On the conditions for nonlinear growth in magnetospheric chorus and triggered emissions

Mark Gołkowski1 and Andrew R. Gibby2
1Department of Electrical Engineering, University of Colorado Denver, Denver, Colorado 80204, USA
2Arion Systems Inc., Chantilly, Virginia 20151, USA

(Received 2 June 2017; accepted 10 August 2017; published online 28 August 2017)

The nonlinear whistler mode instability associated with magnetospheric chorus and VLF triggered emissions continues to be poorly understood. Following up on formulations of other authors, an analytical exploration of the stability of the phenomenon from a new vantage point is given. This exploration derives an additional requirement on the anisotropy of the energetic electron distribution relative to the linear treatment of the instability, and shows that the nonlinear instability is most favorable to increasing growth rate when electrons become initially trapped in the wave potential of a constant frequency wave. These results imply that the initiation of the nonlinear instability at the equator requires a positive frequency sweep rate, while the initiation of the instability by a constant frequency triggering wave must occur at a location downstream of the geomagnetic equator. Published by AIP Publishing. [http://dx.doi.org/10.1063/1.4986225]

I. INTRODUCTION

The VLF phenomena of magnetospheric chorus and triggered emissions have been documented in a number of studies spanning more than 40 years.1–12 These nonlinear phenomena are both characterized by the generation of narrowband, high-intensity, changing frequency plasma emissions in the whistler mode. Additionally, the emissions are “free running,” meaning that once the emissions have developed, the frequency-time characteristics of the emitted wave energy are independent of the frequency-time behavior of any triggering source, if present.

Chorus is characterized by the spontaneous generation of these free running emissions, while triggered emissions are characterized by the presence of an externally imposed “triggering” wave. The triggering source may be natural, as in the case of whistler triggered emissions,1 or man made.9,13–15 During the triggering process, the imposed wave experiences a period of exponential temporal growth (as seen by a stationary observer) to a maximum level, at which time the free running emissions are generated.

The clear nonlinear nature (temporal growth, rapid frequency change) of these phenomena point to significant modification of the hot plasma distribution from interaction with a whistler mode wave. Early works identified the importance of radiating currents comprised of electrons in counter-streaming cyclotron resonance.2,16,17 A unique approach to the theory of chorus emissions worth mentioning was put forth by Trakhtengerts18 who proposed a backward wave oscillator (BWO) model used in device electronics and cyclotron masers. The BWO approach is applicable in the presence of beam-like or step-like hot electron distribution and has been used to interpret chorus observations.19,20 Another approach, originating from work on triggered emissions but applicable to the wider problem of cyclotron resonance growth, depends upon the formation of an electron phase space hole that is created by the potential of the narrowband triggering wave or free running emission.2,4,5,7,10,11,21–25 Specifically, near-resonant energetic electrons streaming toward the magnetic equator are trapped in the wave potential and forced to remain in cyclotron resonance with the wave over long periods. These electrons subsequently follow a phase space trajectory dictated by the resonance condition, which varies continuously due to the inhomogeneity of the static magnetic field and the time-frequency profile of the trapping wave. Furthermore, this trapping process creates a separatrix between these trapped electrons and the ambient electrons. If the electron distribution has a high energy tail with decreasing density as a function of velocity parallel to the geomagnetic field, then the trapped electrons moving toward the equator are guided to a lower parallel velocity range where the flux of the untrapped resonant electrons is higher than that of the trapped resonant electrons. Thus, this phase space region of reduced electron density represents an electron hole, generating resonant currents that act to amplify the trapping wave. Figure 1 shows a schematic illustration of the creation of the phase space hole and its effect on resonant currents.

The complicated nonlinear nature of this interaction limits a full examination of these phenomena to treatment by computer simulation. In particular, the resonant current profile for the interaction is a function of the entire several thousand kilometer trapping history and the wave profile in that region. However, it is possible to significantly simplify the analysis of the problem for a given locality in phase space by assuming that the separatrix is well defined and the electron population within is mixed sufficiently that the depressed electron density within the separatrix may be regarded as uniform as has been done by several authors.25–26 This assumption allows for the integration of the energetic electron distribution in two dimensions: velocity parallel to the ambient magnetic field and Larmor phase. Studies by Omura et al.23 and Omura et al.26 further make the problem completely tractable by assuming that the generation of the free running emissions occurs at the magnetic equator and that the electron distribution may be represented by a delta
function distribution in perpendicular velocity. Using these assumptions the authors produce a series of analytical results and they show that simulations agree well with these results, lending credence to the analytical method.

We seek to extend the method of Omura et al.\textsuperscript{23} to treat the case of the general energetic electron distribution at any given point in phase space, on or off the magnetic equator, and we find the conditions on the energetic electron distribution under which the nonlinear instability can initiate and under which it leads to accelerating wave growth. In this context, we note that Cully et al.\textsuperscript{27} also extended the methods of Omura et al.\textsuperscript{23} to a general distribution function but neglected the spatial inhomogeneity of the background field, which we include. To execute this treatment, we first present a quantitative assessment of the plasma wave instability and its evolution that differs from the typical approach for analyzing plasma wave instabilities. For the illustration of the technique, this method is first applied to the analysis of the linear whistler mode instability (in which no electron hole is formed), yielding the familiar conditions for initiation of the instability. The method is then applied to the nonlinear instability (with the presence of an electron hole) using the technique of Omura et al.\textsuperscript{23} As a result of this analysis, we find that initiation of the nonlinear instability at the equator requires a non-zero frequency sweep rate, consistent with the results of Omura et al.\textsuperscript{23} Conversely, the initiation of the nonlinear instability by a constant frequency triggering wave must occur off the equator. Furthermore, the requirement on distribution function anisotropy for initiation of the nonlinear instability off the equator is changed from the requirement specified by linear theory.

II. ANALYSIS

With wave variations that can be expressed as $e^{i(\omega t - kz)}$, the standard approach for analytically studying wave instabilities in a hot plasma is through the wave dispersion relation, which relates the frequency $\omega$ to the wave number $k$ for the wave mode under study. In the case where the solutions to the dispersion relation allow $\omega$ or $k$ to be complex quantities, the wave can be stably damped or grown without bound. For example, if the wave frequency has real and imaginary components, $\omega = \omega_0 - i \gamma$, the wave variation becomes $e^{i\gamma} e^{i(\omega_0 t - kz)}$, and the wave is damped if $\gamma < 0$, and an instability exists if $\gamma > 0$. Furthermore, the detail of the dispersion relation reveals the magnitude of the exponential growth rate $\gamma$ and the conditions under which it changes signs.

For whistler mode waves, the dispersion relation can be calculated analytically under the assumptions of linearity.\textsuperscript{8,28} Specifically, by assuming that the wave amplitude is small and that the energetic electron distribution can be functionally separated into a constant unperturbed component and a small rapidly varying component, the Vlasov equation may be linearized, and an analytical solution may be obtained. However, this technique is only applicable to the study of the initial stages of an instability when wave amplitudes are small and the linear technique cannot predict nonlinear phenomena such as saturation of the unstable growth.

In contrast, a generally applicable equation relating the growth of whistler mode waves and currents generated in the plasma can be obtained.\textsuperscript{24,16,23,26,29}

$$\frac{\partial B_w}{\partial t} + \sqrt{k(z)} v_E \frac{\partial}{\partial z} \left( \frac{B_w}{\sqrt{k(z)}} \right) = -\frac{\mu_0 v_E J_E}{2},$$

where $B_w$ is the wave magnetic field amplitude, $t$ represents time, $v_E$ is the whistler mode group velocity, $z$ is the spatial coordinate along the magnetic field line, $\mu_0$ is the magnetic permeability of free space, and $J_E$ is the component of the resonant plasma current parallel to the wave electric field. Equation (1) is derived under the assumptions of neglecting the displacement current assuming a parallel propagating
whistler mode wave. Whistler mode emissions observed on the ground typically propagated in ducts which enforce parallel propagation. The surprising utility of the parallel propagation assumption even for moderately oblique whistler mode waves is analyzed by Nunn and Omura.\textsuperscript{40} Equation (1) further assumes a single wave frequency. Specifically, the terms involving the first order time and space derivatives of \(k\) and \(j_E\) are neglected as well as the second order derivatives of \(B_w\). Omura and Matsumoto.\textsuperscript{29} It is common to assume that the inhomogeneity of the plasma is small on the scale of the wavelength allowing one to simplify Eq. (1) to

\[
\frac{\partial B_w}{\partial t} + v_g \frac{\partial B_w}{\partial z} = - \frac{\mu_0 v_g}{2} J_E.
\]

Identifying the left hand side of the equation as the substantial derivative of the wave amplitude, (2) simply states that a wave packet traveling at the group velocity will grow in accordance with the current field that is established by the wave.

By necessity, the amplitude of the resonant current \(J_E\) is a function of the wave amplitude \(B_w\), not only locally but also downstream (where downstream refers to the direction of wave propagation). In other words, the presence of the resonant current is due to the net action of the wave on the hot electron distribution. For this reason, it is important to consider the variation of \(J_E\) with respect to \(B_w\). Indeed this relationship serves as a proxy for how the initial instability in the wave will evolve. From (2), if the magnitude of \(J_E\) is an increasing function of \(B_w\), the frame of the traveling wave packet \(J_E\) is becoming more negative, then growth of the wave will create a stronger resonant current, which will, in turn, cause more wave growth. If, in contrast, the magnitude of \(J_E\) is a decreasing function of \(B_w\), then growth of the wave will act to reduce the resonant current, and the rate of growth of the wave packet will decrease. Thus, the conditions that satisfy \(\partial J_E / \partial B_w < 0\) are the conditions under which wave growth begets more wave growth, and the parameter \(\partial J_E / \partial B_w\) can serve as a proxy for conditions that lead to chorus emissions and triggered emissions. For triggered emissions, the exponential growth rate in time, as observed at a stationary receiver for a constant amplitude triggering pulse, is well established\textsuperscript{14,31} and indicates that the conditions in the magnetosphere change rapidly with time. (If the growth rate were large but constant, the constant amplitude triggering pulse would emerge amplified equally at all points in time). Growth rates for chorus emissions are difficult to quantify because the uncertainty in the seed wave amplitude and how long it grows in the linear regime. However, certain similarities between triggered emissions and chorus risers lead us to believe that conditions of accelerating growth also produce the key features of the naturally occurring emissions. Namely, for triggered emissions, the high growth phase is predictably followed by free running frequency risers,\textsuperscript{14,31,32} like that seen in chorus observations.

It is important to note that the condition \(\partial J_E / \partial B_w < 0\) does not imply that the growth of the wave will be exponential; it merely implies that a growing wave will have an increasing growth rate. We also acknowledge that in strictly mathematical terms it is possible to construct growing nonlinear solutions for \(\partial J_E / \partial B_w > 0\) and even decaying solutions for \(\partial J_E / \partial B_w < 0\). However, such scenarios are characterized by slowing growth of \(B_w\) with time and are inconsistent with the published archive, where whistler mode amplitude changes for chorus and triggered emissions are seen to be explosive and accelerating.

In general, the value of \(\partial J_E / \partial B_w\) at the initiation of an instability will determine how the instability will evolve and whether the unstable conditions will lead to observable emissions. If \(\partial J_E / \partial B_w\) is initially positive, then the presence of the instability implies that conditions are unfavorable to continued wave growth, and the wave growth will saturate. On the other hand, if \(\partial J_E / \partial B_w\) is initially negative, the wave will grow with an increasing growth rate, with growing wave amplitude continually increasing the magnitude of the amplifying resonant current.

In the case of the nonlinear interaction, the applicability of the local quantity \(\partial J_E / \partial B_w\) must be limited to the situation where the local wave field directly influences the formation of the resonant current. Specifically, at the region in phase space where a trap is forming, the local wave field contributes directly to the depth of the trap. In contrast, the value of the distribution function in a deep trap is determined by the wave profile through the entire trapping history. Therefore, the calculation of \(\partial J_E / \partial B_w\) for the nonlinear analysis that follows will be constrained to consideration of the instability as the trap is formed. If the trap is formed quickly and rapidly expands, the assumption of a mixed trap is expected to hold.

We will now calculate \(\partial J_E / \partial B_w\) for the whistler mode instability in two cases. First, we will consider the well-understood linear case, in which the wave amplitude is small enough that the linearized Vlasov equation may be used to calculate the current generating effect of the wave. Then, we will consider the nonlinear case under the assumptions of Omura \textit{et al.}\textsuperscript{28} and Omura \textit{et al.}\textsuperscript{26} By comparing \(\partial J_E / \partial B_w\) for the two cases, we will demonstrate the quantitative differences between the two types of growth and the resulting consequences for the observable phenomena of chorus and triggered emissions.

\subsection*{A. Linear analysis}

We first define a cylindrical reference frame (see Fig. 2) that is rotating at the whistler mode wave frequency \(\omega\) such that the electron velocity vector is divided into components parallel and perpendicular to the static magnetic field, and

\[\phi = \zeta - \omega t + kz,\]

where \(\zeta\) is the angle between the perpendicular velocity vector and the wave magnetic field vector, and \(\phi\) is the angle between the perpendicular velocity vector and the axis defined by the wave magnetic field at \(t = 0, z = 0\) in the stationary reference frame.\textsuperscript{8} The wave is propagating in the direction of the static magnetic field \(B_0 = B_0 e_z\), where \(e_z\) is the unit vector in that direction. The velocity vector is expressed as

\[v = v_{\perp} e_{\perp} + v_{\parallel} e_{\parallel} = -v_{\perp} \sin \zeta e_E + v_{\perp} \cos \zeta e_B - v_{\parallel} e_z,\]

where \(e_E\) and \(e_B\) are unit vectors perpendicular and parallel to the static magnetic field, respectively.
where \( v_\perp \) and \( v_\parallel \) are the electron velocity perpendicular and parallel to the static magnetic field, respectively, and \( \mathbf{e}_\perp \) and \( \mathbf{e}_\parallel \) represent unit vectors in those directions. Additionally, \( \mathbf{e}_E \) and \( \mathbf{e}_B \) are the unit vectors in the direction of the wave electric and magnetic field components, respectively, with \( v_\parallel > 0 \) implying counter-streaming particles. With these definitions, the velocity gradient operator becomes

\[
\nabla_v = \hat{\mathbf{e}}_\perp \frac{\partial}{\partial v_\perp} + \hat{\mathbf{e}}_\phi \frac{\partial}{\partial \phi} + \hat{\mathbf{e}}_z \frac{\partial}{\partial v_z},
\]

where \( \hat{\mathbf{e}}_\phi \) is the unit vector in the \( \phi \)-direction, \( v_z \) is the velocity in the direction of the static magnetic field, and use of the relations

\[
\frac{\partial \phi}{\partial t} = -\omega \quad \text{and} \quad v_z = -v_\parallel \quad (6)
\]

has been made.

Consistent with the assumption of linearity, the Vlasov equation is

\[
\frac{\partial f}{\partial t} + v \cdot \nabla f - \frac{q}{m_e} (v \times B_0) \cdot \nabla f = \frac{q}{m_e} \mathbf{E}_w \cdot \nabla f_0, \quad (7)
\]

where \( f \) is the perturbed electron distribution function, which varies as \( e^{i(\omega - k \cdot v)} \), \(-q \) is the electron charge, \( m_e \) is the electron rest mass, \( \mathbf{E}_w = \frac{q}{m_e} \mathbf{E}_E \) and \( \mathbf{B}_w = \frac{q}{m_e} \mathbf{E}_B \) are the whistler mode wave electric and magnetic field vectors, respectively, and \( f_0 \) is the initial unperturbed electron distribution function. The unit vectors in the perpendicular plane are related by

\[
\hat{\mathbf{e}}_E = -\sin \zeta \hat{\mathbf{e}}_\perp - \cos \zeta \hat{\mathbf{e}}_z, \quad (8)
\]

\[
\hat{\mathbf{e}}_B = \cos \zeta \hat{\mathbf{e}}_\perp - \sin \zeta \hat{\mathbf{e}}_z, \quad (9)
\]

With \( \partial f / \partial t = i \omega f \) and \( \nabla f = -ik \hat{\mathbf{e}}_z \), and applying (5) the left hand size of (7) becomes

\[
\frac{\partial f}{\partial t} + v \cdot \nabla f - \frac{q}{m_e} (v \times B_0) \cdot \nabla f = i \omega f + ikv_\parallel f - i \Omega f, \quad (10)
\]

where \( \Omega = \frac{qB_0}{m_e} \) is the electron cyclotron frequency. Using (8), the term in brackets becomes

\[
[B_w + (v \times B_w)] = B_w \left[ \left( \frac{\omega}{k} + v_\parallel \right) \hat{\mathbf{e}}_E - v_\perp \sin \zeta \hat{\mathbf{e}}_z \right]
\]

\[
= -B_w \sin \zeta \left[ \left( \frac{\omega}{k} + v_\parallel \right) \hat{\mathbf{e}}_E - v_\perp \hat{\mathbf{e}}_z \right]
\]

\[
= -B_w \cos \zeta \left[ \left( \frac{\omega}{k} + v_\parallel \right) \hat{\mathbf{e}}_z \right], \quad (11)
\]

and the right hand side of (7) becomes

\[
\frac{q}{m_e} \left[ E_w + (v \times B_w) \right] \cdot \nabla f_0 = -\frac{qB_w}{m_e} \sin \zeta \left[ \left( \frac{\omega}{k} + v_\parallel \right) \frac{\partial f_0}{\partial v_\perp} - v_\perp \frac{\partial f_0}{\partial v_z} \right], \quad (12)
\]

where the \( \sin \zeta \)-dependence has dropped out due to the azimuthal symmetry of the unperturbed velocity distribution. Equating (10) and (12), the perturbed distribution function can be found directly

\[
f = i \frac{qB_w}{m_e} \left[ -\frac{\sin \zeta}{\omega + k v_\parallel - \Omega} \right] \left\{ \left( \frac{\omega}{k} + v_\parallel \right) \frac{\partial f_0}{\partial v_\perp} - v_\perp \frac{\partial f_0}{\partial v_z} \right\}. \quad (13)
\]

An examination of (13) reveals many of the characteristics of the linear instability. Upon integration, the \( \sin \zeta \)-dependence will result in a resonant current parallel or anti-parallel to the wave electric field. In addition, the singularity at \( v_\parallel = v_{res} = \frac{\Omega - \omega}{k} \) (the familiar whistler mode cyclotron resonance condition) will result in current components that are in phase (real) and out of phase (imaginary) with the wave. The imaginary component is dominated by the cold electrons and represents the cold plasma currents carried by the whistler mode wave. The real component, on the other hand, is dominated by the contribution of the hot electrons near resonance, and can exchange energy with the wave. To find the resonant current

\[
J_E = \text{Re} \left[ \int_0^{2\pi} \int_{-\infty}^{\infty} \left[ q v_\perp^2 \sin \zeta d\omega d v_\perp \right] \right]
\]

\[
\times \int_{-\infty}^{\infty} \left[ \frac{v_\parallel}{v_\parallel - \frac{\Omega - \omega}{k}} \right] \left[ \frac{\omega}{k} + v_\parallel \right] \frac{\partial f_0}{\partial v_\perp} - v_\perp \frac{\partial f_0}{\partial v_z} \right] \right] \right]. \quad (14)
\]
Integrating over $\zeta$, applying the Plemelj relation, and taking the real part

$$J_E = -\frac{q^2 B_w \pi^2}{km_e} \int_0^\infty \left\{ \frac{(\omega - v_i)}{k + v_i} \right\} \left\{ \frac{\partial f_0}{\partial v_{\perp}} - v_{\perp} \frac{\partial f_0}{\partial v_{\parallel}} \right\}_{v_i = v_{\text{res}}} v^2_{\perp} dv_{\perp},$$

(15)

which explicitly relates the resonant current $J_E$ to the wave amplitude $B_w$. Finally, taking the derivative with respect to the wave amplitude,

$$\frac{\partial J_E}{\partial B_w} = -\frac{q^2 \pi^2}{km_e} \int_0^\infty \left\{ \frac{(\omega - v_i)}{k + v_i} \right\} \left\{ \frac{\partial f_0}{\partial v_{\perp}} - v_{\perp} \frac{\partial f_0}{\partial v_{\parallel}} \right\}_{v_i = v_{\text{res}}} v^2_{\perp} dv_{\perp},$$

(16)

which has considerable similarity to the linear growth rate, which is proportional to the anisotropy $A$ of the distribution given by

$$A = \frac{\int_0^\infty \left\{ \frac{\partial f_0}{\partial v_{\perp}} - v_{\perp} \frac{\partial f_0}{\partial v_{\parallel}} \right\}_{v_i = v_{\text{res}}} v^2_{\perp} dv_{\perp}}{\int_0^\infty \left\{ f \right\}_{v_i = v_{\text{res}}} v_{\perp} dv_{\perp}^\perp}.$$  

(17)

We will later compare (16) to the result for the nonlinear instability to distinguish between the phenomena.

**B. Nonlinear analysis**

We next derive an expression similar to (16) for the nonlinear case, following the assumptions of Omura et al. Specifically, the nonlinear interaction is assumed to be dominated by the effect of particle trapping in the potential of the large amplitude whistler mode wave. The Lorentz force keeps the trapped electrons in near resonance with the wave

$$v_{\parallel} \approx v_{\text{res}} = \frac{\Omega(z) - \omega}{k(z)},$$  

(18)

where the cyclotron frequency and wave number are explicit functions of position along the field line due to the inhomogeneous geomagnetic field. The trapped electrons are forced to satisfy Eq. (18), and therefore, their parallel velocity is changed from what it would be without the presence of the wave. This diversion of electrons from their adiabatic (unperturbed) paths results in a local depression of phase space density within the region of the particle trap under the conditions of a high energy tail with decreasing density as a function of parallel velocity. To represent this, the nonlinear perturbed distribution function may be expressed as the combination of the initial distribution function $f_0$ and a disturbed distribution function $f_d$, which represents the phase space depression (see Fig. 1)

$$f(v_{\perp}, v_{\parallel}, \zeta) = f_0(v_{\perp}, v_{\parallel}) - f_d(v_{\perp}, v_{\parallel}, \zeta)$$  

(19)

with

$$f_d(v_{\perp}, v_{\parallel}, \zeta) = h(v_{\perp}) g_l(v_{\parallel}, \zeta),$$  

(20)

where we have separated the distribution function into a component $h(v_{\perp})$ which is determined by the value of the distribution function within the particle trap and a component $g_l(v_{\parallel}, \zeta)$ which describes the shape of the trap. In particular if it is assumed that the trap is well mixed and uniform then we can express, $g_l(v_{\parallel}, \zeta) = 1$ within the trapping region and $g_l(v_{\parallel}, \zeta) = 0$ otherwise. Thus

$$\int_0^{2\pi} \int_{-\infty}^{\infty} g_l(v_{\parallel}, \zeta) \sin \zeta dv_{\parallel} d\zeta = 2\frac{q\omega_e}{\omega_e} \int_0^{\frac{\pi}{2}} \sqrt{2 \cos \zeta_1 - \cos \zeta + (\zeta - \zeta_1)S} \sin \zeta d\zeta,$$

(21)

where $e$ is the free space speed of light, $\omega_e = \sqrt{k \varepsilon_0 q B_w/m_e}$ is the trapping frequency and $S$ is the collective inhomogeneity factor

$$S = \frac{1}{\omega_e^2} \left\{ \left( 1 - \frac{v_{\text{res}}}{v_g} \right) \frac{2 \partial \omega}{\partial \Omega} + \left[ \frac{k \varepsilon_0}{2 \Omega} \frac{3}{2} \frac{v_{\text{res}}}{v_{\text{res}}} \right] \frac{\partial \Omega}{\partial \zeta} \right\},$$

(22)

which represents the ability of the wave to trap particles. Specifically, when $|S| < 1$ particles trajectories will oscillate in phase space in what is called a particle trap between the Larmor angles $\zeta_1$ and $\zeta_2$, which are found by solving the system of equations

$$\sin \zeta_1 = -S,$$

$$\cos \zeta_1 - \cos \zeta_2 = (\zeta_1 - \zeta_2)S.$$  

(23)

Noting that (23) has two solutions over the range $-\pi < \zeta_1 < \pi$, $\zeta_1$ is the solution with the smaller magnitude (the other solution represents the stable equilibrium point within the trap). The integral in (21) can be evaluated numerically, and we define the quantities $J_0 = (2\pi)^{1/2} B_w^{1/2} \Omega^{1/2} O_E(S)$ and

$$O_E(S) \equiv \int_{\zeta_1}^{\zeta_2} \sqrt{2 \cos \zeta_1 - \cos \zeta + (\zeta - \zeta_1)S} \sin \zeta d\zeta,$$

(24)

such that

$$\int_0^{2\pi} \int_{-\infty}^{\infty} g_l(v_{\parallel}, \zeta) \sin \zeta dv_{\parallel} d\zeta = J_0 B_w^{1/2} \Omega^{1/2} O_E(S),$$

(25)

the value of which depends on $v_{\perp}$ and $B_w$ explicitly as well as implicitly through the dependence of $S$ on these quantities. $O_E$ is plotted as a function $S$ in Fig. 3. (Note that the definition of the quantity $J_0$ here is slightly different than that used by Omura et al. in that the wave amplitude is not included).

Using the definition of the nonlinear distribution function, (19) and (20), we calculate the resonant current

$$J_E = \int_0^{2\pi} \int_{-\infty}^{\infty} q v_{\perp}^2 \sin \zeta f_d(v_{\perp}, v_{\parallel}, \zeta) dv_{\perp} dv_{\parallel} d\zeta$$

$$= -q \int_0^{\infty} h(v_{\perp}) \left\{ \int_0^{2\pi} \int_{-\infty}^{\infty} g_l(v_{\parallel}, \zeta) \sin \zeta dv_{\parallel} d\zeta \right\} dv_{\perp}$$

$$= -q \int_0^{\infty} h(v_{\perp}) v_{\perp}^2 B_w^{1/2} O_E(S) dv_{\perp},$$

(26)

(27)
Taking the derivative of (27) with respect to $B_w$, and defining the quantity

$$G_E(S) = \frac{\partial}{\partial B_w} \left( O_E(S) B_w^{1/2} \right)$$

(28)

for succinctness, yields

$$\frac{\partial J_E}{\partial B_w} = -qJ_0 \int_0^\infty \left\{ h(v_\perp) G_E(S) + B_w^{1/4} \frac{\partial h(v_\perp)}{\partial B_w} O_E(S) \right\} v_\perp^2 dv_\perp$$

(29)

from which we note that the sign is determined by $h(v_\perp)$ and its derivative and $G_E(S)$, since $O_E(S) \geq 0$.

Noting that $S$ is inversely proportional to $B_w$, and neglecting the effect of the wave amplitude on the frequency, (28) can be evaluated and expressed as

$$G_E(S) = \frac{1}{\sqrt{B_w}} \left( \frac{O_E(S)}{2} - S \frac{\partial O_E}{\partial S} \right),$$

(30)

which is proportional to

$$G_E(S) \propto \sqrt{S} \left( \frac{O_E(S)}{2} - S \frac{\partial O_E}{\partial S} \right),$$

(31)

which can be evaluated numerically to arbitrary precision and is normalized and plotted in Fig. 4 over the range $-1 < S < 0$. Figure 4 shows that $G_E(S)$ achieves its maximum value at $S \approx -0.9$, decreasing as $S$ increases until $G_E(S)$ becomes negative at $S \approx -0.25$. We discuss below the consequences of this behavior.

To conclude our exploration of the nonlinear instability, we derive expressions for $h(v_\perp)$ and its derivative that relates the magnitude of the phase space depression associated with the particle trap to the initial distribution function $f_0$. In particular, the value $h(v_\perp)$ depends on the ambient distribution function at the point in phase space where the particles are trapped. If we assume that the trapping point is some distance $\Delta z$ up the field line

$$h(v_\perp) = f_0(z, v_\perp, z),$$

$$-f_0(z + \Delta z, v_\perp(z + \Delta z), v_\perp(z + \Delta z, v_\perp)),$$

(32)

where the first term represents the ambient value of the distribution function at point $z$ along the field line, and the second term represents the ambient value of the distribution function at $z + \Delta z$ and the point in velocity space where the parallel velocity is locally resonant with the wave, and the perpendicular velocity is $v_\perp$.

For succinctness, yields

$$\frac{\partial J_E}{\partial B_w} = -qJ_0 \int_0^\infty \left\{ h(v_\perp) G_E(S) + B_w^{1/4} \frac{\partial h(v_\perp)}{\partial B_w} O_E(S) \right\} v_\perp^2 dv_\perp$$

(29)

from which we note that the sign is determined by $h(v_\perp)$ and its derivative and $G_E(S)$, since $O_E(S) \geq 0$.

Noting that $S$ is inversely proportional to $B_w$, and neglecting the effect of the wave amplitude on the frequency, (28) can be evaluated and expressed as

$$G_E(S) = \frac{1}{\sqrt{B_w}} \left( \frac{O_E(S)}{2} - S \frac{\partial O_E}{\partial S} \right),$$

(30)

which is proportional to

$$G_E(S) \propto \sqrt{S} \left( \frac{O_E(S)}{2} - S \frac{\partial O_E}{\partial S} \right),$$

(31)

which can be evaluated numerically to arbitrary precision and is normalized and plotted in Fig. 4 over the range $-1 < S < 0$. Figure 4 shows that $G_E(S)$ achieves its maximum value at $S \approx -0.9$, decreasing as $S$ increases until $G_E(S)$ becomes negative at $S \approx -0.25$. We discuss below the consequences of this behavior.

To conclude our exploration of the nonlinear instability, we derive expressions for $h(v_\perp)$ and its derivative that relates the magnitude of the phase space depression associated with the particle trap to the initial distribution function $f_0$. In particular, the value $h(v_\perp)$ depends on the ambient distribution function at the point in phase space where the particles are trapped. If we assume that the trapping point is some distance $\Delta z$ up the field line

$$h(v_\perp) = f_0(z, v_\perp(z), v_\perp(z))$$

$$-f_0(z + \Delta z, v_\perp(z + \Delta z), v_\perp(z + \Delta z, v_\perp)),$$

(32)

FIG. 3. The value of the integral in (25), which is computed by numerically integrating over the shape of the electron trap in electron velocity parallel to the static magnetic field and Larmor phase.

FIG. 4. The value of the function in (31), (top) expressed as a function of wave amplitude $B_w$, and (bottom) normalized so that the maximum value is unity.
The change in the resonant velocity between the two points is

\[ \Delta v_{\text{res}} = \int_z^{z+\Delta z} \frac{d\Omega_{\text{res}}}{dz} dz = \int_z^{z+\Delta z} \frac{3}{2k} \frac{d\Omega}{dz} dz, \]

(38)

where the second equality can be demonstrated using (18) and the expression for the whistler mode wave number

\[ k = \frac{\omega_p}{c} \left( \frac{\omega}{\Omega - \omega} \right)^{1/2}. \]

Through the interaction with the whistler mode wave, particles are forced to follow contours in phase space described by

\[ \frac{dv_{\perp}}{dv_{\parallel}} = -\frac{(\omega/k + v_{\parallel})}{v_{\perp}}, \]

so the change in perpendicular velocity due to trapping becomes

\[ \Delta v_{\perp,\text{res}} = \int_z^{z+\Delta z} \frac{d\Omega_{\text{res}}}{dz} dz = -\int_z^{z+\Delta z} \frac{\left( \frac{\omega}{k} + v_{\parallel} \right)}{v_{\perp}} \frac{d\Omega}{dz} dz \]

(39)

Finally, the changes in particle velocity due to adiabatic motion can be expressed as

\[ \Delta v_{\parallel,\text{adiab}} = \int_{z+\Delta z}^{z} \frac{v_{\parallel}^2}{2\Omega} d\frac{d\Omega}{dz} dz = -\int_{z}^{z+\Delta z} \frac{v_{\perp}^2}{2\Omega} d\frac{d\Omega}{dz} dz, \]

(40)

\[ \Delta v_{\perp,\text{adiab}} = \int_{z+\Delta z}^{z} \frac{v_{\perp}}{2\Omega} d\frac{d\Omega}{dz} dz = -\int_{z}^{z+\Delta z} \frac{v_{\perp}}{2\Omega} d\frac{d\Omega}{dz} dz. \]

(41)

Combining (37)–(41) gives the expression for \( h(v_{\perp}) \)

\[ h(v_{\perp}) = f_0(z, v_{\text{res}}(z), v_{\perp}) \]

\[ -f_0 \left( z, v_{\text{res}}(z) + \int_z^{z+\Delta z} \frac{3}{2k} \frac{v_{\parallel}^2}{2\Omega} d\frac{d\Omega}{dz} dz, v_{\perp} \right), \]

\[ -\int_z^{z+\Delta z} \left( \frac{\left( \frac{\omega}{k} + v_{\parallel} \right)}{v_{\perp}} \frac{3}{2k} \frac{v_{\parallel}^2}{2\Omega} d\frac{d\Omega}{dz} dz \right), \]

(42)

which is a general expression for the value of the distribution function in the particle trap. The dependence of \( h(v_{\perp}) \) on the wave amplitude \( B_w \) is through \( \Delta z \). Specifically, \( \Delta z \) corresponds to the distance up the field line where \( S = -1 \) for a given trapped particle, which will increase as the wave amplitude increases.

We next specialize our consideration to the initiation of the nonlinear instability, to explore the conditions under which the nonlinear instability will yield an accelerating growth rate. Thus, for an assumed location downstream of the equator, we consider \( \Delta z \) to be small, so that

\[ h(v_{\perp}) = \frac{3}{2k} \frac{d\Omega}{dz} H, \]

(44)

where

\[ H = \left\{ \begin{array}{l} \frac{1}{v_{\perp}} \left[ \left( \frac{\omega}{k} + v_{\parallel} \right) \frac{\partial f_0}{\partial v_{\parallel}} - v_{\perp} \frac{\partial f_0}{\partial v_{\perp}} \right] \\
+ \frac{k v_{\perp}}{3\Omega v_{\parallel}} \left[ v_{\parallel} \frac{\partial f_0}{\partial v_{\parallel}} - v_{\perp} \frac{\partial f_0}{\partial v_{\perp}} \right] \end{array} \right\} \]

(45)

is a factor that is related to the anisotropy of the distribution function. The first term in square brackets is the gradient of the distribution function along whistler contours and the second term in square brackets is the gradient of the distribution function along constant energy contours.

From (44) we can see that, at the initiation of the instability, the deformation of the distribution function depends on three factors: the small distance up the field line where particles are trapped, the magnitude of the inhomogeneity, and the local (i.e., near-resonant) shape of the distribution function.

Equation (44) furthermore allows us to analytically determine the function \( \partial h(v_{\perp})/\partial B_w \). As preliminary steps in the derivation, we note that \( S \) is proportional to the inverse of the wave amplitude, such that

\[ \frac{\partial S}{\partial B_w} = -\frac{S}{B_w} \]

(46)

and again assuming minimal drift in frequency, we evaluate the quantity

\[ \frac{\partial}{\partial B_w} \left( \frac{\partial S}{\partial B_w} \right) = \frac{\partial}{\partial z} \left( \frac{\partial S}{\partial B_w} \right) = -\frac{\partial}{\partial z} \left( \frac{S}{B_w} \right) \]

\[ = -\frac{1}{B_w^2} \left[ B_w \frac{\partial S}{\partial z} - S \frac{\partial B_w}{\partial z} \right] \]

\[ = -\frac{1}{B_w^2} \left[ B_w \frac{\partial S}{\partial z} - S \frac{\partial B_w}{\partial z} \right] \]

\[ = -\frac{2}{B_w^2} \left[ B_w S - S \frac{\partial B_w}{\partial z} \right] \]

(47)
With $\Delta z$ as the distance required for $S$ to go from its local value to $S = -1$, to first order

$$\Delta z = -1 - S \frac{\partial S}{\partial S}. \tag{48}$$

Taking the partial differential of (48) with respect to $B_w$ yields

$$\frac{\partial \Delta z}{\partial B_w} = \left( \frac{\partial S}{\partial S} \right)^2 - \left[ \frac{\partial S}{\partial S} \right] \frac{\partial (1 + S)}{\partial B_w} \left( \frac{\partial S}{\partial S} \right) \right] \tag{49}$$

and substituting the results from (46) and (47) into (49)

$$\frac{\partial \Delta z}{\partial B_w} = \left( \frac{\partial S}{\partial S} \right)^{-1} \left[ \frac{S}{B_w} - 2 + 2S \right] \right] \tag{50}$$

and substituting (48) into (50)

$$\frac{\partial \Delta z}{\partial B_w} = \frac{1 + 2 S}{B_w} \frac{\partial S}{\partial S}. \tag{51}$$

Finally, we substitute the results from (44) and (51) into (29)

$$\frac{\partial I_E}{\partial B_w} = -q_0 \frac{3 d\Omega}{2k \Delta z} \times \int_0^\infty H \left\{ G_E(S) + B_w^{1/2} \frac{2 + S}{1 + S} O_E(S) \right\} e^{S/2} d\varphi, \tag{52}$$

which is the expression for the instability factor at the initiation of the nonlinear instability. In the preceding analysis we have not considered significant frequency change during the interaction, except through change in the value of $S$ [see Eq. (22)]. This is justified by the fact that we are primarily interested in the initial growth phase and its dynamics. Experimental data presented by Paschal and Helliewell,32 and more recently by Golkowski et al.,31 show that the initial growth phase occurs over a small bandwidth before free running emissions (most often risers) are triggered. In the context of chorus waves, Omura et al.,25 claim that the rising frequency-time signature is self selected from a broadband background of noise based on optimization of nonlinear growth and the frequency of wave does not change as it propagates. Omura and Nunn35 on the other hand, claim that the frequency of the chorus packet changes as the wave propagates. When frequency change occurs in chorus is harder to interpret from observations. A study by Santolík et al.36 on in situ chorus observations reports a relationship of peak amplitude decreasing with frequency. However, the Santolík et al.36 frequency-amplitude correlation has a significant spread and is based on peak amplitude versus normalized frequency, where the normalization is to the background cyclotron frequency. Therefore, their result it is not an amplitude versus frequency change of individual emissions but an aggregate amplitude versus normalized frequency compilation. As the study herein focuses on the inception of the nonlinear instability, the assumption of small frequency change as a function of wave amplitude made in the derivation of (30) is seen as justifiable.

III. DISCUSSION

Equations (16) and (52) are expressions for $\partial I_E/\partial B_w$, the variation of the amplitude of the resonant current parallel to the wave electric field with respect to the amplitude of the wave, that are applicable at the initiation of the linear and nonlinear phases of the whistler mode instability, respectively. For both cases, the sign of $\partial I_E/\partial B_w$ is controlled by the value of the integral. In particular, when the sign of the integral is positive, $\partial I_E/\partial B_w$ is negative, leading to unstable growth.

In the linear case, as shown in expression (16), noting $v_+ \geq 0$, the sign of the integral is determined solely by terms in curly braces, which is the familiar condition of anisotropy that is required for the linear whistler mode instability.

For the nonlinear case, three factors contribute to the sign of $\partial I_E/\partial B_w$. First, the sign of $\partial I_E/\partial B_w$ is determined by the nature of the inhomogeneity $d\Omega$, which is non-negative for the formation of an electron hole. A negative value of $d\Omega$ corresponds to a value of $S \geq 0$, which is inconsistent with the assumptions made by deriving (25). In particular, the relationship $\zeta_1 < \zeta_2$ is assumed. In the case of a positive $S$, the shape of the trap is flipped in $\zeta$-space, and the limits of integration must be reversed. Therefore, although $d\Omega$ is negative, the entire result in (52) changes sign, and the effect of the inhomogeneity on the sign of $\partial I_E/\partial B_w$ is the same. This situation corresponds to the case of the formation of an electron ‘hill’ or ‘island’ as the particle trap forms. The second factor $H$ is an anisotropy requirement, similar to the corresponding term in curly braces in (16) yet differing in that there is an additional term associated with the anisotropy along constant energy contours. This term arises because the electrons in the particle trap come from resonant flux farther up the field line. As the gradient along constant energy contours increases in magnitude, the rate at which the resonant flux decreases as a function of location along the field line similarly increases, leading to a ‘deeper’ electron hole.

Theoretically, this additional anisotropy term indicates that there could be a distribution function that is stable to the linear instability, yet unstable to the nonlinear instability. However, the contours defined by the bracketed terms in (45) are similar enough that is difficult to imagine when this situation would occur in practice.

The third and final term in (52) that could affect the sign of $\partial I_E/\partial B_w$ is the quantity in curly braces

$$G_E(S) + B_w^{1/2} \frac{2 + S}{1 + S} O_E(S)$$

which is normalized and plotted in Fig. 5. From the figure, it is clear that this term is always positive, indicating that the stability of the nonlinear interaction will be controlled by the anisotropy of the distribution function. Furthermore, this
term is maximum near $S \approx -0.86$, indicating that the growth rate is changing fastest just after trapping is initiated, where low wave amplitudes, wave frequency changes, and the inhomogeneity of the magnetic field combine to reduce the effectiveness of the trapping process.

These results describing the initiation of the instability depend on the derivation of (52), which assumes that the length of the trap $\Delta z$ is small. Specifically, the magnitude of $\partial J_E/\partial B_w$ in (52) is proportional to $\Delta z \partial^2 J_E/\partial t^2$, and the dependence on the distribution function can be expressed in terms of gradients. If the short trap assumption is relaxed, (42) must be used to express the depth of the phase space hole. In that case, the value of $\partial J_E/\partial B_w$ will change in particular, with regard to the functional dependence on the unperturbed distribution function, but the general conclusions about the instability will remain unaffected.

For the specialized case analyzed by Omura et al. and Omura et al., the distribution function was assumed to be a delta function in perpendicular velocity and the nonlinear instability was assumed to occur near the equator (that is, $\partial^2 J_E/\partial t^2 \approx 0$). In this case, the value of $\partial J_E/\partial B_w$ will remain small unless $S \approx -0.86$, where the term in (53) becomes large. Considering (22), the only way for that requirement to be satisfied when $\partial^2 J_E/\partial t^2 \approx 0$ is if $\partial J_B/\partial t > 0$, which is consistent with the findings of the study.

A marked contrast arises when considering a case that is frequently seen in the data from controlled experiments of triggered emissions. There, it was found that a constant frequency input signal (that is $\partial J_B/\partial t = 0$) regularly initiates the nonlinear instability, which only later evolves to a “free running,” usually, rising frequency emission. With a constant frequency trapping wave, the value of $S$ is controlled by the inhomogeneity in the static magnetic field. For small $S$ (near the equator), the term in (53) becomes small. Thus, we can conclude that the instability must initiate off the equator for a constant frequency triggering input. This is not to say, that a trap cannot exist at the magnetic equator; it merely cannot lead to accelerating growth when initiated by a constant frequency wave. This finding is consistent with the numerical results of Gibby et al. and Harid et al., who showed that the nonlinear instability for a constant frequency triggering source initiated off the equator.

The comparison of these two cases implies two distinct phases of the instability, when it is triggered by a constant frequency input. The first phase is triggered growth, in which the instability starts at a location downstream of the equator and expands until the equator is included in the trapping region. This is displayed in the experimental data as the initial growth phase of the input wave when the phase advance is small. The second phase is the generation of a free running emission, which is established at the equator with a varying wave frequency $\partial J_B/\partial t > 0$ just as in naturally occurring chorus emissions. An experimental observation of controlled wave injection using the High Frequency Active Auroral Research Program (HAARP) facility described by Gołkowski et al. shows that the transition from the first phase to the second phase is remarkably consistent even for greater than 10 dB differences in amplitudes of the injected input waves. The input waves with lower amplitudes were found to grow in the linear regime for a longer duration, but once their amplitude exceeded the nonlinear amplitude threshold, the subsequent amplitude and frequency characteristics were identical to that of the higher input amplitude waves.

This transition between constant frequency temporal growth and free running emission most likely occurs near or at the equator. Specifically, Omura et al. showed that a wave packet does not experience significant frequency change when $|J_E| \approx |J_B|$, where $J_B$ is the component of the resonant current parallel to the wave magnetic field. Omura et al. additionally show that this condition on the resonant current is generally true whenever conditions are favorable for growth ($S \leq -0.4$). However, as we have shown, the conditions near the equator for a constant frequency wave are not favorable for growth, and the ratio $J_B/J_E$ can be significantly greater than unity when $S \approx 0$. This may open a small window under which the currents driving frequency change may cause the transition between the two phases of the instability.

Our heretofore description of the instability, however, leaves one additional aspect of the phenomenon undescribed. So far, we have not addressed the case for which the emission frequency follows $\partial J_B/\partial t < 0$, which is observed in the experimental data. Falling emissions have been investigated by Nunn et al., Nunn and Omura, Trakhtengerts et al., and Demekhov. Many approaches like the recent work by Nunn and Omura suggest that the fallers are generated before the wave reaches the equator in region of positive $S$ values that results in the formation of a “hill” instead of hole in phase space. In the constraints of the current article, we address the condition of the falling frequency emission in qualitative terms. The experimental data from controlled triggering experiments show that falling frequency emissions are more likely to occur for shorter duration input signals. From the description of the triggered growth phase above, this implies that the input signal is terminated before the trapping region expands to include the
equator. If the instability occurs at a position along the field line that is far enough from the equator that the second term in (22) drives $S$ to be less than $-1$, then the only manner in which the instability can continue is by achieving a value of $\partial \sigma / \partial t < 0$.

IV. CONCLUSIONS

In summary, we have presented a new method of regarding the nonlinear whistler mode instability that has led us to make a number of analytical conclusions. Specifically, instead of regarding the instability in terms of the dispersion relation and an imaginary growth term, we have presented the instability in terms of the evolution of the resonant current with respect to the wave amplitude. This, along with a simplification of the current integral presented by Omura et al., has given us a description of the conditions required for a rapidly increasing growth rate of the nonlinear instability.

In particular, the requirements on the anisotropy of the distribution function are modified relative to the requirements for the linear instability. Furthermore, we have shown that the initiation of the nonlinear instability depends on the condition that singular values of $S_{-0.9}$ as being the dominant condition of nonlinear growth. Taking $S = -0.4$ as the dominant condition allows for calculation of a frequency sweep rate and has been used to put forth concepts of an “optimum amplitude” and “threshold amplitude” where the former derives from maintaining this sweep rate and the latter is defined by the necessary spatial gradient of wave amplitude to maintain the $S = -0.4$ condition as the wave propagates away from the equator. While $S = -0.4$ conditions emphasize the maximum possible nonlinear growth rate, the $S = -0.9$ conditions put forth here highlight the portion of the wave-particle interaction where the amplitude and growth rate are changing the fastest and the interaction is at the edge of its nonlinear phase. In this context, it is important to note that singular values of $S$ are convenient for theory but in the physical situation may be less well defined. However, since the phase space hole model has recently been considered to also be the main process in generation of less coherent hiss emissions, it is important that all regimes in the model are thoroughly investigated.

Although we have demonstrated that our results are consistent with observations and recent simulations of the nonlinear instability, it is clear that new in situ observations are needed to firmly validate existing nonlinear theories. In particular, simultaneous multi-spacecraft measurements with probe separation on the scale of a small number of whistler mode wavelengths would allow for observing the nonlinear process as it evolves.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation with Grant Nos. AGS 1254365 and PLR 1542608 to the University of Colorado Denver.