Nonlinear saturation and oscillations of collisionless zonal flows

Hongxuan Zhu¹ ², Yao Zhou³ and I Y Dodin¹ ²

¹ Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, United States of America
² Princeton Plasma Physics Laboratory, Princeton, NJ 08543, United States of America

E-mail: hongxuan@princeton.edu

Keywords: collisionless zonal flows, modulational instability, nonlinear stage, predator–prey oscillations

Abstract

In homogeneous drift-wave turbulence, zonal flows (ZFs) can be generated via a modulational instability (MI) that either saturates monotonically or leads to oscillations of the ZF energy at the nonlinear stage. This dynamics is often attributed as the predator–prey oscillations induced by ZF collisional damping; however, similar dynamics is also observed in collisionless ZFs, in which case a different mechanism must be involved. Here, we propose a semi-analytic theory that explains the transition between the oscillations and saturation of collisionless ZFs within the quasilinear Hasegawa–Mima model. By analyzing phase-space trajectories of DW quanta (driftons) within the geometrical-optics (GO) approximation, we argue that the parameter that controls this transition is \( N \sim \gamma_{MI}/\omega_{DW} \), where \( \gamma_{MI} \) is the MI growth rate and \( \omega_{DW} \) is the linear DW frequency. We argue that at \( N \ll 1 \), ZFs oscillate due to the presence of so-called passing drifton trajectories, and we derive an approximate formula for the ZF amplitude as a function of time in this regime. We also show that at \( N \gtrsim 1 \), the passing trajectories vanish and ZFs saturate monotonically, which can be attributed to phase mixing of higher-order sidebands. A modification of \( N \) that accounts for effects beyond the GO limit is also proposed. These analytic results are tested against both quasilinear and fully-nonlinear simulations. They also explain the earlier numerical results by Connaughton \textit{et al} (2010 J. Fluid Mech. 654 207) and Gallagher \textit{et al} (2012 Phys. Plasmas 19 122115) and offer a different perspective on what the control parameter actually is that determines the transition from the oscillations to saturation of collisionless ZFs.

1. Introduction

Zonal flows (ZFs) are banded sheared flows that can spontaneously emerge from drift-wave (DW) turbulence in magnetized plasmas [1–4] and, similarly, from Rossby-wave turbulence in the atmospheres of rotating planets [5]. They are considered important as regulators of turbulent transport and thus have been studied actively by many researchers. One mechanism of the ZF generation is the secondary (or zonostrophic) instability [3, 6–14], and the modulational instability (MI) as a special case when the DW is monochromatic [15–29]. The linear stage of the MI is generally understood, but the dynamics of ZFs at the nonlinear stage is not sufficiently explored. Some progress in this area has been made by applying quasilinear (QL) models (section 2.4), such as the second-order cumulant expansion theory (CE2) [9–14], or the stochastic structural stability theory [30–32]; however, those are not particularly intuitive. A more intuitive paradigm was proposed based on a simpler QL model known as the wave-kinetic equation (WKE) [6, 7, 33–39]. The WKE treats DW turbulence as a collection of DW quanta (‘driftons’), for which the ZF velocity serves as a collective field. Then, the ZF–DW interactions can be viewed as the predator–prey dynamics, which can result in either monotonic or oscillatory energy exchange between the ZFs and DWs [40–44]. Such dynamics is indeed observed in simulations [45]. However, the existing paradigm substantially relies on collisional damping of ZFs, whereas simulations indicate that nonlinear saturation and oscillations are also possible when the ZF is collisionless [17, 18, 25–27]. This leads to the question of how this dynamics can be explained within a collisionless theory.

© 2019 The Author(s). Published by IOP Publishing Ltd on behalf of the Institute of Physics and Deutsche Physikalische Gesellschaft
Here, we propose a semi-analytic theory that explains the transition between the oscillations and saturation of collisionless ZFs within the QL Hasegawa–Mima model. By analyzing phase-space trajectories of DW quanta (driftons) within the geometrical-optics (GO) approximation, we argue that the parameter that controls this transition is $N \sim \gamma_{\text{MI}}/\omega_{\text{DW}}$, where $\gamma_{\text{MI}}$ is the MI growth rate and $\omega_{\text{DW}}$ is the linear DW frequency. We argue that at $N \ll 1$, ZFs oscillate due to the presence of so-called passing drifton trajectories, and we derive an approximate formula for the ZF amplitude as a function of time in this regime. In doing so, we also extend the applicability of the popular ‘four-mode truncation’ (4MT) model [16–29], which is commonly used for the linear stage, to nonlinear ZF–DW interactions. We also show that at $N \gtrsim 1$, the passing trajectories vanish and ZFs saturate monotonically, which can be attributed to phase mixing of higher-order sidebands when the 4MT ceases to be a reasonable approximation. The identification of this control parameter $N$ is our main result. A modification of $N$ that accounts for effects beyond the GO limit is also proposed. These analytic results are tested against both QL and fully-nonlinear (NL) simulations. They also explain the earlier numerical results by Connaughton et al [17] and Gallagher et al [18] and offer a different perspective on what the control parameter actually is that determines the transition from oscillations to saturation of collisionless ZFs. More specifically, our control parameter captures the correct dependence on the ZF wavenumber and leads to more robust quantitative predictions than the parameter proposed in [17, 18].

Our paper is organized as follows. In section 2, we introduce the basic equations, including the Hasegawa–Mima equation (HME), the 4MT, the QL approximation, the Wigner function, and the WKE that we use later on. In section 3, we discuss the nonlinear stage of the MI using both the 4MT description and the WKE, and we also propose our control parameter $N$. In section 4, we test our control parameter $N$ using both QL and NL HME simulations. In section 5, we compare our results with those in [17, 18]. In section 6 we summarize our main conclusions.

2. Basic equations

2.1. Hasegawa–Mima equation

Both electrostatic DW turbulence in plasmas and Rossby-wave turbulence in the atmospheres of rotating planets are often modeled by the HME [46]:

$$\frac{\partial w}{\partial t} + (\hat{\xi} \times \nabla \varphi) \cdot \nabla w + \beta \frac{\partial \varphi}{\partial x} = 0,$$

(2.1)

$$w = (\nabla^2 - L_D^{-2} \hat{\alpha}) \varphi. \quad (2.2)$$

(In geophysics, it is also known as the Obukhov–Charney equation [47].) Here, the geophysics coordinate convention is used to simplify comparisons with the earlier relevant studies (section 5). The HME describes wave turbulence on a two-dimensional plane $(x, y)$, where ZFs develop along the $x$-direction, and $\nabla^2 \equiv \partial_x^2 + \partial_y^2$ is the Laplacian. (We use the symbol $\hat{\alpha}$ to denote definitions.) In the plasma-physics context, the system is assumed to be immersed in a uniform magnetic field along the $z$ axis, the plasma is assumed to have a constant density gradient scale length along the $y$ axis, $\beta$ is a scalar constant proportional to this gradient, $L_D$ is the ion-sound radius, and $\varphi$ is the perturbation of the electrostatic potential. In the geophysical context, the constant $\beta$ is proportional to the latitudinal gradient of the vertical rotation frequency, $L_D$ is the deformation radius, and $\varphi$ is the stream function.

The operator $\hat{\alpha}$ is an identity operator in the original HME (oHME). The so-called modified HME (mHME), which is also known as ‘generalized’ [33, 48] or ‘extended’ [18, 19], uses $\hat{\alpha} = -\langle f \rangle$, where $f = f(x, y, t)$ is any field quantity and

$$\langle f \rangle = \frac{1}{L_x} \int_0^{L_x} f \, dx$$

denotes the zonal average of $f$, and $L_x$ is the system length in the $x$-direction. (For more details, see, for example, [29, 49].) Below, we consider both the oHME and mHME and treat them on the same footing by using general $\hat{\alpha}$. Notably, the HME used in this paper describes a conservative system, where the primary DW amplitude is prescribed rather than driven by a certain primary instability. We choose this model because it suffices to describe the specific ZF dynamics that is addressed in our paper. This is understood from the fact that although a primary instability and dissipation would contribute additional terms in our WKE model (section 2.5), they would not affect the drifton phase-space trajectories that we discuss here. For example, in a recent study on the formation of solitary zonal structures [50], we demonstrate numerically that the structure formation is not significantly affected by external forcing and dissipation. Also, the trapping of driftions near phase-space equilibria, which will be discussed in section 4.1, can help explain the localization of DW activity reported recently within the Hasegawa–Wakatani model in [51] and may explain the results of some earlier studies of gyrokinetic plasmas [3, 52].
2.2. Fourier decomposition and the 4MT

It is common to approach the dynamics of $\varphi$ in the Fourier representation, $\varphi = \sum_k \varphi_k(t) \exp(ik \cdot x)$, where $k = (k_x, k_y)$ and $x = (x, y)$. This leads to the following equation for $\varphi_k$:

$$\frac{\partial \varphi_k}{\partial t} = i\omega_k \varphi_k + \frac{1}{2} \sum_{k_i, k_f} T(k, k_i, k_f) \varphi_{k_i} \varphi_{k_f} \delta_{k_i + k_f, k}.$$  

(2.3)

Here, $\delta_{k_i, k_f}$ equals one only if $k_i = k_f + k$ and is zero otherwise,

$$\omega_k = -\frac{\beta k_x}{k_D^2}$$  

(2.4)

is the linear DW frequency, and

$$T(k, k_i, k_f) = -\frac{k_D^2 - k_D^2}{k_D^2}(k_i \times k_f) \cdot \hat{z}$$  

(2.5)

are the coefficients that govern the nonlinear mode coupling. Also,

$$k_D^2 = \|k\|^2 + L_D^{-2} \alpha_k,$$  

(2.6)

and $k_{n,D}^2$ (where $n = 1, 2$) are similarly defined as

$$k_{n,D}^2 = \|k\|^2 + L_D^{-2} \alpha_k.$$  

(2.7)

Finally, $\alpha_k$ is the Fourier representation of the operator $\tilde{\alpha}$; namely, for the oHME $\alpha_k$ is unity, and for the mHME $\alpha_k$ equals zero if $k_x = 0$ and equals unity if $k_x = 0$.

It is also common to introduce the 4MT, which is a truncation of the system (2.3) that retains only four Fourier harmonics, namely, those with wave vectors $p, q$, and $p_c = p \pm q$. (Since $\varphi$ is real, one has $\varphi(x, y) = \varphi^*(x, y)$, hence, the harmonics with wave vectors $-p, -q$, and $-p_c$ are included too.) Then, for $\Phi_k = \varphi_k \exp(-i\omega_k t)$, (2.3) gives

$$\partial_t \Phi_p = T(p, q, p) \Phi_q \Phi_p e^{i\Delta_{+}}, \quad \partial_t \Phi_q = T(q, p, q) \Phi_p \Phi_q e^{-i\Delta_{-}}, \quad \partial_t \Phi_{p_c} = T(p_c, q, p_c) \Phi_q \Phi_q e^{-i\Delta_{-}}, \quad \partial_t \Phi_{p_c} = T(p_c, p, q) \Phi_p \Phi^* q e^{i\Delta_{+}}.$$  

(2.8a)- (2.8d)

where $\Delta_{\pm} = \omega_p \pm \omega_q - \omega_{p_c}$.

2.3. Modulational instability

Suppose a perturbation on a primary wave $\Phi_p = \Phi_0$,

$$\begin{pmatrix} \Phi_p \\ \Phi_q \\ \Phi_{p_c} \\ \Phi_{p_c} \end{pmatrix} = \begin{pmatrix} \Phi_0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \epsilon \begin{pmatrix} 0 \\ A_q e^{-i\delta q} \\ A_p e^{-i\delta p} \\ A_{p_c} e^{-i\delta_{p_c}} \end{pmatrix},$$  

(2.9)

where $\epsilon$ is small. Then, the linearized (2.8a)-(2.8d) give $\Omega_{p \pm} = \Delta_{\pm} \pm \Omega_q$ together with the following dispersion relation:

$$\Omega_q + \frac{|\Phi_0|^2 T(q, -p, -q) T(p, q, p)}{\Delta_{+} + \Omega_q} = -\frac{|\Phi_0|^2 T(q, -p, q) T(p, -q, p)}{\Delta_{-} - \Omega_q} = 0.$$  

(2.10)

(The derivation can be found, for example, in [18].) This equation can have a complex solution for $\Omega_q$ with $\text{Im} \Omega_q > 0$, which signifies the presence of the MI. As shown in [17, 28], the 4MT is indeed often a good approximation at the linear stage of the MI.

In the following, we restrict our discussions to the case when $p = (p, 0)$ and $q = (0, q)$. In this case, $\omega_q = 0$ and

$$\Delta_{\pm} \equiv \Delta = \frac{\beta pq^2}{p_D^2 (p_D^2 + q^2)}, \quad p_D^2 \equiv L_D^{-2} + p^2.$$  

(2.11)

Then, one finds that $\Omega_q^2$ is real, and hence the MI has a positive growth rate $\gamma_{\text{MI}}$ if $\Omega_q^2$ is negative; namely, $\gamma_{\text{MI}} \equiv -\Omega_q^2$, which can be explicitly written as
\[ \gamma_{\text{MI}}^2 = \frac{2|\Phi_0|^2 q^2}{(1 + \alpha L_D^{-2} q^2)} \frac{\delta' - 1}{\delta + 1} - \Delta^2. \] (2.12)

Here,
\[ \delta \approx \frac{p_D^2}{q^2}, \quad \delta' \approx \frac{p_D^2 - \alpha L_D^{-2}}{q^2}, \] (2.13)
with \( \alpha = 1 \) for the oHME and \( \alpha = 0 \) for the mHME. Notably, this implies that the mHME gives much larger growth rates than the oHME does. Also, a necessary condition for the modulation to be unstable is \( \delta' > 1 \), which requires
\[ q^2 < p_D^2 - \alpha L_D^{-2}. \] (2.14)

### 2.4. Quasilinear approximation

In order to describe the nonlinear stage of the MI analytically, we proceed as follows. Let us decompose the field quantities into the zonal-averaged part and the fluctuation part, \( f = \langle f \rangle + \tilde{f} \). Then, (2.1) and (2.2) become [12]
\[
\frac{\partial \tilde{\omega}}{\partial t} + U \frac{\partial \tilde{\omega}}{\partial x} + \left( \beta - \left( \frac{\partial^2}{\partial y^2} - \alpha L_D^{-2} \right) U \right) \frac{\partial \tilde{\varphi}}{\partial x} = \tilde{f}_{\text{NL}},
\] (2.15)
\[
\left[ 1 - \alpha L_D^{-2} \frac{\partial}{\partial y} \right] \frac{\partial U}{\partial t} = - \frac{\partial}{\partial y} \langle \tilde{v}_x \tilde{y} \rangle.
\] (2.16)

Here, \( U(y, t) \equiv - \partial_x \langle \varphi \rangle \) is the ZF velocity, \( \tilde{\varphi} \equiv \tilde{\varphi} \times \nabla \varphi \) is the fluctuation velocity, \( \partial_j^{-2} \) is an operator that in the wave-vector (Fourier) representation is simply a multiplication by \( -k_j^{-2} \), and \( \tilde{f}_{\text{NL}} \equiv \tilde{\varphi} \times \nabla \tilde{\varphi} - \langle \tilde{\varphi} \times \nabla \tilde{\varphi} \rangle \) describes self-interactions of DWs, or eddy–eddy interactions. We shall simplify the problem by ignoring these interactions, i.e. by adopting \( \tilde{f}_{\text{NL}} = 0 \). This is the commonly-used QL approximation, which often yields an adequate description of ZF–DW interactions [12]. (We shall also discuss the applicability of this approximation in section 4.2.) Notably, once the QL approximation is adopted, the 4MT model becomes the exact description of the linear MI (see section 3.1).

We also assume, for simplicity, that the ZF is sinusoidal and non-propagating (assuming \( \gamma_{\text{MI}} \) is real),
\[ U = u(t) \cos qy, \] (2.17)
where for clarity we choose the origin on the \( y \) axis such that \( y = 0 \) corresponds to the maximum of \( U \).

(Propagating zonal structures are studied in detail in our [50].) The assumption of spatially-monochromatic ZF holds approximately if the ZF’s second and higher harmonics do not outpace the fundamental harmonic during the linear MI. Hence, the ansatz (2.17) implies that the value of \( q \) is close to the one that maximizes \( \gamma_{\text{MI}} \). Then, the DW field can be represented as \( \tilde{\varphi} = \text{Re} \exp[i\sum_{m=-\infty}^{\infty} \varphi_m(t) \exp(ipx + imy)] \), where \( \varphi_0(t = 0) = \Phi_0 \) is the primary-wave amplitude, and the DW equation (2.15) becomes
\[
\frac{d\varphi_m}{dt} = i\omega_m \varphi_m + \frac{ip}{2} \left( \frac{q^2 + \alpha L_D^{-2} - k_{m+1,D}^2}{k_{m,D}^2} \right) \varphi_{m+1} + \frac{ip}{2} \left( \frac{q^2 + \alpha L_D^{-2} - k_{m-1,D}^2}{k_{m,D}^2} \right) \varphi_{m-1},
\] (2.18)
where
\[ \omega_m \equiv - \frac{k_p}{k_{m,D}^2}, \quad k_{m,D}^2 \equiv p_D^2 + (mq)^2. \] (2.19)

Note that the definition of \( k_{m,D} \) is consistent with that given by (2.7).

### 2.5. Wigner function and WKE

Equation (2.18) describes the mode-coupling among different Fourier modes of DWs due to the ZF. Now, we seek to interpret this equation in terms of the drifton phase-space dynamics. Consider the zonal-averaged Wigner function of DWs
\[
W(y, k, t) = \int d^2x \ e^{-ikx} \left\{ \tilde{\omega} \left( x + \frac{y}{2}, t \right) \tilde{\omega} \left( x - \frac{y}{2}, t \right) \right\}.
\] (2.20)

It can be understood as the spectral representation of the DW two-point correlation function, such as the one used in the CE2. It can also be viewed as the quasiprobability of the drifton distribution in the \((y, k)\) space, where the prefix ‘quasi’ reflects the fact that \( W \) is not necessarily positive-definite. That said, it becomes positive-definite in the GO limit defined as (i) \( \alpha L_D^{-2} + q^2 \ll L_D^{-2} + p^2 \) and (ii) \( \partial_y W \ll q^{-1}W \), when it can be considered as the true distribution function of driftons [53]. Then, the Wigner function can be shown to satisfy the following partial differential equation [33, 37]:
\[ \partial_t W(y, k, t) = \{ \mathcal{H}, W \} + 2\Gamma W. \]  

(2.21)

Here, \( \{, \} \) is the Poisson bracket:

\[ [A, B] \equiv \frac{\partial A}{\partial x} \cdot \frac{\partial B}{\partial k} - \frac{\partial A}{\partial k} \cdot \frac{\partial B}{\partial x}, \]

(2.22)

and the functions \( \mathcal{H} \) and \( \Gamma \) can be interpreted as the drifton Hamiltonian and the drifton dissipation rate, respectively. Specifically, for the sinusoidal ZF (2.17) they are given by [33, 36]

\[ \mathcal{H}(y, k, t) = -\frac{k_x \beta}{k_0^2} + k_y \left( 1 - \frac{q^2 + \alpha L_D^{-2}}{k_0^2} \right) u(t) \cos qy, \]

(2.23)

\[ \Gamma(y, k, t) = -\frac{k_x k_y q(q^2 + \alpha L_D^{-2})}{k_0^2} u(t) \sin qy. \]

(2.24)

Equation (2.21) is the WKE that describes the phase-space dynamics of driftons (DW quanta). They obey Hamilton’s equations and move on constant-\( \mathcal{H} \) contours in the phase-space \((y, k)\) if the ZF is stationary. (The x-momentum \( k_x = p \) serves only as a parameter.) In the GO limit, \( \Gamma \) is small and unimportant for the discussions below. However, in general, keeping \( \Gamma \) is necessary to ensure that the WKE preserves the conservation of the fundamental integrals of the HME [37].

2.6. Evolution of \( W \) beyond the GO approximation

Beyond the GO approximation, the Wigner function satisfies a pseudo-differential equation known as the Wigner–Moyal equation [33]. But here, we use a somewhat different formulation. Following the procedure in [33], we consider the spectrum of \( W \),

\[ W_\lambda(k_x, t) \equiv \int W(y, k_x, t) \exp(-i\lambda y) dy, \]

(2.25)

which is a function in the double-momentum space \((\lambda, k_x)\). Then, (2.18) is transformed into (see [33] for more details)

\[
\partial_t W_\lambda(k_x, t) = i \partial_\lambda \left( \frac{1}{\kappa_\lambda^2} - 1 \right) W_{\lambda,0} - \frac{i m p}{2} \left( Q_{\lambda+2q} W_{\lambda+2q, -q} - Q_{\lambda-2q} W_{\lambda-2q, q} \right) + \frac{i m p}{2} \left( Q_{\lambda+2q} W_{\lambda+2q, -q} - Q_{\lambda-2q} W_{\lambda-2q, q} \right),
\]

(2.26)

where we introduced

\[ W_{\lambda,0} = W_\lambda(p, k_y + \frac{b}{2}, t), \quad \kappa_\lambda^2 = p_x^2 + \left( k_y + \frac{a}{2} \right)^2, \]

(2.27)

and

\[ Q_\lambda \equiv 1 - \frac{\alpha L_D^{-2} + q^2}{\kappa_\lambda^2}. \]

(2.28)

Equation (2.26) is equivalent to (2.18), but describes the dynamics in the double-momentum space. The DW momentum flux can also be expressed through \( W_\lambda \), whose gradient drives the Fourier component of the ZF, \( U_{\lambda=q} \equiv \pi u(t) \). Specifically, from (2.16), the following equation is obtained [33]:

\[
\frac{dU_q}{dt} = \frac{i}{(1 + \alpha L_D^{-2} q^{-2})} \int \frac{d^2k}{(2\pi)^2} \kappa_q^2 q^{2} W_q(k_x, t).
\]

(2.29)

3. Nonlinear stage of the modulation instability

3.1. A preliminary toy model based on the 4MT

In order to describe the nonlinear stage of the MI, let us first consider this instability within the 4MT model. We shall show that the 4MT has an exact analytic solution that not only gives the linear growth rate (2.12), but also predicts the reversal of the ZF growth. We shall also propose a toy-model modification of the 4MT that qualitatively explains the transition from ZF oscillations to saturation. A more rigorous quantitative explanation of the transition will be given in section 3.2.

For the MI, the initial condition is \( \varphi(x, t = 0) = \Phi_0 \exp(ipx) + \text{c.c.} \), which corresponds to a delta function in the double-momentum space (see (2.20))
\[ W_\lambda(k, t = 0) = a_0 \delta(\lambda) \delta(k) \{ \delta(k_x - p) + \delta(k_x + p) \}, \]  
\[ a_0 = \pm 8\pi^3 (L_D^2 + p^2)^2 |\Phi_0|^2. \]  
According to (2.26), \( W \) remains delta-shaped also at \( t > 0 \), so we search for it in the form
\[ W_\lambda(p, k_x, t) = \sum_{m,n} W_{m,n}(t) \delta(\lambda - mq) \delta(k_x - nq/2), \]
where the factor \( \delta(k_x - p) \) is omitted, and \( W(-k, t) = W(k, t) \). The 4MT model (section 2.2) corresponds to keeping the following nine peaks:
\[ \tilde{W}_{0,0} \equiv a(t), \quad \tilde{W}_{\pm 1,\pm 2} = \tilde{d}(t), \quad \tilde{W}_{1,\pm 1} \equiv b(t) \pm ic(t), \quad \tilde{W}_{-1,\pm 1} \equiv b(t) \mp ic(t). \]
Here, \( a, b, c, d, \) and \( \tilde{d} \) are real, and the initial value of \( a(t) \) is \( a_0 \). Using (2.26), one obtains
\[ \hat{a} = 2puc \frac{\delta'}{1 + \delta}, \quad \hat{b} = \Delta \hat{c}, \quad \hat{\epsilon} = -\Delta \hat{b} + \frac{pua}{2} \frac{1 - \delta'}{\delta} + pud \frac{\delta'}{1 + \delta}, \quad \hat{d} = puc \frac{1 - \delta'}{\delta}. \]  
Here, \( \Delta, \delta, \) and \( \delta' \) are defined by (2.11) and (2.13). Also, the ZF amplitude is governed by (see (2.29))
\[ u = -\frac{pq^2 c}{\epsilon p_0^2 (p_D^2 + q^2)}, \quad \epsilon \equiv 2\pi^3 (1 + \alpha L_D^2 q^2). \]  
Note that DWs with \( k_x = p \) and \( k_x = -p \) contribute equally to \( \tilde{u} \). Next, we introduce dimensionless variables
\[ \tau = t/T, \quad a = a/A, \quad \hat{b} = b/B, \quad \hat{\epsilon} = \epsilon/C, \quad u = u/V, \]
where we choose \( B = C = \beta k v \) and
\[ T = \frac{q^2 \beta (\delta + 1)}{\beta p}, \quad A = \frac{\beta^2 \epsilon}{q^2 (\delta + (\delta')^2 - 1)}, \quad V = \sqrt{\frac{A}{2e q^2 \beta^2 \delta' \epsilon}}. \]  
Then, we obtain the time-evolution of \( \tilde{u} \) as (appendix)
\[ \frac{d^2 \hat{u}}{d\tau^2} = -\frac{d\tilde{T}}{du} \tilde{u}, \quad \Theta(\tilde{u}) \equiv -\frac{g^2 \tilde{u}^2}{2} + \frac{\tilde{u}^4}{8}. \]  
Here,
\[ g^2 \equiv \frac{\pi_0}{2} - 1, \quad \pi_0 \equiv \frac{a_0}{A}, \]  
and \( a_0 \) is given by (3.2).

The effective potential \( \Theta(\tilde{u}) \) is plotted in figure 1(a). (A small nonzero initial value of \( \tilde{u} \) causes a slight alteration of \( \Theta \), but the qualitative picture remains the same.) The steady-state solution \( \tilde{u} = 0 \) is unstable if \( g^2 > 0 \), which signifies the presence of a linear instability (namely, the MI) with the growth rate
\[ \gamma_{\text{MI}} = \frac{g^2}{T^2} = \frac{1}{T^2} \left( \frac{a_0}{2A} - 1 \right), \]
which is in agreement with (2.12). Beyond the linear regime, i.e. when \( \tilde{u}^4 \) is no longer negligible, (3.8) can also be integrated exactly, yielding
\[ \tilde{u} = 2g \sech(g \tau), \quad g = \sqrt{\frac{a_0}{2} - 1}. \]
This solution corresponds to the initial condition \( \tilde{a}(\tau \to -\infty) = \tilde{a}_0 \) and \( \tilde{u}(\tau \to -\infty) = 0 \), and the origin on the time axis is chosen such that the ZF attains the maximum amplitude at \( \tau = 0 \). In our original variables, this maximum amplitude is given by
\[ u_{\text{max}} = 2Vg = 2VT \gamma_{\text{MI}}, \]
or more explicitly
\[ u_{\text{max}} = \sqrt{\frac{2g (\delta + 1)}{\beta (\delta' - 1)}} \frac{\gamma_{\text{MI}}}{p}. \]  

The 4MT dynamics is numerically illustrated in figure 1(b). Unlike the exact solution (3.11), a finite initial perturbation of \( \tilde{u} \) results in oscillations with a finite period (figure 1(b), blue curve with circles). A comparison between the 4MT solution and numerical simulations of HWE will also be presented in section 4.1, where it is shown (in figures 6 and 7) that the 4MT solution (3.11) indeed describes the maximum ZF amplitude and the reversal of the ZF growth when the ZF is oscillating. Therefore, we can conclude that the ZF oscillates when the system can be approximately described by the 4MT. Meanwhile, we expect the ZF to saturate when the 4MT approximation fails. To illustrate the transition from the oscillations to the monotonic saturation of \( \tilde{u} \), we use a
toy model that assumes an ad hoc damping of the sidebands to mimic their coupling to higher harmonics (not to be confused with collisional damping). It is shown in appendix that in the large-ν limit the evolution of \( \bar{u} \) is

\[
\frac{d\bar{u}}{d\tau} \approx \frac{\bar{a}_0 \bar{u}}{2\nu} \left( 1 - \frac{\bar{u}^2}{\bar{a}_0} \right).
\]

(3.14)

This equation also has an exact solution, namely,

\[
\bar{u} = \sqrt{\frac{\bar{a}_0}{2}} \exp \left( \frac{\bar{a}_0}{4\nu \tau} \right) \sqrt{\text{sech} \left( \frac{\bar{a}_0}{2\nu \tau} \right)},
\]

(3.15)

which describes monotonic saturation of the ZF (figure 1(b), green curve with squares). This shows that, qualitatively, the ZF saturation can be explained as a result of phase mixing that occurs due to the coupling of the primary DW to higher harmonics.

However, the ad hoc damping is introduced for illustrative purposes only. The toy model cannot determine whether or not the 4MT is a good approximation of the QL system (2.18) and (2.26) at the nonlinear stage, and hence cannot quantitatively determine the criterion for ZF oscillations and saturation. In order to make quantitative predictions, a more rigorous approach is proposed below based on exploring the drifton phase-space dynamics.

### 3.2. A more rigorous explanation based on the phase-space dynamics of driftons

Imagine the QL system (2.18) and (2.26) as an infinite chain of oscillators (DW harmonics), each being coupled to its neighbors through the ZF amplitude \( u \). At the nonlinear stage of the MI, the energy of the primary DW at the center of the chain will propagate to higher harmonics. In the following, we show that although all DW harmonics are coupled together, only few of them actually receive a substantial amount of energy when \( u \) is smaller than a critical value \( u_{c,1} \) (see (3.16) below). Then, the 4MT can be a good approximation when \( u \ll u_{c,1} \).

The above conclusion is made based on studying the WKE (2.21), which naturally accounts for the whole DW spectrum. According to the WKE, driftons obey Hamilton’s equations and move along contours of

---

Figure 1. (a) The effective potential \( \Theta(\bar{u}) \) (see (3.9)) with \( \bar{a}_0 = 4 \) (and hence \( a^2 = 1 > 0 \)). The MI corresponds to an initial condition \( \bar{u} = \bar{a}_0 \) near the origin (black cross, \( \bar{a}_0 = 10^{-3} \)). At the nonlinear stage, \( \bar{u} \) is constrained by the conservation of the ‘energy’, \( E = (\bar{a}, \bar{u})^2/2 + \Theta \) (dashed line), and hence will start to decrease when reaching \( \bar{u} \approx 2 \) (black circle). (b) Numerical solutions of the 4MT system governed by (3.4) and (3.5) with dimensionless variables given by (3.6). The coefficient \( \nu \) describes an ad hoc damping that mimics the coupling to other DW sidebands beyond the 4MT (see (A.11)). The initial conditions are \( \bar{a} = \bar{a}_0 = 4, a = 10^{-3} \), and \( \bar{b} = \bar{c} = 0 \). A transition from oscillations to saturation of \( \bar{u} \) is observed as \( \nu \) increases.
constant $\mathcal{H}$ in the $(y, k_y)$ phase space when the ZF is stationary. As a starting point, we consider a sinusoidal ZF (2.17), when $\mathcal{H}$ is given by (2.23). Then, three types of drifton phase-space trajectories can be identified: trapped, passing, and runaway [36]. These trajectories are illustrated in figure 2(d) and are labeled 'T', 'P', and 'R', respectively. As shown in figure 2(d), trapped trajectories are closed orbits around $y = \pm \pi/q$, passing trajectories are open and periodically traverse the whole domain, and runaway trajectories move to infinity along the $k_y$ axis while retaining finite $y$. It was also found in [36] that the drifton phase-space topology can vary depending on how the ZF amplitude relates to the following two critical values:

$$u_{c1,1} = \frac{\beta}{2p_D^2 - (\alpha L_D^{-2} + q^2)}, \quad u_{c1,2} = \frac{\beta}{\alpha L_D^{-2} + q^2}. \quad (3.16)$$

(The GO limit corresponds to $u_{c1,1} \ll u_{c1,2}$.) At $u < u_{c1,1}$ (regime 1, figure 2(d)), trajectories of all three types are possible; at $u_{c1,1} \leq u < u_{c1,2}$ (regime 2, figure 2(c)), passing trajectories vanish; and finally, at $u > u_{c1,2}$ (regime 3, not shown here), trapped trajectories also vanish, leaving only runaway trajectories.

To show how considering the drifton trajectories can help understand the behavior of the QL system, we numerically simulate the evolution of $W_\alpha$ from (2.26) at fixed ZF velocity, namely, $U = u \cos qy$ with constant $u$ and $q$. The results are shown in figures 2 and 3. Under the GO assumption, both the mHME and oHME can be approximately described by their corresponding WKEs (which only differ in the value of $\alpha$) and hence can be treated on the same footing. So for clarity, we consider only the mHME ($\alpha = 0$). Since $u$ is assumed stationary, (2.26) is linear in $W$, and the initial condition is chosen as (see (3.3))
\[ W_{m,n}(t = 0) = \delta_{m,0}\delta_{n,0}, \]  
(3.17)

where \( \delta_{ij} \) is the Kronecker delta. Figure 2 corresponds to \( u < u_{c,1} \) (regime 1). In this case, the distribution \( W(y, k) \) in the phase space is confined to small \( |k_y| \) regions enclosed by the contours of passing and trapped trajectories, and no driftons reside on runaway trajectories (transport along \( k_y \) is suppressed). The corresponding \( W_{m,n} \) are vanishingly small at large \( m \) and \( n \); i.e. only a finite number of harmonics are coupled. In contrast, figure 3 corresponds to \( u_{c,1} < u < u_{c,2} \) (regime 2). In this case, passing trajectories are replaced by runaway trajectories, so driftons can propagate to much larger \( |k_y| \) (transport along \( k_y \) is not suppressed). Then, the corresponding \( W_{m,n} \) has a much wider distribution; i.e. many harmonics are coupled simultaneously.

Now, let us consider the ZF evolution that would be driven by the above dynamics (which we now consider prescribed for simplicity). As seen from \((2.16)\), the ZF is driven by the gradient of the (minus) DW momentum flux

\[ R(y, t) = -\langle \tilde{v}_x \tilde{v}_y \rangle(y, t), \]  
(3.18)

which can be numerically calculated from the Wigner function \((33)\):

\[ R(y, t) = \int \frac{d^2k}{(2\pi)^3} \frac{k_x k_y}{\kappa^2_{\omega,\lambda} k^2_{\omega,\lambda}} W_k(k, t) e^{i\lambda y}. \]  
(3.19)

The values of \( R(y, t) \) are plotted in the lowest rows in figure 2 and 3. Since \( \kappa^2_{\omega,\lambda} = p_{D,\lambda}^2 + (k_y \pm \lambda/2)^2 \), runaways with large \( |k_y| \) contribute little to \( R \), so the global dynamics is largely determined by passing driftions. Particularly, consider the slope of \( R \) at the ZF peak \((y = 0) \). In regime 1, this slope oscillates (figures 2(g)–(i)) and causes oscillations of the ZF amplitude. In contrast, in regime 2, this slope quickly flattens and stays zero indefinitely (figures 3(g)–(i)). This is due to the fact that driftions largely accumulate on runaway and trapped trajectories near the ZF troughs \((y = \pm \pi/\lambda)\) and thus cannot influence the ZF peak anymore.

From the above results, we conclude that only few DW harmonics are involved at \( u < u_{c,1} \); hence, the ZF oscillates. On the other hand, the 4MT fails to be a good approximation at \( u \ll u_{c,1} \), and hence the ZF oscillates. On the other hand, the 4MT fails to be a good approximation at \( u \ll u_{c,1} \)
approximation when \( u > u_{c,1} \); then, the ZF saturation can be attributed to the presence of runaways. Hence, whether a ZF will oscillate or saturate monotonically depends on whether the ‘control parameter’ \( N \equiv u_{\text{max}} / u_{c,1} \) is smaller or larger than unity. One can also make this estimate more quantitative as follows. Since the 4MT can be considered as a reasonable model, we use (3.13) to estimate \( u_{\text{max}} \). In the GO limit, we obtain

\[
\begin{align*}
\frac{u_{\text{max}}}{u_{c,1}} & \approx \frac{\sqrt{2} \gamma_{\text{MI}}}{p}, \\
\end{align*}
\]

where we adopted \( \delta, \delta' \gg 1 \) for the GO limit. Then, \( N \) can be expressed as

\[
N \approx \frac{2\sqrt{2} \gamma_{\text{MI}}}{\omega_{\text{DW}}} \cdot \omega_{\text{DW}} \approx \frac{\beta p}{p_{D}^{2}},
\]

where \( \omega_{\text{DW}} \) can be recognized as the (absolute value of) the characteristic DW frequency. In summary, the ZF oscillates if \( \gamma_{\text{MI}} \ll \omega_{\text{DW}} \) and monotonically saturates otherwise.

### 3.3. Modifications due to full-wave effects

At large enough \( q \), the WKE (2.21) ceases to be valid and one must take into account deviations from the GO approximation. These deviations are called full-wave effects and can be understood from the coupling between harmonics in (2.26). At large \( q \), the coupling coefficient \( Q \) deviates from unity and becomes inhomogeneous (see (2.28)). Recall that

\[
\kappa_{a}^{2} = p_{D}^{2} + \left( k_{y} + \frac{a}{2} \right)^{2},
\]

hence the minimum value of \( Q_{a} \) is achieved at \( k_{y} + \frac{a}{2} = 0 \), and

\[
\min Q = 1 - \delta_{a}^{-1}, \quad \delta_{a} = \frac{p_{D}^{2}}{\alpha L_{D}^{2} + q^{2}}.
\]

(Note that \( \delta_{a} > 1 \) is a necessary condition for the MI, and hence \( \min Q > 0 \).) When \( \min Q \ll 1 \), it can be seen from (2.28) that the modes along the diagonals \( k_{y} = \pm \lambda / 2 \) are decoupled from the rest and the initial perturbation \( \bar{W}_{0,0} \) propagates mainly along the diagonals, as demonstrated in figure 4.

When full-wave effects are important, the critical ZF amplitude can deviate from \( u_{c,1} \) given by (3.16). Here, we study the critical ZF amplitude by numerically integrating (2.26) with stationary \( u \) and recording the ZF drive at \( y = 0 \), namely

\[
R(t) \equiv \partial_{y} R(y, t)|_{y=0} = -\partial_{y} \langle \tilde{R}_{y} \tilde{R}_{y} \rangle (y, t)|_{y=0}.
\]

Specifically, we calculate

\[
R \equiv \sqrt{\frac{\langle \tilde{R}^{2}(t) \rangle}{\max R^{2}(t)}},
\]

where \( \bar{\cdot} \) is the time-average. (To exclude the initial transient dynamics, only the interval \( 0.25T < t < T \) is used for the time-averaging, where \( T \) is the total integration time.) The critical amplitude is defined as the value of \( u \) beyond which \( R \) becomes negligible.

As shown in figures 5(c) and (d), the critical amplitude starts to deviate from \( u_{c,1} \) as \( q \) increases. The following empirical correction can be adopted to account for the finite \( q \)-dependence of the critical amplitude:

\[
u_{c,1} \rightarrow u_{c,1,\text{eff}} \equiv \frac{u_{c,1}}{1 - 0.5 \delta_{a}^{-1}},
\]

as seen in figures 5(c) and (d). Then, the control parameter (3.21) becomes

\[
N = \frac{2\sqrt{2} \gamma_{\text{MI}}}{\omega_{\text{DW}}} \sqrt{\frac{\delta (\delta + 1)}{\delta (\delta' - 1)} \left( 1 - \frac{1}{2 \delta_{a}} \right)},
\]

where \( \delta, \delta', \) and \( \delta_{a} \) are given by (2.13) and (2.23). The ZF oscillates at \( N \ll 1 \) and saturates at \( N \gtrsim 1 \).

### 4. Numerical simulations

#### 4.1. Parameter scan

In order to test the above theory of the ZF fate beyond the linear stage, we numerically integrated both the the QL system (see (2.15) and (2.16)) and the NL system (see (2.1) and (2.2)) for various parameters such that
The second requirement means \( \delta_0 \) shall not be too large, i.e. \( q_2 \) shall not be too small, because then ZF harmonics with wave numbers that are multiples of \( q \) have higher growth rates and outpace the fundamental harmonic at the linear stage (see more discussions in section 5.1). Overall there are five parameters that determine the system dynamics at the nonlinear stage: \( L_D, p, \beta, q, \) and \( \Phi_0 \). For the mHME \((\alpha = 0)\), we vary only \( L_D, \beta, q, \) and \( \Phi_0 \), because \( L_D \) and \( p \) mainly appear as a combination \( L_D^2 + p^2 \), and the remaining \( p \) in (2.26) only defines the time scale and can be absorbed by a variable transformation \( t \to pt \). For the oHME \((\alpha = 1)\), we vary only \( p, \beta, q, \) and \( \Phi_0 \), because it is hard to satisfy the requirement (4.1) when \( L_D \) is varied with other parameters kept fixed. For both the oHME and the mHME, the initial conditions are chosen to be

\[
0 < N \lesssim 1, \quad \delta_0 \gtrsim 1. \tag{4.1}
\]

The simulation results of the QL and NL systems are shown in figure 6 (mHME) and 7 (oHME), where we plot the ZF energy

\[
E_{ZF}(t) \doteq \frac{1}{2} \int dy [U(y, t)]^2
\]

versus time \( t \). As predicted by our theory, in QL simulations, ZFs oscillate at the nonlinear stage if \( N \ll 1 \) and largely saturate monotonically if \( N \gtrsim 1 \). For comparison, we also plot the corresponding 4MT solutions (3.11) for the smallest-\( N \) cases in each figure. The fact that the 4MT solutions agree with predictions of the QL theory confirms the 4MT applicability at the nonlinear stage of the MI in the weak-ZF limit.

Figure 4. Same as figure 2 but for \( q = 2 \) and \( u = 0.2 \). Due to the large ZF wavenumber \( q \), full-wave effects are significant. Namely, \( W(y, t) \) are localized along the diagonals \( k_y = \pm \lambda/2 \). Also, \( W(y, t) \) cannot be easily interpreted as the drifton distribution function. Instead, each drifton is smeared out in the phase space. (The associated dataset is available at http://doi.org/10.5281/zenodo.2563449.)
approximation is satisfied with earlier simulations based on the two-fluid [51] and gyrokinetic models [3, 52]. It appears then

$$E_{\text{DW}}(t) = \frac{1}{2} \int \frac{d^2 k}{(2\pi)^2} \frac{W(y, k, t)}{k^3_3}. \quad (4.4)$$

Due to the total energy conservation, this leaves more energy for the ZFs.

In figures 6 and 7, NL simulation results are also shown for the same parameters. The transition from ZF oscillations to saturation is also recovered from these simulations. However, in terms of the oscillation amplitudes and frequencies, good agreement between QL and NL simulations occurs only when the GO oscillations to saturation is also recovered from these simulations. However, in terms of the oscillation amplitude, when $R(t)$ decays to zero, becomes larger than $u_{c, 1}$ and is estimated from (3.26). Figure 5. (a) and (b): numerical solutions of (2.26) at fixed ZF amplitude $u$ for $\alpha = 0$ (mHME) and $\beta = L = p = 1$. (a) The ZF drive $R(t)$ (see (3.24)) at $y = 0$ for various values of $u$ versus $t$ at $q = 0.1$, which corresponds to $u_{c, 1} \approx 0.25$ (see (3.16)). $R(t)$ oscillates when $u < u_{c, 1}$ and decays to zero when $u > u_{c, 1}$. (b) Same as (a) but for $q = 1.0$, which corresponds to $u_{c, 1} \approx 0.33$. The critical amplitude, when $R(t)$ decays to zero, becomes larger than $u_{c, 1}$ and is estimated from (3.26). (c) and (d): a parameter scan over $q$ and $u$ for determining the critical ZF amplitude above which $R(t)$ decays to zero. The parameters are $\beta = L = 1, \alpha = 0$ (mHME) and $p = 1$ in (c), and $\alpha = 1$ (oHME) and $p = 2$ in (d). Shown in color is the value of $R(t)$ (see (3.25)). The white dashed and solid curves are $u = u_{c, 1}$ and $u = u_{c, \text{eff}}$. The latter roughly matches the threshold beyond which $R$ is negligible. (The associated dataset is available at http://doi.org/10.5281/zenodo.2563449, [54])

At smaller $\delta_0$, NL simulations show much smaller amplitudes of the ZF oscillations due to the additional DW–DW self-interactions. A brief explanation of the discrepancy between QL and NL simulations is given in section 4.2.

In figure 8, we show a snapshot of the DW vorticity $\bar{\omega}(x, y)$, the corresponding Wigner function $W(y, k)$, and the ZF profile $U(y)$ from a QL mHME simulation. The snapshot is taken at $\gamma_{M} t = 7$ of the $\beta = 2.0$ case in figure 6(b). This corresponds to the time when the ZF energy reaches the maximum and is about to start decreasing. The DW vortex structure at $U < 0$ is clearly seen and corresponds to trapped driftons in the phase space. In contrast, there is almost no DW activity at $U > 0$, because driftons follow passing trajectories and have left this region. At this stage, the ZF velocity is no longer sinusoidal, and have a deep trough and a flat peak. This shape of the ZF is due to the larger ZF drive (see (3.19)) induced by the trapped trajectories at the ZF trough, and hence cause the ZF to have a larger local amplitude. After this moment of time, passing driftons return to the ZF top since the system is periodic in $y$, and reduce the ZF amplitude; correspondingly, the ZF energy oscillates, as shown in figure 6(b). We also note that the localization of DW structures near phase-space equilibria is largely consistent with earlier simulations based on the two-fluid [51] and gyrokinetic models [3, 52]. It appears then
that the concept of phase-space trajectories, which we have inferred from the HME, remains relevant also in these more complete descriptions of DW turbulence.

Finally, in order to further illustrate the reversal of the ZF growth caused by passing driftons and the ZF saturation caused by runaways, we provide two videos of the corresponding dynamics. The videos are available in the HTML, and two snapshots from them are shown in figure 9. The two cases shown correspond to the $\Phi_0 = 0.18$ case and the $\Phi_0 = 0.35$ case in figure 6. The videos and figures show the Wigner function $W(y,ky,t)$ and ZF velocity $U(y,t)$. They also illustrate the dynamics of test driftons, which are added post hoc. Their trajectories are calculated using the ray equations (55), where the drifton Hamiltonian $\mathcal{H}$ is determined by the ZF velocity $U(y,t)$. They also illustrate the dynamics of test driftons, which are added post hoc. Their trajectories are calculated using the ray equations [55], where the drifton Hamiltonian $\mathcal{H}$ is determined by the ZF velocity $U(y,t)$.

4.2. Difference between the QL and NL models

Here, we discuss why good agreement between QL and NL systems can be achieved when the GO approximation is well satisfied, and why discrepancies arise otherwise. Recall that the MI growth rate $\gamma_{MI}$ (2.12) derived from the 4MT is exact for the QL model but not for the corresponding NL model. Therefore, it is expected that the discrepancy can be attributed, at least partly, to the difference between the QL and NL growth rates. We compare these growth rates by comparing the relative amplitude between the second DW sideband and the first DW sideband at the linear stage, the former being excluded from the 4MT. From (2.3), we have

$$
\epsilon_{NL} \approx \frac{T(2p + q, p + q, p, q)\Phi_0}{T(p + q, q, p)\Phi_0}
$$

Figure 6. Numerical simulations of the QL (solid lines) and the NL mHWE (dotted lines, using the same parameters). The blue circles are the corresponding 4MT solutions (3.11). Shown is the ZF energy $E_{ZF}$ (see (4.3)) in units $E_{ZF}(t = 0)$ versus time in units $\gamma_{MI}t$ (see (2.12)). The specific parameters are presented in the corresponding figures. The initial conditions are given by (4.2). The best agreement between the QL and NL simulations can be found in the $\beta$-scan (figure (b)), when the GO approximation is satisfied with the highest accuracy ($\delta_\alpha = 5$). (The associated dataset is available at http://doi.org/10.5281/zenodo.2563449.[54])
as a measure of the importance of NL effects. Also, $|\varphi_p^0| = |\Phi_0|$ is assumed at the linear stage. From (2.5), the coupling coefficients are

$$|T(2p + q, p + q, p)| = \frac{q^2}{L_D^2 + 4p^2 + q^2} pq,$$

(4.7)

$$|T(p + q, q, p)| = \frac{(1 - \alpha)L_D^2 + p^2 - q^2}{L_D^2 + p^2 + q^2} pq.$$  

(4.8)

Then

$$\epsilon_{NL} \sim \left| \frac{\varphi_{p+q}}{\varphi_q} \right| = \frac{q^2}{L_D^2 + 4p^2 + q^2} \frac{L_D^2 + p^2 + q^2}{(1 - \alpha)L_D^2 + p^2 - q^2}. \quad (4.9)$$

In the above expression for $\epsilon_{NL}$, the first term $|\varphi_{p+q}/\varphi_q|$ can be estimated from the $4MT$ equations (2.8a)-(2.8d), which give

$$\left| \frac{\varphi_{p+q}}{\varphi_q} \right| \sim \left| \frac{T(p + q, q, p)\Phi_0\varphi_q}{T(q, -p, p + q)\Phi_0\varphi_{p+q}} \right|. \quad (4.10)$$

Since $\varphi_{p+q}/\varphi_q \sim \varphi_{p+q}/\varphi_q$ and $|T(q, -p, p + q)| = pq$, this gives

$$\left| \frac{\varphi_{p+q}}{\varphi_q} \right| \sim \sqrt{\frac{(1 - \alpha)L_D^2 + p^2 - q^2}{L_D^2 + p^2 + q^2}}, \quad (4.11)$$

Figure 7. Same as in figure 6, but for the oHME (\(\alpha = 1\)). Good agreement between the QL and NL simulations can be found in the $\beta$-scan and $\Phi_0$-scan (figures (c) and (d)), when the GO approximation is well satisfied ($\delta_\alpha = 2.5$). (The associated dataset is available at http://doi.org/10.5281/zenodo.2563449. [54])

New J. Phys. 21 (2019) 063009 H Zhu et al
and thus

\[ \epsilon_{NL} \sim \frac{q^2}{L^2 \delta^2 + 4p^2 + q^2} \left( \delta + 1 \right)^{1/2}, \]

(4.12)

where \( \delta \) and \( \delta_0 \) are given by (2.13) and (3.23), respectively. When \( \delta, \delta_0 \gg 1 \), i.e. \( q^2 \ll p^2 + (1 - \alpha) L^2 \), one obtains \( \epsilon_{NL} \ll 1 \). Then, NL effect is expected to be small, and hence QL and NL simulations produce similar
results. In contrast, when \( \delta_a \) approaches unity, \( \varepsilon_{NL} \) can become of order one. Then, the QL model ceases to be an adequate approximation to the NL model.

The above estimate also indicates that, at least for the mHME, when \( p^2 \gg L_D^{-2} \), one has

\[
\varepsilon_{NL} \approx \frac{\varepsilon_{GO}}{4} \sqrt{\frac{1 + \varepsilon_{GO}}{1 - \varepsilon_{GO}}} \tag{4.13}
\]

where \( \varepsilon_{GO} \sim \delta^{-1} \ll 1 \) under the GO approximation. Therefore, \( \varepsilon_{NL} \ll 1 \) at \( \varepsilon_{GO} \ll 1 \) and \( \varepsilon_{NL} \gg 1 \) when \( \varepsilon_{GO} \) approaches unity (and the MI vanishes when \( \varepsilon_{GO} > 1 \)). Hence, the applicability domain of the QL approximation is roughly the same as that of the GO approximation.

5. Comparison with previous studies

5.1. Control parameters: \( N \) versus \( M \)

Let us also compare our results with those in [17, 18] where related simulations were performed. Specifically, [17] reports NL oHME simulations for \( p^2 \gg q^2 \) and \( L_D^{-2} = 0 \). Also, [18] reports NL mHME simulations; however, the parameter is chosen such that \( p^2 \gg q^2 \), \( L_D^{-2} \), and the resulting dynamics is almost identical to that in the oHME. In both cases, the GO assumption is well satisfied (\( b, \beta', \delta_a \approx 100 \gg 1 \)). Hence, the difference between NL and QL simulations is expected to be small (see section 4.2), and the results of [17, 18] can be compared with ours within the scope of the QL approximation. (We have indeed been able to reproduce all the related results from NL simulations using the same parameters as in [17, 18], but we choose not to duplicate the figures here.) Due to the similar choice of the parameters in [17] and [18], we compare with [17] only.

Within the GO limit, our control parameter \( N \) (see (3.27)) is

\[
N \approx \frac{2\sqrt{2} p \gamma_{MI}}{\beta}, \quad \gamma_{MI} \approx \sqrt{2 p^2 q^2 |\Phi_0|^2 - \left( \frac{3q^3}{p^3} \right)^2}. \tag{5.1}
\]

A different parameter was proposed in [17], namely

\[
M \approx \frac{p^2 \Phi_0}{\beta}. \tag{5.2}
\]

There, the authors argue that \( M \ll 1/3 \) corresponds to ZF oscillations and \( M \gg 1/3 \) corresponds to monotonic ZF saturation. However, in contrast to our quantitative derivations, only a qualitative argument is provided in [17] (also see section 5.2). As a result, the parameter \( M \) cannot describe the sensitive dependence on \( q \) or \( L_D \), as our \( N \) does in figures 6(a), (c), and 7(b). Our \( N \) is also better in terms of the predictive power. For example, for the case of \( \beta = 1.2 \) in figure 7(c), one has \( M = 0.47 > 1/3 \), so the ZF is supposed to saturate; however, the ZF actually oscillates, which can be predicted by our \( N = 0.36 < 1 \).

We note that our theory can also predict the outcomes of three numerical examples used in [17], where \( M = 0.1, 1, \) and 10, respectively. For \( M = 0.1 \) and \( M = 10 \), our control parameter takes the values \( N = 0.028 \) and \( N = 3.94 \), respectively; hence, our theory predicts that the ZF oscillates in the former case and saturates monotonically in the latter case, which is indeed observed in [17]. However, [17] reports ZF saturation at \( M = 1 \), while our \( N = 0.39 < 1 \). It may seem then that our theory predicts ZF oscillations instead of saturation, which would be incorrect. But actually, due to the small \( q \) in this case, high harmonics of the ZF have much larger linear growth rates and outpace the fundamental harmonic (figure 10). This makes our analytic model inapplicable, for it assumes a quasi-sinusoidal ZF. That said, if one calculates \( N \) by replacing \( q \) with the wave number of the dominant harmonic, then \( N \) becomes larger than unity and the ZF saturation is readily anticipated.

5.2. Relevance of the Rayleigh–Kuo threshold for the ZF saturation

The authors of [17] proposed a brief explanation of the physical meaning of their parameter \( M \), which is as follows. First, they assumed that DWs transfer an order–one fraction of their energy to ZFs, so the ZF maximum amplitude is

\[
u_{\text{max}} \sim p \Phi_0. \tag{5.3}
\]

Second, they speculated that the ZF saturates when it reaches the Rayleigh–Kuo (RK) threshold [56]

\[
\partial_y^2 U(y, t) = \beta > 0. \tag{5.4}
\]

We believe that this explanation is problematic, namely, for two reasons. First, the estimate (5.3) contradicts the 4MT estimate (3.13) that we have confirmed numerically (figures 6 and 7). A more accurate estimate within the...
GO regime would be $u_{\text{max}} \gtrsim 2\beta \phi_0$. (Note that this estimate reinstates the dependence on $q$, which is absent in $M$.) Second, the RK criterion does not describe the ZF saturation but rather determines the threshold of the instability of the Kelvin–Helmholtz type that destroys the ZF. (It is also called the ‘tertiary instability’ by some authors [3, 35, 36, 55–59], and in our earlier studies we showed that this instability does not exist in the GO limit [35, 36, 55].) Since the RK threshold corresponds to $u > u_{c,2}$ (section 3.2), and $u_{c,2} \gg u_{c,1}$ in the GO regime, we claim that ZFs saturate before the RK threshold is reached. In summary, we believe that our parameter $N$ is more substantiated than the parameter $M$ introduced in [17, 18]. We also emphasize that our theory withstands the test of numerical simulations, as shown in section 4.1.

6. Conclusions

In this paper we propose a semi-analytic theory that explains the transition between the oscillations and saturation of collisionless ZFs within the QL HME. By analyzing phase-space trajectories of driftions within the GO approximation, we argue that the parameter that controls this transition is $N \sim \gamma_{\text{MI}}/\omega_{\text{DW}}$, where $\gamma_{\text{MI}}$ is the MI growth rate and $\omega_{\text{DW}}$ is the linear DW frequency. We argue that at $N \ll 1$, ZFs oscillate due to the presence of so-called passing driftion trajectories, and we derive an approximate formula for the ZF amplitude as a function of time in this regime. In doing so, we also extend the applicability of the popular 4MT model, which is commonly used for the linear stage, to nonlinear ZF–DW interactions. We also show that at $N \gtrsim 1$, the passing trajectories vanish and ZFs saturate monotonically, which can be attributed to phase mixing of higher-order sidebands when the 4MT ceases to be a reasonable approximation. A modification of $N$ that accounts for effects beyond the GO limit is also proposed. These analytic results are tested against both QL and NL simulations. They also explain the earlier numerical results by [17, 18] and offer a different perspective on what the control parameter actually is that determines the transition from oscillations to saturation of collisionless ZFs.

Acknowledgments

This work was supported by the US Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences, and also by the US DOE through Contract No. DE-AC02-09CH11466. Digital data can also be found in DataSpace of Princeton University (http://arks.princeton.edu/ark:/88435/dsp01m613n138p).
Appendix. Simplification of the 4MT equations

In the following, we simplify (3.4) and (3.5) and obtain the time-evolution of \( u \). First, notice that

\[
\frac{d}{\dot{a}} = \frac{(1 - \delta')(1 + \delta)}{266'},
\]

(A.1)

Thus, \( d \) can be expressed through \( a \):

\[
d = \frac{(1 - \delta')(1 + \delta)}{266'}(a - a_0),
\]

(A.2)

where \( a_0 \) is given by (3.2). Then, using the dimensionless variables (3.6), (3.4) and (3.5) become

\[
\frac{d\bar{a}}{d\tau} = \alpha \bar{c}, \quad \frac{d\bar{c}}{d\tau} = -\bar{c},
\]

(A.3)

\[
\frac{d\bar{b}}{d\tau} = \bar{c}, \quad \frac{d\bar{c}}{d\tau} = -\bar{b} - \left( \bar{a} - \frac{\bar{a}_0}{2} \right) \bar{u},
\]

(A.4)

where \( \bar{a}_0 = a_0 / A \). This leads to

\[
\frac{d^3\bar{u}}{d\tau^3} = \frac{d^2\bar{c}}{d\tau^2} = \frac{d\bar{b}}{d\tau} + \left( \bar{a} - \frac{\bar{a}_0}{2} \right) \frac{d\bar{u}}{d\tau} + \frac{d\bar{a}}{d\tau} u
\]

\[
= \left[ 1 - \left( \bar{a} - \frac{\bar{a}_0}{2} \right) \frac{d\bar{a}}{d\tau} - \frac{d\bar{a}}{d\tau} u \right] = -\frac{d}{d\tau} (\bar{f} \bar{u}),
\]

(A.5)

where

\[
\bar{f} = 1 - \left( a - \frac{\bar{a}_0}{2} \right).
\]

(A.6)

Assuming infinitesimally small initial \( \bar{u} \), we have

\[
\frac{d^2\bar{u}}{d\tau^2} + \bar{f} \bar{u} = 0.
\]

(A.7)

Finally, since

\[
\frac{d\bar{f}}{d\tau} = -\frac{d\bar{a}}{d\tau} = -\alpha \bar{c} = \frac{1}{2} \frac{d(\bar{u}^2)}{d\tau},
\]

(A.8)

one has that \( \bar{f} - \bar{u}^2 / 2 \) is conserved; namely

\[
\bar{f} - \frac{\bar{u}^2}{2} = -\bar{g}^2 \equiv 1 - \frac{\bar{a}_0}{2}
\]

(A.9)

is a constant. (Having \( \bar{g}^2 < 0 \) is possible but does not lead to an instability.) Hence, we obtain that \( \bar{u} \) satisfies a nonlinear-oscillator equation

\[
\frac{d^2\bar{u}}{d\tau^2} = -\frac{d\Theta}{d\bar{u}}, \quad \Theta(\bar{u}) = -\frac{\bar{g}^2 \bar{u}^2}{2} + \frac{\bar{u}^4}{8}.
\]

(A.10)

An exact solution of \( \bar{u} \) is given in section 3.1 (see (3.11)).

In section 3.1 we also propose a toy model to illustrate the transition from oscillations to saturation of \( \bar{u} \), which adds an ad hoc damping \( \nu \) of the sidebands to mimic their coupling to higher harmonics. Specifically, we replace (A.4) with

\[
\frac{d\bar{b}}{d\tau} = \bar{c} - \nu \bar{b}, \quad \frac{d\bar{c}}{d\tau} = -\bar{b} - \left( \bar{a} - \frac{\bar{a}_0}{2} \right) \bar{u} - \nu \bar{c},
\]

(A.11)

where \( \nu \) is some positive constant. Then, as we increase \( \nu \), a gradual transition from oscillations to saturation of \( \bar{u} \) is observed as shown in figure 1(b). In particular, at very large \( \nu \) such that \( \nu \gg d/d\tau \), one has

\[
\bar{b} \approx \frac{\bar{c}}{\nu}, \quad \bar{c} \approx -\frac{\bar{u}}{\nu} \left( \bar{a} - \frac{\bar{a}_0}{2} \right).
\]

(A.12)

and (A.11) gives

\[
\frac{d\bar{a}}{d\tau} \approx \frac{\bar{a}_0 \bar{u}}{2\nu} \left( 1 - \frac{\bar{u}^2}{\bar{a}_0} \right).
\]

(A.13)

An exact solution of \( \bar{u} \) in this case is also given in section 3.1 (see (3.15)).
References

[1] Lin Z, Hahm T S, Lee W W, Tang W M and White R B 1998 Science 281 1835
[2] Diamond P H, Itoh S-I, Itoh K and Hahm T S 2005 Plasma Phys. Control. Fusion 47 R35
[3] Rogers B N, Dorland W and Kotschenreuther M 2000 Phys. Rev. Lett. 85 5336
[4] Jenko F, Dorland W, Kotschenreuther M and Rogers B N 2000 Phys. Plasmas 7 1904
[5] Yasavada A R and Showman A P 2005 Rep. Prog. Phys. 68 1935
[6] Smolyakov A I, Diamond P H and Shevchenko V I 2000 Phys. Plasmas 7 1349
[7] Smolyakov A I, Diamond P H and Malkov M 2000 Phys. Rev. Lett. 84 491
[8] Strintzi D and Jenko F 2007 Phys. Plasmas 14 042305
[9] Srinivasan K and Young W R 2012 J. Atmos. Sci. 69 1633
[10] Parker J B and Krommes J A 2013 Phys. Plasmas 20 100703
[11] Parker J B and Krommes J A 2014 New J. Phys. 16 053006
[12] Parker J B 2014 PhD Thesis Princeton University
[13] St-Orge D A and Krommes J A 2017 Phys. Plasmas 24 042107
[14] Marston J B, Chini G P and Tobias S M 2016 Phys. Rev. Lett. 116 214501
[15] Qi D and Majda A arXiv:1901.08590
[16] Chen L, Lin Z and White R 2000 Phys. Plasmas 7 3129
[17] Connaughton C P, Nadiga B T, Nazarenko S V and Quinn B E 2010 J. Fluid Mech. 654 207
[18] Gallagher S, Hnat B, Connaughton C, Nazarenko S and Rowlands G 2012 Phys. Plasmas 19 122113
[19] Connaughton C, Nazarenko S and Quinn B 2015 Phys. Rep. 604 1
[20] Champeaux S and Diamond P H 2001 Phys. Lett. A 288 214
[21] Koshkarov O, Smolyakov A I and Mendonça J D 2016 Plasma Phys. Rep. 42 769
[22] Lashkin V M 2008 Phys. Plasmas 15 124502
[23] Jenko F 2006 Phys. Lett. A 351 417
[24] Shukla P K and Stenflo L 2002 Eur. Phys. J. D 20 103
[25] Manfredi G, Roach C M and Dendy R O 2001 Plasma Phys. Control. Fusion 43 825
[26] Lashmore-Davies C N, McCarthy D R and Thiyagaraja A 2001 Phys. Plasmas 8 5121
[27] Mahanti A C 1981 Arch. Meteorol. Geophys. Bioclimatol. Ser. A 30 211
[28] Gill A E 1974 Geophys. Astrophys. Fluid Dyn. 6 29
[29] St-Orge D A 2017 J. Plasma Phys. 83 905830304
[30] Constantinou N C and Parker J B 2018 Astrophys. J. 863 46
[31] Farrell B F and Ioannou P J 2007 J. Atmos. Sci. 64 3652
[32] Bakas N A, Constantinou N C and Ioannou P J 2015 J. Atmos. Sci. 72 1689
[33] Ruiz D E, Parker J B, Shi E L and Dodin I Y 2016 Phys. Plasmas 23 122304
[34] Ruiz D E, Glinsky M E and Dodin I Y 2019 J. Plasma Phys. 85 905850101
[35] Zhu H, Zhou Y, Ruiz D E and Dodin I Y 2018 Phys. Rev. E 97 053210
[36] Zhu H, Zhou Y and Dodin I Y 2018 Phys. Plasmas 25 072121
[37] Parker J B 2016 J. Plasma Phys. 82 593820602
[38] Smolyakov A I and Diamond P H 1999 Phys. Plasmas 6 4410
[39] Kaw P, Singh R and Diamond P H 2002 Plasma Phys. Control. Fusion 44 51
[40] Malkov M A, Diamond P H and Smolyakov A 2001 Phys. Plasmas 8 1553
[41] Miki K, Diamond P H, Gürcan O D, Tynan G R, Estrada T, Schmitz L and Xu G S 2012 Phys. Plasmas 19 092306
[42] Diamond P H and Liang Y M 1994 Phys. Rev. Lett. 72 2565
[43] Kim E J and Diamond P H 2003 Phys. Rev. Lett. 90 185006
[44] Berionni V and Gürcan O D 2011 Phys. Plasmas 18 112301
[45] Kobayashi S, Gürcan O D and Diamond P H 2015 Phys. Plasmas 22 090702
[46] Hasegawa A and Mima K 1977 Phys. Rev. Lett. 39 205
[47] Dolzhansky F V 2013 Fundamentals of Geophysical Hydromechanics (Berlin: Springer) ch 7
[48] Krommes J A and Kim C-B 2000 Phys. Rev. E 62 8508
[49] Hammett G W, Beer M A, Dorland W, Cowley S C and Smith S A 1993 Plasma Phys. Control. Fusion 35 973
[50] Zhou Y, Zhu H and Dodin I Y 2019 Plasma Phys. Control. Fusion 61 075003
[51] Kim B-C, Min B and An C-Y 2019 Plasma Phys. Control. Fusion 61 035002
[52] Kobayashi S and Rogers B N 2012 Plasma Phys. 54 012315
[53] Cartwright N D 1976 Physica A 83 210
[54] Zhu H, Zhou Y and Dodin I Y 2019 Nonlinear saturation and oscillations of collisionless zonal flows, Zenodo http://doi.org/10.5281/zenodo.2563449
[55] Zhu H, Zhou Y and Dodin I Y 2018 Phys. Plasmas 25 082121
[56] Kuo H-L 1949 J. Meteorol. 6 105
[57] Numata R, Ball R and Dewar R L 2007 Phys. Plasmas 14 102312
[58] Kim E J and Diamond P H 2002 Phys. Plasmas 9 4530
[59] Rath F, Peeters A G, Buchholz R, Grosshauser S R, Seiferling F and Weidk A 2018 Phys. Plasmas 25 052102