order to get $\partial_x^3 \overline{\varphi} > 2$ in the ZF minima results in a fast ITG instability there. This tertiary mode is not the one described in Appendix C.1 and familiar from the work of Rogers *et al.* (2000). A further investigation of it is left for future work. Its effect in the quasi-stationary zonal-staircase-dominated state is to push the system in the vicinity of the ZF minima towards a state in which the zonal-temperature perturbations cancel the equilibrium temperature gradient (see Appendix C.4).

C.3.2. *Tertiary Instability at ZF Maxima*

In contrast to the ZF minima, the ZF maxima are hosts to a wimpy ITG instability with a growth rate an order of magnitude smaller than that near the ZF minima: $\gamma_3^{\text{wimpy}} = 0.006$ (see Figure 27, left panel). Setting $\kappa_T = 0$ eliminates this instability, confirming that it is, indeed, an ITG instability. The ZF satisfies $\partial_x^3 \overline{\varphi} < 0$ at the ZF maxima, so there is no unstable KH mode there; this is confirmed numerically. The poloidal wavelength of the turbulence at the ZF maxima is significantly longer than that at the ZF minima. In fact, in SimL and SimH, it is determined by the poloidal box size L_y. By increasing L_y, we are able to obtain a saturated state with a well-defined wavelength at the ZF maxima that is smaller than the poloidal box size: see Figure 28.

The suppression of the ITG instability at the ZF maxima might be due to the localisation width Δ, naïvely given by (C 6). Using the observed values of $\partial_x^3 \overline{\varphi}$ and $\partial_x \overline{T}$ in (C 6), we find $\Delta \approx 1.8$ for the fastest mode of (C 3). This suggests that these ITG modes should have radial wavenumbers of the order of $k_x \sim 2\pi/\Delta \approx 3.5$, which are deep in the stable region for the parameters considered (see Figure 2, right panel). We leave the detailed investigation of these modes for future work. Their effect in the quasi-stationary zonal-staircase-dominated state is to seed turbulence in the shear zones through the emission of travelling structures (ferdinons) when the zonal shear in the shear zones has been sufficiently weakened by viscosity (see Sections 3.1.2 and 3.3).
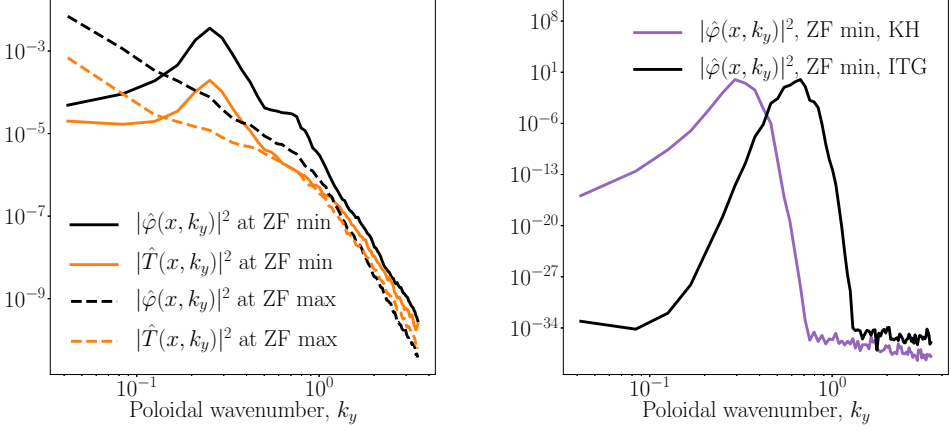
FIGURE 26. **Left:** Time-averaged poloidal spectra $\langle|\hat{\varphi}(x,k_y)|^2\rangle$ (black) and $\langle|\hat{T}(x,k_y)|^2\rangle$ (orange), as defined by (C 8), at a fixed radial location x for a ZF minimum (solid) and a ZF maximum (dashed). The time average is performed over the entire time period shown in Figure 11 (top panel). A peak in the spectrum is observed around $k_y \approx 0.25$ for the turbulence at the ZF minima. This agrees with the wavelength evident in Figure 9. In contrast, the ITG turbulence at the ZF maxima saturates at the largest available poloidal length in the box. **Right:** Poloidal spectra $\langle|\hat{\varphi}(x,k_y)|^2\rangle$ of the unstable modes at a ZF minimum for the KH instability that develops in the Hasegawa-Mima equation (purple) and the fast ITG instability without zonal temperature perturbations (black). The spectra have been normalised to the maximal value of 1. We see that both modes are approximately monochromatic, peaked around $k_y \approx 0.29$ for the KH mode and $k_y \approx 0.66$ for the ITG mode. The spectrum of saturated tubulence at the ZF minima peaks around $k_y \approx 0.25$ (left panel). This supports the case for the poloidal wavenumber at the ZF minima to be determined by the KH instability there.
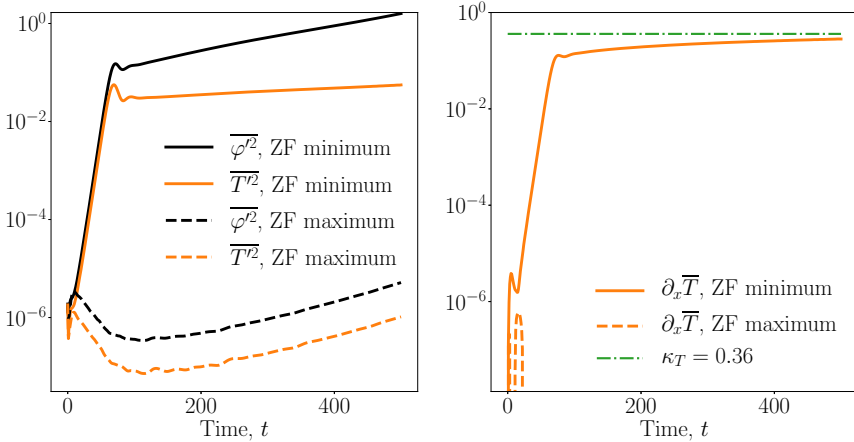

FIGURE 27. Time evolution of the perturbations in the ZF extrema of the zonal staircase of SimH. **Left:** $\overline{\varphi'^2}$ (solid black) and $\overline{T'^2}$ (solid orange) at the ZF minimum at $x \approx 24.5$. Dashed lines show the same quantities at the ZF maximum at $x \approx 47$. See Figure 8 (top panel) for the ZF profile. **Right:** Zonal temperature gradient at the ZF minimum (the ZF maximum has a very low and negative zonal temperature gradient, not shown). After the initial fast instability, the zonal temperature gradient settles at $\partial_x \overline{T} \approx \kappa_T = 0.36$.
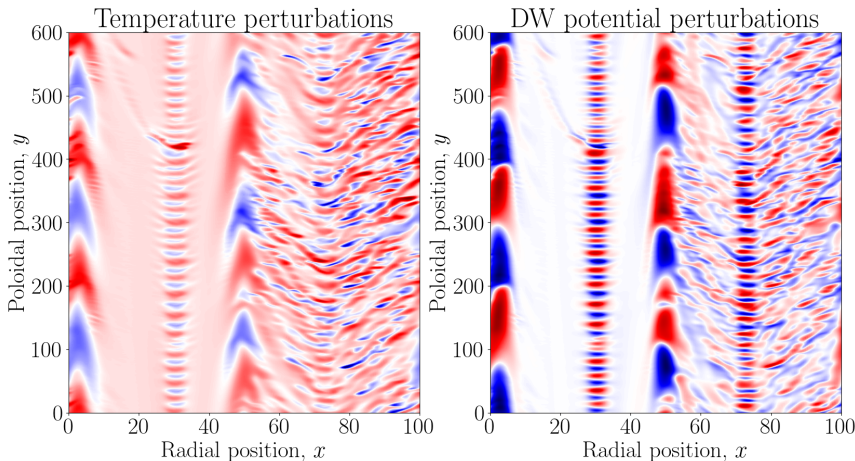
FIGURE 28. Snapshots of T and φ perturbations for SimH parameters, viz., $\kappa_T = 0.36$, $\chi = 0.1$, but for $L_y = 600$. Evident is a discrepancy in the poloidal wavenumbers of the nonzonal perturbations at the ZF maxima ($x \approx 5, 50$) and ZF minima ($x \approx 30, 75$). The dominant wavenumber at the ZF minima is $k_y \approx 0.25$, just as in the shorter box (see Figure 26), and $k_y \approx 0.03$ at the ZF maxima. The poloidal wavelength at the ZF maxima is shorter than the poloidal size of the box.

C.4. *Zonal Temperature Saturation and Equilibrium Gradient Flattening*

Is it inevitable that the temperature profile is flattened in the regions of vigorous instability? Intuitively, one might expect this to be a consequence of the high level of radial transport by the nonzonal (φ') eddies. In that case, the flattening ought to depend on some measure of the strength of the radial $\boldsymbol{E} \times \boldsymbol{B}$ velocity. Let us translate this intuition into equations.

The zonal part of (2.18) is

$$\partial_t \overline{T} = \partial_x \overline{(T\partial_y\varphi + \chi\partial_x T)} = -\partial_x(Q_t + Q_d), \tag{C 9}$$

where $Q_t = -\overline{T\partial_y\varphi}$ and $Q_d = -\chi\partial_x\overline{T}$ are the local turbulent and diffusive radial heat fluxes, respectively. This is a conservation equation for the zonal temperature perturbations. The fact that the zonal temperature stays constant $\partial_t\overline{T} = 0$ during the quiescent periods implies a balance of the local radial heat fluxes:

$$\partial_x(Q_t + Q_d) = 0 \implies Q_t = -Q_d + Q \implies \partial_x\overline{T} = \frac{Q_t - Q}{\chi}, \tag{C 10}$$

where the integration constant Q is the total box-averaged radial turbulent heat flux defined by (2.38). Indeed, this balance is observed in the quiescent zonal staircase (see Figure 29).

Consider the steady state ($\partial_t = 0$) solution of (C 2). Since we are interested in the behaviour of the temperature perturbations in the ZF extrema, where the zonal shear vanishes, we drop the effect of zonal shear (formally speaking, we transform to a frame moving with the local ZF velocity). We find

$$0 = \partial_t T' = \chi\nabla^2 T' - (\kappa_T - \partial_x\overline{T})\partial_y\varphi'. \tag{C 11}$$

Substituting (C 10) into (C 11) gives us

$$(\chi\kappa_T - Q_t + Q)\partial_y\varphi' = \chi^2\nabla^2 T'. \tag{C 12}$$
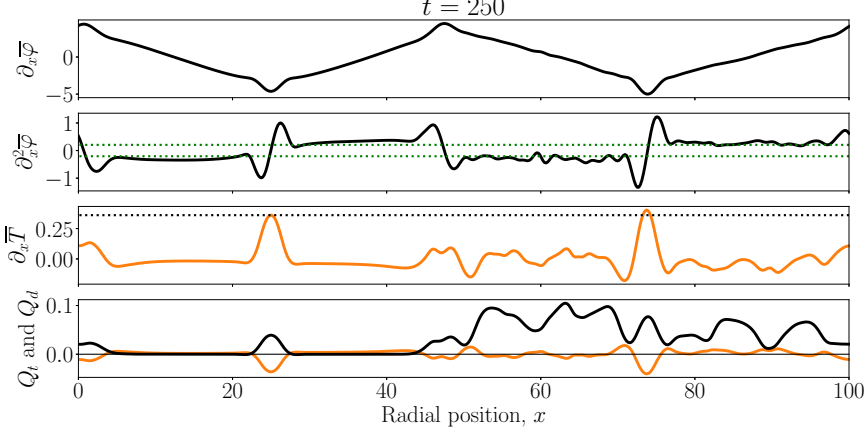
FIGURE 29. Same as the bottom panel of Figure 8, but with plots of the turbulent $Q_t = -\overline{T\partial_y\varphi}$ (black) and diffusive $Q_d = -\chi\partial_x\overline{T}$ (orange) heat fluxes as well. We can see that the balance (C 10) between Q_t and Q_d holds in the saturated staircase (left half of the domain), but does not in the turbulent regions (right half of the domain), where the turbulence generates a significant turbulent heat flux.

Multiplying both sides of (C 12) by the radial $\boldsymbol{E} \times \boldsymbol{B}$ velocity $u_x = -\partial_y\varphi'$ and performing a poloidal average yields

$$(\chi\kappa_T - Q_t + Q)\overline{u_x^2} = \chi^2 \overline{\left(\frac{\nabla^2 T'}{T'}\right)T'\partial_y\varphi'} \approx \chi^2 Q_t k_\perp^2, \qquad (C\,13)$$

where we have approximated the thermal diffusion as that of a locally monochromatic temperature perturbation T' with a wavenumber $k_\perp$. Then

$$Q_t = \frac{Q + \chi\kappa_T}{1 + \chi^2 k_\perp^2/\overline{u_x^2}}. \qquad (C\,14)$$

Substituting (C 14) into the rightmost expression in (C 10) yields the following prediction for the zonal temperature gradient

$$\partial_x\overline{T} = \frac{\kappa_T - Q\chi k_\perp^2/\overline{u_x^2}}{1 + \chi^2 k_\perp^2/\overline{u_x^2}}. \qquad (C\,15)$$

Therefore, for $Q > 0$, steady-state saturation implies $\partial_x\overline{T} < \kappa_T$. In the case of strong radial transport, $\overline{u_x^2} \gg \chi^2 k_\perp^2$, we obtain

$$Q_t \approx Q + \chi\kappa_T, \quad \partial_x\overline{T} \approx \kappa_T. \qquad (C\,16)$$

Thus, perturbations with a strong radial $\boldsymbol{E} \times \boldsymbol{B}$ velocity flatten out the temperature gradient by generating a strong zonal temperature perturbation, as intuitively expected. Indeed, (C 14) agrees with the numerical data for the heat flux at the ZF minima (see Figure 30), where we find the fast ITG tertiary instability that saturated via flattening of the equilibrium temperature gradient, as discussed in Appendix C.3.1.
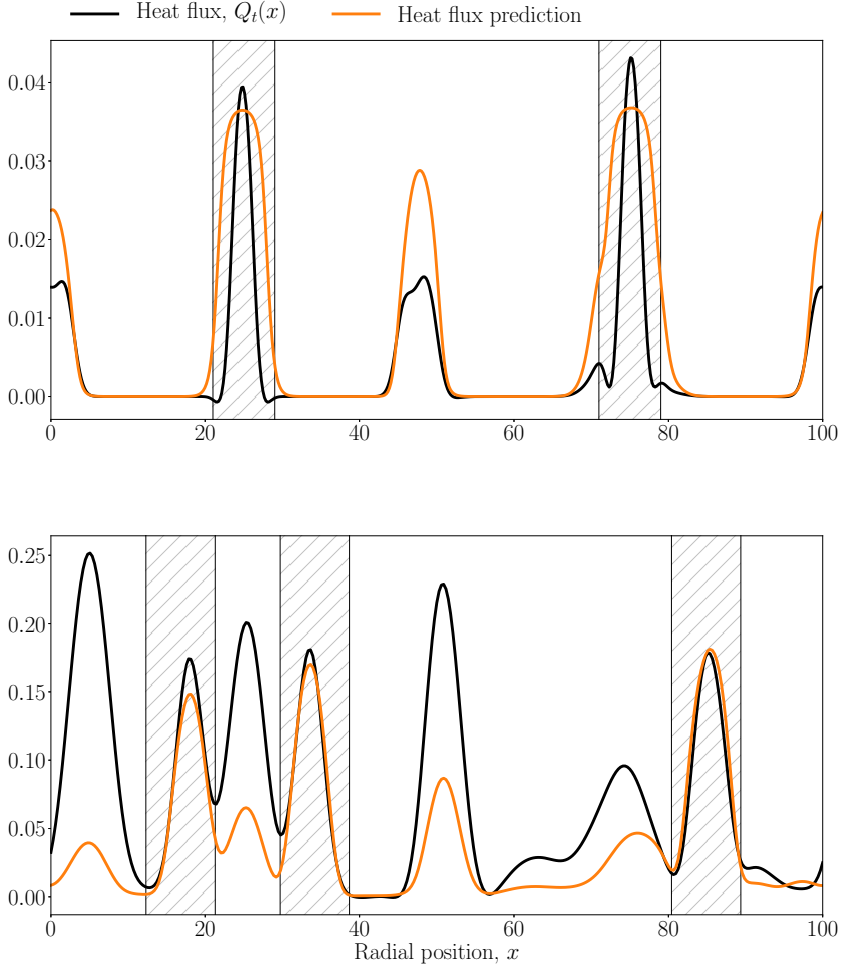
FIGURE 30. Comparison of the estimate (C 14) (orange) and observed turbulent heat flux (black) of the saturated zonal staircase. To choose $k_\perp$, we estimated k_x and k_y using the observed radial width δ of the convection zones and the poloidal spectrum of the turbulence there. The shaded areas highlight the ZF minima and have a width δ. **Top:** Data from SimH with $\delta = 8$ (so $k_x = 2\pi/8 \approx 0.8$) and $k_y = 0.25$, corresponding to the spectral peak in Figure 26. This gives $k_\perp \approx 0.8$. **Bottom:** Same as SimH, but with increased collisionality $\chi = 1$. We estimated $\delta = 9$ (so $k_x \approx 0.7$) and $k_y = 0.25$, corresponding to the peak of the poloidal spectrum at the ZF minima in that simulation. This gives $k_\perp \approx 0.7$. The agreement is better for the higher value of collisionality.

REFERENCES

ABEL, IG & COWLEY, SC 2013 Multiscale gyrokinetics for rotating tokamak plasmas: II. Reduced models for electron dynamics. *New J. Phys* **15**, 023041.

ABEL, IG, PLUNK, GG, WANG, E, BARNES, M, COWLEY, SC, DORLAND, W & SCHEKOCHIHIN, AA 2013 Multiscale gyrokinetics for rotating tokamak plasmas: fluctuations, transport and energy flows. *Rep. Prog. Phys* **76**, 116201.

BARNES, M, PARRA, FI & SCHEKOCHIHIN, AA 2011 Critically balanced ion temperature gradient turbulence in fusion plasmas. *Phys. Rev. Lett.* **107**, 115003.

BEER, MA 1995 Gyrofluid Models of Turbulent Transport in Tokamaks. PhD thesis, Princeton University.

CATTO, PJ 2019 Practical gyrokinetics. *J. Plasma Phys.* **85**, 925850301.

COLYER, GJ, SCHEKOCHIHIN, AA, PARRA, FI, ROACH, CM, BARNES, MA, GHIM, Y-C & DORLAND, W 2017 Collisionality scaling of the electron heat flux in ETG turbulence. *Plasma Phys. Control. Fusion* **59**, 055002.

COWLEY, SC, KULSRUD, RM & SUDAN, R 1991 Considerations of ion-temperature-gradient-driven turbulence. *Phys. Fluids B* **3**, 2767.

DIAMOND, PH, ITOH, S-I, ITOH, K & HAHM, TS 2005 Zonal flows in plasma—a review. *Plasma Phys. Control. Fusion* **47**, R35.

DIF-PRADALIER, G, DIAMOND, PH, GRANDGIRARD, V, SARAZIN, Y, ABITEBOUL, J, GARBET, X, GHENDRIH, PH, STRUGAREK, A, KU, S & CHANG, CS 2010 On the validity of the local diffusive paradigm in turbulent plasma transport. *Phys. Rev. E* **82**, 025401.

DIF-PRADALIER, G, HORNUNG, G, GARBET, X, GHENDRIH, PH, GRANDGIRARD, V, LATU, G & SARAZIN, Y 2017 The exb staircase of magnetised plasmas. *Nuclear Fusion* **57**, 066026.

DIF-PRADALIER, G, HORNUNG, G, GHENDRIH, PH, SARAZIN, Y, CLAIRET, F, VERMARE, L, DIAMOND, PH, ABITEBOUL, J, CARTIER-MICHAUD, T, EHRLACHER, C, ESTÈVE, D, GARBET, X, GRANDGIRARD, V, GÜRCAN, Ö D, HENNEQUIN, P, KOSUGA, Y, LATU, G, MAGET, P, MOREL, P, NORSCINI, C, SABOT, R & STORELLI, A 2015 Finding the elusive exb staircase in magnetized plasmas. *Phys. Rev. Lett.* **114**, 085004.

DIMITS, AM, BATEMAN, G, BEER, MA, COHEN, BI, DORLAND, W, HAMMETT, GW, KIM, C, KINSEY, JE, KOTSCHENREUTHER, M, KRITZ, AH, LAO, LL, MANDREKAS, J, NEVINS, WM, PARKER, SE, REDD, AJ, SHUMAKER, DE, SYDORA, R & WEILAND, J 2000 Comparisons and physics basis of tokamak transport models and turbulence simulations. *Phys. Plasmas* **7**, 969.

DORLAND, W & HAMMETT, GW 1993 Gyrofluid turbulence models with kinetic effects. *Phys. Fluids B* **5**, 812.

DORLAND, W, JENKO, F, KOTSCHENREUTHER, M & ROGERS, BN 2000 Electron temperature gradient turbulence. *Phys. Rev. Lett.* **85**, 5579.

FRIEMAN, EA & CHEN, L 1982 Nonlinear gyrokinetic equations for low-frequency electromagnetic waves in general plasma equilibria. *Phys. Fluids* **25**, 502.

FRISCH, U 1995 *Turbulence: The Legacy of A. N. Kolmogorov*. CUP.

GENTLE, KW & HE, H 2008 Texas helimak. *Plasma Sci. Technol.* **10**, 284.

HAMMETT, GW, BEER, MA, DORLAND, W, COWLEY, SC & SMITH, SA 1993 Developments in the gyrofluid approach to tokamak turbulence simulations. *Plasma Phys. Control. Fusion* **35**, 973.

HASEGAWA, A & MIMA, K 1978 Pseudo-three-dimensional turbulence in magnetized nonuniform plasma. *Phys. Fluids* **21**, 87.

HASEGAWA, A & WAKATANI, M 1983 Plasma edge turbulence. *Phys. Rev. Lett.* **50**, 682.

HELANDER, P & SIGMAR, DJ 2002 *Collisional Transport in Magnetized Plasmas*. CUP.

HORTON, W, CHOI, D & TANG, WM 1981 Toroidal drift modes driven by ion pressure gradients. *Phys. Fluids* **24**, 1077.

HORTON, W & HASEGAWA, A 1994 Quasi-two-dimensional dynamics of plasmas and fluids. *Chaos* **4**, 227.

JENKO, F, DORLAND, W, KOTSCHENREUTHER, M & ROGERS, BN 2000 Electron temperature gradient driven turbulence. *Phys. Plasmas* **7**, 1904.

KIM, E & DIAMOND, PH 2002 Dynamics of zonal flow saturation in strong collisionless drift wave turbulence. *Phys. Plasmas* **9**, 4530.

Kinsey, JE, Waltz, RE & Candy, J 2005 Nonlinear gyrokinetic turbulence simulations of exb shear quenching of transport. *Phys. Plasmas* **12**, 062302.

Kobayashi, S & Rogers, BN 2012 The quench rule, dimits shift, and eigenmode localization by small-scale zonal flows. *Phys. Plasmas* **19**, 012315.

Kotschenreuther, M, Dorland, W, Beer, MA & Hammett, GW 1995*a* Quantitative predictions of tokamak energy confinement from first-principles simulations with kinetic effects. *Phys. Plasmas* **2**, 2381.

Kotschenreuther, M, Rewoldt, G & Tang, WM 1995*b* Comparison of initial value and eigenvalue codes for kinetic toroidal plasma instabilities. *Comput. Phys. Commun.* **88**, 128.

Kraichnan, RH 1967 Inertial ranges in two-dimensional turbulence. *Phys. Fluids* **10**, 1417.

Kuo, H 1949 Dynamic instability of two-dimensional nondivergent flow in a barotropic atmosphere. *J. Meteorol.* **6**, 105.

Majda, AJ, Qi, D & Cerfon, AJ 2018 A flux-balanced fluid model for collisional plasma edge turbulence: Model derivation and basic physical features. *Phys. Plasmas* **25**, 102307.

McMillan, BF, Hill, P, Bottino, A, Jolliet, S, Vernay, T & Villard, L 2011 Interaction of large scale flow structures with gyrokinetic turbulence. *Phys. Plasmas* **18**, 112503.

McMillan, BF, Jolliet, S, Tran, TM, Villard, L, Bottino, A & Angelino, P 2009 Avalanchelike bursts in global gyrokinetic simulations. *Phys. Plasmas* **16**, 022310.

McMillan, BF, Pringle, CCT & Teaca, B 2018 Simple advecting structures and the edge of chaos in subcritical tokamak plasmas. *J. Plasma Phys.* **84**, 905840611.

Newton, SL, Cowley, SC & Loureiro, NF 2010 Understanding the effect of sheared flow on microinstabilities. *Plasma Phys. Control. Fusion* **52**, 125001.

Numata, R, Ball, R & Dewar, RL 2007 Bifurcation in electrostatic resistive drift wave turbulence. *Phys. Plasmas* **14**, 102312.

Parra, FI, Barnes, M & Peeters, AG 2011 Up-down symmetry of the turbulent transport of toroidal angular momentum in tokamaks. *Phys. Plasmas* **18**, 062501.

Plunk, GG & Bañón Navarro, A 2017 Nonlinear growth of zonal flows by secondary instability in general magnetic geometry. *New J. Phys* **19**, 025009.

Plunk, GG, Cowley, SC, Schekochihin, AA & Tatsuno, T 2010 Two-dimensional gyrokinetic turbulence. *J. Fluid Mech.* **664**, 407.

Qi, D, Majda, AJ & Cerfon, AJ 2019 A flux-balanced fluid model for collisional plasma edge turbulence: Numerical simulations with different aspect ratios. *Phys. Plasmas* **26**, 082303.

Rath, F, Peeters, AG, Buchholz, R, Grosshauser, SR, Migliano, P, Weikl, A & Strintzi, D 2016 Comparison of gradient and flux driven gyro-kinetic turbulent transport. *Phys. Plasmas* **23**, 052309.

Ricci, P, Rogers, BN & Dorland, W 2006 Small-scale turbulence in a closed-field-line geometry. *Phys. Rev. Lett.* **97**, 245001.

Rogers, BN & Dorland, W 2005 Noncurvature-driven modes in a transport barrier. *Phys. Plasmas* **12**, 062511.

Rogers, BN, Dorland, W & Kotschenreuther, M 2000 Generation and stability of zonal flows in ion-temperature-gradient mode turbulence. *Phys. Rev. Lett.* **85**, 5336.

Ruiz, DE, Glinsky, ME & Dodin, IY 2019 Wave kinetic equation for inhomogeneous drift-wave turbulence beyond the quasilinear approximation. *J. Plasma Phys.* **85**, 905850101.

Ruiz, DE, Parker, JB, Shi, EL & Dodin, IY 2016 Zonal-flow dynamics from a phase-space perspective. *Phys. Plasmas* **23**, 122304.

Schekochihin, AA, Cowley, SC, Dorland, W, Hammett, GW, Howes, GG, Quataert, E & Tatsuno, T 2009 Astrophysical gyrokinetics: kinetic and fluid turbulent cascades in magnetized weakly collisional plasmas. *Astrophys. J. Suppl. S.* **182**, 310.

Srinivasan, K & Young, WR 2012 Zonostrophic instability. *J. Atmospheric Sci.* **69**, 1633.

St-Onge, DA 2017 On non-local energy transfer via zonal flow in the Dimits shift. *J. Plasma Phys.* **83**, 905830504.

Strintzi, D & Jenko, F 2007 On the relation between secondary and modulational instabilities. *Phys. Plasmas* **14**, 042305.

Sugama, H & Horton, W 1997 Transport processes and entropy production in toroidally rotating plasmas with electrostatic turbulence. *Phys. Plasmas* **4**, 405.

Sugama, H & Horton, W 1998 Nonlinear electromagnetic gyrokinetic equation for plasmas with large mean flows. *Phys. Plasmas* **5**, 2560.

Sugama, H, Okamoto, M, Horton, W & Wakatani, M 1996 Transport processes and entropy production in toroidal plasmas with gyrokinetic electromagnetic turbulence. *Phys. Plasmas* **3**, 2379.

Terry, PW & Horton, W 1983 Drift wave turbulence in a low-order k space. *Phys. Fluids* **26**, 106.

Vallis, GK 2017 *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-Scale Circulation*, 2nd edn. CUP.

Villard, L, Angelino, P, Bottino, A, Brunner, S, Jolliet, S, McMillan, BF, Tran, TM & Vernay, T 2013 Global gyrokinetic ion temperature gradient turbulence simulations of ITER. *Plasma Phys. Control. Fusion* **55**, 074017.

Villard, L, McMillan, BF, Sauter, O, Hariri, F, Dominski, J, Merlo, G, Brunner, S & Tran, TM 2014 Turbulence and zonal flow structures in the core and l-mode pedestal of tokamak plasmas. *J. Phys. Conf. Ser.* **561**, 012022.

Waltz, RE 1988 Three-dimensional global numerical simulation of ion temperature gradient mode turbulence. *Phys. Fluids* **31**, 1962.

Waltz, RE, Dewar, RL & Garbet, X 1998 Theory and simulation of rotational shear stabilization of turbulence. *Phys. Plasmas* **5**, 1784.

Waltz, RE, Kerbel, GD & Milovich, J 1994 Toroidal gyro-landau fluid model turbulence simulations in a nonlinear ballooning mode representation with radial modes. *Phys. Plasmas* **1**, 2229.

van Wyk, F, Highcock, EG, Field, AR, Roach, CM, Schekochihin, AA, Parra, FI & Dorland, W 2017 Ion-scale turbulence in MAST: anomalous transport, subcritical transitions, and comparison to BES measurements. *Plasma Phys. Control. Fusion* **59**, 114003.

van Wyk, F, Highcock, EG, Schekochihin, AA, Roach, CM, Field, AR & Dorland, W 2016 Transition to subcritical turbulence in a tokamak plasma. *J. Plasma Phys.* **82** (6), 905820609.

Zhou, Y, Zhu, H & Dodin, IY 2019 Formation of solitary zonal structures via the modulational instability of drift waves. *Plasma Phys. Control. Fusion* **61**, 075003.

Zhou, Y, Zhu, H & Dodin, IY 2020 Solitary zonal structures in subcritical drift waves: a minimum model. *Plasma Phys. Control. Fusion* **62**, 045021.

Zhu, H, Zhou, Y & Dodin, IY 2018a On the rayleigh–kuo criterion for the tertiary instability of zonal flows. *Phys. Plasmas* **25**, 082121.

Zhu, H, Zhou, Y & Dodin, IY 2018b On the structure of the drifton phase space and its relation to the rayleigh–kuo criterion of the zonal-flow stability. *Phys. Plasmas* **25**, 072121.

Zhu, H, Zhou, Y & Dodin, IY 2019 Nonlinear saturation and oscillations of collisionless zonal flows. *New J. Phys.* **21**, 063009.

Zhu, H, Zhou, Y & Dodin, IY 2020 Theory of the tertiary instability and the dimits shift from reduced drift-wave models. *Phys. Rev. Lett.* **124**, 055002.