



## Saturation effects in the VLF-triggered emission process

A. R. Gibby,<sup>1</sup> U. S. Inan,<sup>1</sup> and T. F. Bell<sup>1</sup>

Received 31 March 2008; revised 20 July 2008; accepted 2 September 2008; published 22 November 2008.

[1] An in-depth study is performed of the saturation characteristics of the VLF-triggered emission process, a plasma instability associated with the amplification of whistler mode signals in the magnetosphere. A survey of data from the 1986 operating year of the Siple VLF wave injection experiment reveals that long-period oscillations (characterized by a pattern of growth to saturation then subsequent suppression of the wave), short-period oscillations (previously identified as sidebands), and generation of incoherent wave energy around the frequency of the triggering signal are all characteristics of the instability at saturation. A model is developed to study these phenomena. The model is in some respects similar to the Vlasov Hybrid Simulation but modified so that saturation is a natural consequence of the modeled growth process. Results from the model indicate that the growth and eventual saturation are caused by wave amplitude gradients, most notably gradients that indicate a transition from an amplitude where the wave cannot trap electrons in its potential in the presence of the inhomogeneous magnetospheric magnetic field to an amplitude where such trapping is possible. That is, resonant currents are enhanced in the region where the phase space electron hole created by the trapping condition is allowed to mix with the ambient energetic electrons. This growth process is found to naturally lead to saturation of the instability.

**Citation:** Gibby, A. R., U. S. Inan, and T. F. Bell (2008), Saturation effects in the VLF-triggered emission process, *J. Geophys. Res.*, 113, A11215, doi:10.1029/2008JA013233.

### 1. Introduction

[2] VLF-triggered emissions [Helliwell, 1967; Omura *et al.*, 1991] occur when a whistler mode signal excites the magnetospheric plasma in such a manner as to cause growth of the initial wave and generation of coherent (but often varying in frequency) wave energy at frequencies other than that of the initial wave. It is generally thought that this instability is caused by a cyclotron resonance interaction between ducted whistler mode waves and counter-streaming energetic radiation belt electrons. Linear plasma theory predicts that cyclotron resonance amplification of any whistler mode wave will occur in a magnetoplasma whose velocity distribution function is sufficiently anisotropic [Kennel and Petschek, 1966; Liemohn, 1967; Matsumoto and Kimura, 1971]. The resonance interaction can be further enhanced (in a nonlinear manner) by the trapping of resonant particles in the wave potential in the presence of the inhomogeneous magnetospheric magnetic field [Nunn, 1974]. That is, instead of following adiabatic trajectories dictated by the ambient magnetic field gradients, a certain fraction of the near resonant particles are constrained in gyrophase and velocity parallel to the ambient magnetic field in such a manner that they remain in resonance with the wave. The perturbed particle motion due to this trapping

process enhances currents that radiate in a manner so as to amplify the wave field.

[3] A characteristic of VLF-triggered emissions is one of exponential growth of the initial whistler mode wave to a saturation level [Helliwell and Katsufurakis, 1974], which typically remains constant during the course of observation episodes (many minutes to tens of minutes). If the amplification process were linear, all parts of a finite pulse would be amplified equally, and no growth would be observable at a stationary receiver. Since this amplification is observable as temporal growth at a stationary receiver, the amplification must be driven by a nonlinear process, neglecting any initial rise-time effects. Consequently, saturation is expected to result from some disruption of the nonlinear growth process.

[4] An obvious cause of saturation of any plasma instability is the removal or exhaustion of the free energy for growth. Owing to the results of linear theory, it has long been thought [e.g., Das, 1968] that the exhaustion of free energy for the instability involves the destruction of the anisotropy of the resonant particle distribution. With varying details, many authors agree that the reduction of anisotropy is caused by the phase-space mixing of trapped particles in the wave potential well [Das, 1968; Dysthe, 1971; Cornilleau-Wehrin and Gendrin, 1979; Matsumoto *et al.*, 1980]. That is, particle trapping initially enhances resonant currents causing wave growth. As the particles are trapped, they begin to oscillate in the wave potential, the frequency of which is determined by the wave amplitude and the particle's velocity perpendicular to the ambient

<sup>1</sup>Space, Telecommunications, and Radioscience Laboratory, Stanford University, Stanford, California, USA.

magnetic field. Subsequently, the wave amplitude increases owing to the action of the enhanced resonant currents, causing more rapid oscillations and mixing of trapped particles. This mixing acts to reduce the anisotropy of the trapped particles, reducing the magnitude of the enhanced currents, and eventually slowing growth so that the wave amplitude reaches a maximum level.

[5] *Dowden et al.* [1978] and *Omura and Summers* [2006] note that trapping of the particles creates a separatrix in phase space. According to these authors, the whistler mode wave is expected to reach its highest amplitude after it has passed the magnetic equator. If the amplitude reached via linear growth is sufficiently large to trap particles, then the separatrix is formed. This separatrix prevents mixing between the trapped and untrapped populations, which in turn generates a phase space hole as the trapped particles travel toward the equator on non-adiabatic paths into regions of higher phase space density. Despite the rapid mixing of particles within the separatrix, resonant currents are enhanced. If the region in which particle trapping occurs expands along the field line back toward the equator, the total wave growth increases. This growth proceeds until the expanding trapping region reaches the equator, at which point the location of the hole in phase space transitions to a position where the lower density of the hole generates resonant currents that reduce the wave amplitude, causing saturation.

[6] Another possible mechanism for saturation when a particle trap is formed beyond the equator is due to energy exchange between the particles and the wave. Non-trapped resonant particles interact with the wave in such a manner as to lose energy to the wave on average, while the trapped particles are, on average, accelerated by the wave [*Bell and Inan*, 1981]. As the wave amplitude increases and the trapping region expands, the number of trapped particles accelerated by the wave increases. Conversely, the increase wave amplitude scatters the untrapped population more strongly as well. If more energy is expended in accelerating the trapped population than is gained from the untrapped population, wave growth slows, eventually stalling.

[7] Recent self-consistent computer simulations reproduce many of the characteristics of whistler mode amplification that are also associated with the phenomenon of VLF-triggered emissions, including exponential growth to saturation [*Omura and Summers*, 2004] and generation of rising frequency emissions [*Kato and Omura*, 2006]. However, these simulations do not elucidate the underlying components of the instability that control saturation.

[8] For example, *Omura and Summers* [2004] explore the amplification of wideband whistler mode waves by a relativistic, anisotropic distribution. Their simulation produces strong amplification over a narrow frequency range that is characterized exponential growth to saturation. To explain this saturation, the authors compare the electron distributions at the start and finish of the simulation and note that the difference is qualitatively similar to differences that would be caused if quasilinear diffusion were the active process. However, the authors do not explain how this diffusion occurs dynamically as the amplification process progresses. Thus, since these large, self-consistent simulations appear to contain all of the relevant physical processes, they may effectively reproduce the macroscopic character of

the modeled events. However, because of their completeness, it is difficult for such simulations to delineate and identify portions of the physics which control different aspects of the triggered emission process, such as saturation.

[9] Computer simulations using the Vlasov Hybrid model [*Nunn*, 1990, 1993; *Nunn et al.*, 1997, 2003, 2005; *Smith and Nunn*, 1998] have also reproduced many observed aspects of triggered emissions, but have been unable to demonstrate the saturation effect. Instead, saturation is imposed artificially with the supposition that it is caused by effects not encapsulated by the model, such as electrostatic damping or leakage from the whistler mode duct.

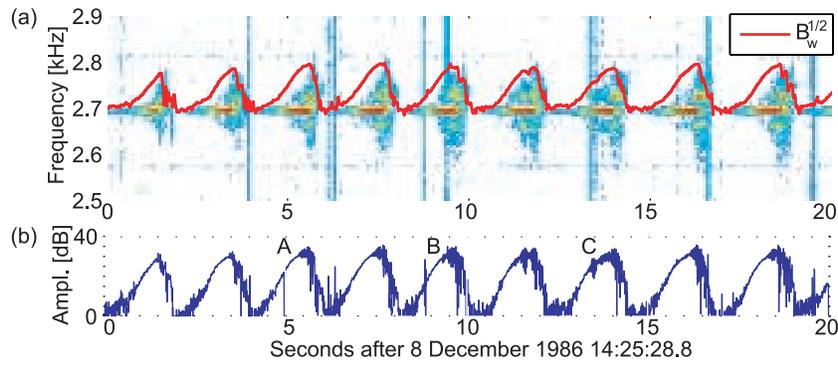
[10] The many competing explanations of saturation of growth in this coherent instability underscore the need to understand the manner in which saturation occurs dynamically. We start by presenting observational evidence for two effects at saturation: the generation of amplitude oscillations and the generation of wideband whistler mode energy. We then develop a numerical model similar to the Vlasov Hybrid model to simulate the wave-particle interaction process. Using the new model we are able to reproduce the saturation characteristics of the VLF-triggered emission process and explain the physical process by which saturation occurs.

## 2. Observations

[11] Data from the Siple Experiment [*Helliwell*, 1988] provides the primary source of experimental observations of VLF-triggered emissions. Operating at  $L = 4.3$ , the Siple transmitter injected waves from the ground into the magnetosphere, providing a controlled experiment for the study of triggered emissions. In the later stages of the experiment, the transmitter could be modulated in power and frequency and could transmit multiple frequencies simultaneously. Recently, an effort has been undertaken to transfer the Siple data from its legacy analog magnetic media to digital formats. The existence of the Siple data in digital form allows more detailed and rapid analysis of the data. The availability of digital analysis sheds light on previously unexplained phenomena, such as the saturation effect.

[12] Examination of the Siple data from 1986 reveals many events that clearly display saturation. During this period, Siple transmission formats were designed to explore a variety of effects, which is reflected in the complexity of the transmitted formats. To isolate the saturation phenomenon, we only consider simple transmissions that were typically embedded within more complex formats. Specifically, we select constant-frequency transmissions when Siple was transmitting only a single frequency in the form of a continuous wave. During this period, virtually all constant frequency transmissions contained a power ramp, i.e., the transmission power was increased slowly (typically at 10 dB/s) from some initial level to a full-power level. The purpose of this power ramp was to study the instability threshold, in which temporal growth at the receiver is not initially observed from low-amplitude input waves, but growth subsequently begins as the input wave amplitude is increased [*Helliwell et al.*, 1980].

[13] Additionally, because the signature of saturation could easily be masked by interference between signals received with different delays, events are selected which



**Figure 1.** Siple transmission from 8 December 1986 showing saturation characteristics. The Siple transmission is a constant frequency key-down signal with a transmission frequency of 2700 Hz. (a) Time-frequency spectrogram showing 20 s from the event. The development of wideband wave energy is clearly visible. Overlaid is a plot of the square root of the narrowband amplitude at the transmitted frequency. The frequency spread of the wave energy is proportional to the square root of the wave amplitude and is thus likely caused by the trapping of resonant electrons in the wave potential well. (b) Narrowband amplitude plot from a 120 Hz bandwidth filter centered at 2700 Hz.

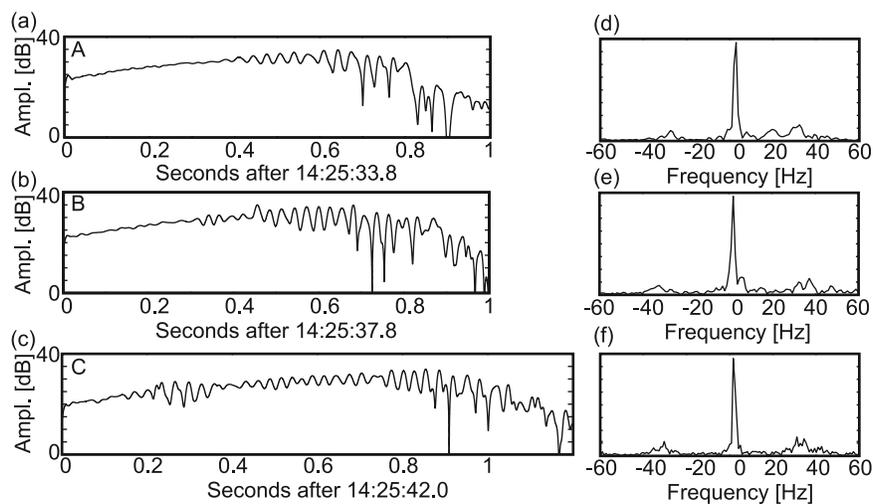
show minimal evidence of excitation of multiple magnetospheric ducts. Only data from periods of low background noise, with signal-to-noise ratios greater than 20 dB, are considered.

**2.1. Amplitude Effects at Saturation**

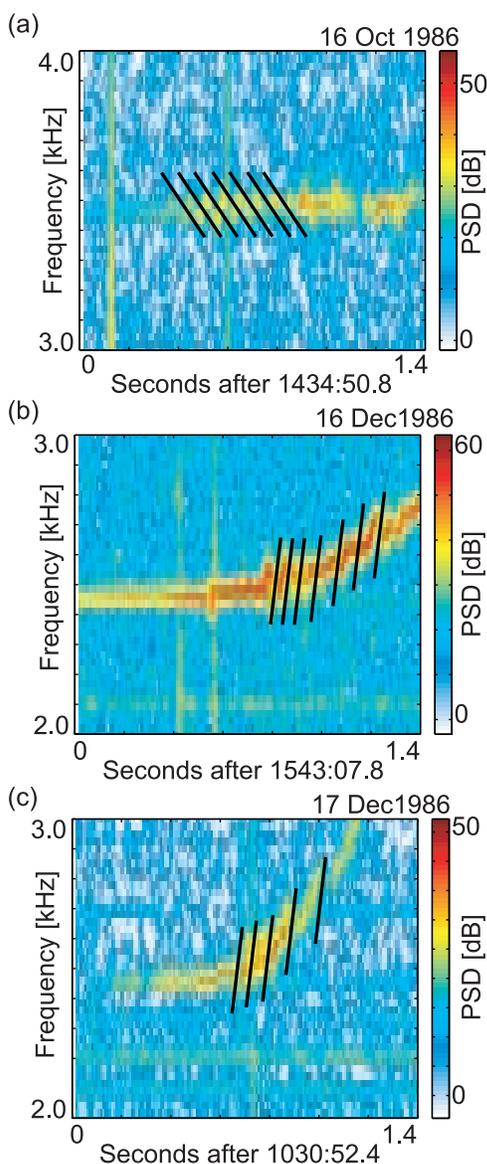
[14] Oscillations in saturated whistler mode signals were first observed by *Bell and Helliwell* [1971]. Since saturation and subsequent re-growth is arguably caused by the removal and subsequent restoration of free energy for wave growth, it is natural that an oscillatory phenomenon is associated with saturation.

[15] Figure 1 shows an event from 8 December 1986. Figure 1a is a 20-s time-frequency spectrogram, showing

the reception of a long, 2700 Hz, key-down transmission at Lake Mistissini, Quebec, conjugate to Siple Station. The transmission is one of a series of key-downs that were regularly broadcast to evaluate the tuning of the Siple transmitter. Figure 1b shows the signal amplitude in a 120 Hz bandwidth during the same 20-s period, centered at 2700 Hz. The transmission is initially received at the beginning of the record and grows exponentially to saturation after approximately 1.5 s. At that point, the signal starts to exhibit a regular oscillation, characterized by a period of 200–500 ms at saturation, followed by rapid suppression of the signal back to the noise level, and concluded by renewed growth to saturation at the same rate as the initial



**Figure 2.** Siple transmission from 8 December 1986 showing saturation characteristics. The Siple transmission is a constant frequency key-down signal with a transmission frequency of 2700 Hz. (a–c) Expanded amplitude plots of maxima labeled “A,” “B,” and “C,” respectively, in Figure 1b. The plots show short-period oscillations develop near saturation. (d–f) Fourier transforms of the data shown in Figures 2a, 2b, and 2c, respectively. The frequency content of the short-period oscillations in all three cases is concentrated near 30 Hz.



**Figure 3.** Time-frequency spectrograms of three events showing short-period amplitude oscillations associated with saturation. Minima in the oscillations are highlighted by lines. The angle of the lines is attributed to whistler mode dispersion.

growth. The period of the oscillation is  $\sim 2$  s, and we term this pattern the “long-period saturation oscillation.”

[16] Suppression of the wave has been reported in the presence of free running plasma emissions triggered after saturation [Stiles and Helliwell, 1977; Chang and Helliwell, 1979], in which the wave grows to saturation, triggers a rising frequency emission, and the wave intensity at the transmitted frequency is suppressed until the emission rises significantly in frequency. Once the emission has risen in frequency by  $\geq 100$  Hz, the wave energy reappears at the transmitted frequency, growing at the same rate as it was initially. The event in Figure 1 is consistent with these characteristics, except that the suppression occurs without the triggering of a free running emission. The occurrence of

suppression without triggering indicates that triggering is not a necessary component of this type of suppression phenomenon. Instead, it appears to be an inherent part of the saturation process.

[17] Raghuram *et al.* [1977] found that whistler mode two-hop echoes from Siple transmissions can cause a disruption or suppression of the growth process. Typical two-hop magnetospheric echo times for Siple signals were  $\sim 4$  s. Because of the shorter  $\sim 1.5$  s onset of the long-period oscillation shown here, it cannot be attributed to such echo-induced suppression.

[18] In addition to the long-period oscillation, a short-period oscillation is also observed at saturation. Figures 2a–2c show expanded amplitude plots at saturation, corresponding to the periods marked “A,” “B,” and “C” in the 20-s amplitude plot of Figure 1b. In all three plots, a clear, nearly-sinusoidal modulation develops near saturation. The frequency of oscillation varies slightly between different maxima, but is measured at  $\sim 20$ –40 Hz.

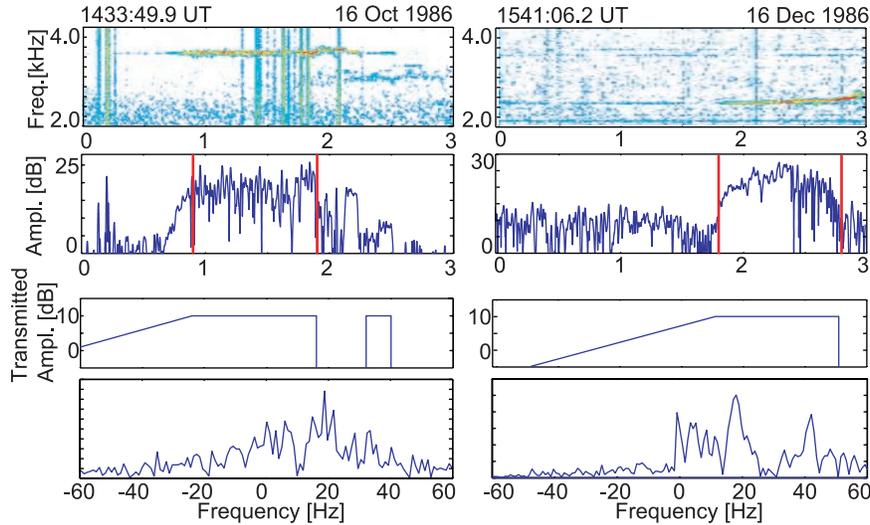
[19] Figure 3 shows spectrograms of three additional events in which short-period oscillations are visible. In each of these, low-amplitude troughs of the oscillation are highlighted by tilted lines. The tilt of the troughs on the spectrogram is attributed to dispersion and confounds the measurement of clear oscillations using standard narrow-band filtering. Thus, the tilt in Figure 3a is in the opposite sense of the tilts in Figures 3b and 3c because the transmitted frequencies are below and above the frequency that corresponds to the minimum path delay along the field line, respectively.

[20] Both Siple transmissions in Figures 3b and 3c trigger rising emissions. Note that the oscillation continues through the development of the emission. Also, as the emission rises in frequency, the oscillation period increases. The signal in Figure 3a remains relatively constant in frequency, and little variation in the oscillation period is observed.

## 2.2. Frequency Effects at Saturation

[21] Sidebands in the triggered emissions process were discussed by Brinca [1972] and were first reported for the Siple transmitter experiments by Park [1981]. Spacecraft observations and measurements of sidebands and their associated characteristics were also reported by Bell [1985]. Detailed frequency measurements of sidebands were subsequently performed by Sa and Helliwell [1988]. This work was later expanded into a wave-wave interaction theory [Sa, 1990] proposing that sideband generation was a result of nonlinear interactions of two or more waves closely spaced in frequency. In the case that single frequency transmissions created sidebands, the existence of undetectable power line radiation was posited as the facilitating second wave for sideband generation.

[22] A simpler explanation for the existence of sidebands might be to recognize them as the frequency domain manifestation of the inherent modulation of the amplified signal at saturation. That is, the Fourier decomposition of the saturation oscillations is revealed in the spectrogram as sidebands. Figures 2d, 2e, and 2f are the Fourier transforms of the time series data shown in Figures 2a, 2b, and 2c, respectively. Figures 2d, 2e, and 2f all show sidebands  $\sim 30$  Hz above and below the frequency of the transmitted signal. These sidebands are frequency domain manifesta-



**Figure 4.** Two events showing the development of sidebands associated with short-period oscillations. The top two plots are time-frequency spectrograms between 3 and 4 kHz. The second set of plots are the narrowband amplitude plots of the signals received at Lake Mistissini. The third set of plots represent the amplitude as a function of time of the signal as transmitted. The bottom plots are the Fourier transform of the data between the vertical lines in the 3-s amplitude plots.

tions of the amplitude modulation associated with the short-period oscillations that are visible in the amplitude plots.

[23] Sidebands are a common occurrence at saturation. Figure 4 shows two similar events from 16 October 1986 and 16 December 1986. Both events are from constant frequency pulses that are ramped in amplitude. The transmitted signals remain below their respective amplification thresholds until just before they reach their full-power levels. Exponential growth to saturation subsequently ensues, leading to the development of short-period oscillations, albeit less clearly than in the 8 December case. The frequency plots allow us to measure the short-period oscillation frequency, with the predominant modulation being at 18 Hz above the transmitted frequency in both cases. Consistent with *Park* [1981], the sideband development is asymmetric, with more intense sidebands above the transmitted frequency.

[24] Other frequency domain effects are also detectable. In addition to sidebands, the spectrogram in Figure 1a also shows wave energy spreading in a continuous band approximately 100 Hz above and below the transmission frequency. Overlaid is a plot of the square root of the normalized wave amplitude,  $B_w^{1/2}$ , in a 20 Hz bandwidth around 2700 Hz. It is apparent that the observable frequency spread (that is, the bandwidth of wave energy within 20 dB of the saturation amplitude) is proportional to  $B_w^{1/2}$ .

[25] This measurement of the frequency spread of the wave allows us to estimate wave amplitudes during the growth and saturation process. Since the frequency spread is proportional to  $B_w^{1/2}$ , and consequently to the trapping bandwidth (see equation (1)), the spread is likely caused by the scattering of electrons trapped in the potential well of the wave. This observation in turn allows us to bound the trapping bandwidth on the lower end (larger frequency spreads are potentially lost in the noise), and thus obtain an estimate of in-duct wave amplitudes at saturation. With

the observed omission bandwidth  $f_i$  defined to be twice the trapping frequency, we have [e.g., *Omura et al.*, 1991]

$$f_i \equiv \frac{1}{\pi} \left( kv_{\perp} \frac{e}{m_e} B_w \right)^{1/2} \quad (1)$$

the cyclotron resonance condition

$$f_c - f = \frac{k_{\parallel} v_{\parallel}}{2\pi} \quad (2)$$

and the pitch angle of interacting particles  $\alpha$

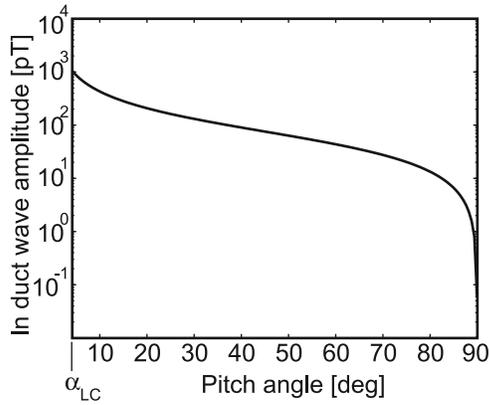
$$\tan \alpha \equiv \frac{v_{\perp}}{v_{\parallel}} \quad (3)$$

we obtain a formula for the wave amplitude as a function of the trapping bandwidth

$$B_w = \frac{\pi m_e}{2q} \frac{f_i^2}{f_c - f} \cot \alpha \quad (4)$$

where we have assumed parallel propagation, with wave number  $k_{\parallel}$ ;  $e$  is the fundamental charge;  $m_e$  is the mass of the electron;  $v_{\parallel}$  and  $v_{\perp}$  are the parallel and perpendicular velocities of the interacting particles;  $B_w$  is the wave amplitude;  $f_c$  is the electron gyrofrequency;  $f$  is the transmitted wave frequency; and  $f_i$  is the trapping bandwidth.

[26] The observable frequency spread measured in Figure 1a is  $\sim 220$  Hz around the transmitted frequency. From group delay measurements of stepped-frequency Siple transmissions during the same hour and using whistler diagnostic techniques [*Park*, 1972; *Sazhin et al.*, 1992], we find that the observed signals propagated near  $L = 4.7$



**Figure 5.** Estimate of the minimum wave amplitude required to produce the frequency spread observed in Figure 1a as a function of the pitch angle of trapped particles.

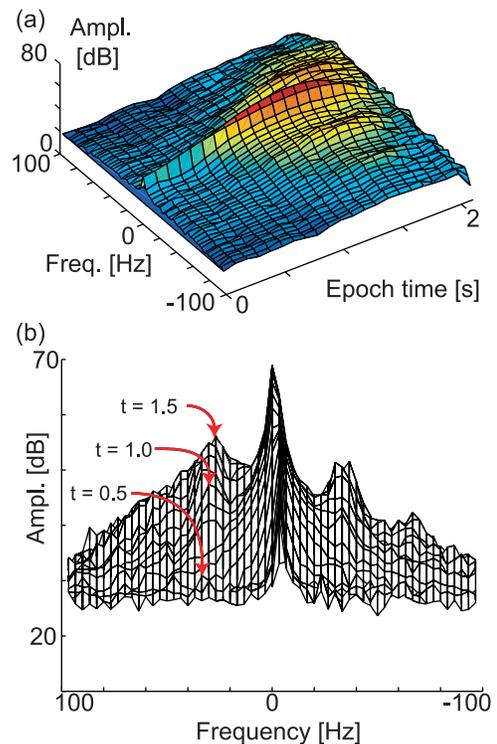
and obtain the result shown in Figure 5. The minimum wave amplitude for trapping particles at the loss cone for that bandwidth is  $\sim 1000$  pT. However, it is more likely that particles with such low pitch angles do not contribute to the observable frequency spread. Trapping particles with pitch angles of approximately 40–60 degrees at that bandwidth requires a minimum in-duct wave amplitude at saturation of  $\sim 50$ –100 pT in order to produce the observed frequency spread. This range of pitch angles is considered by some to be dominant in the amplification and triggering process [Nunn *et al.*, 2005]. On the other hand, POLAR spacecraft observations [Bell *et al.*, 2000] suggest that particles with pitch angles  $>70$  degrees are dominant in the triggering process, in which case, the in-duct wave amplitudes would be  $\leq 5$  pT at saturation.

[27] In addition to direct measurement of the frequency spread, the temporal development of this frequency spread can be observed. Because of the uniformity of each of the long-period oscillation cycles, we can use epoch averaging to more precisely evaluate the time evolution of the frequency behavior of the event in the average sense. Using each long-period oscillation from the 8 December event as an epoch, the Fourier transform of 300 ms windows within the epoch is taken, with each window overlapped by 200 ms. The Fourier amplitudes are then averaged over all the epochs, leading to the results in Figure 6. Figure 6 is, effectively, an average spectrogram of all of the long-period oscillations, plotted as a three-dimensional surface. Figure 6a shows the surface from an isometric perspective. Looking along the time axis, Figure 6b shows the evolution of the frequency behavior during the growth and saturation phases of the event. During growth, the bandwidth increases with a roughly exponential envelope around the transmitted frequency. Furthermore, as the wave amplitude at the transmitted frequency increases, the rate of falloff with frequency decreases. This wideband wave energy appears to be continuously generated through the growth phase of the wave, steadily increasing in intensity and bandwidth as the process approaches saturation. In contrast, the sidebands, which are visible as discrete peaks in the spectrum approximately 35 Hz above and below the central frequency, develop once the wave saturates, as a result of the short-

period modulation. Thus, we identify the spectral presence of sidebands as the frequency-domain manifestation of saturation, rather than being caused by the more exotic processes proposed by Brinca [1972] or Sa [1990].

### 3. Model Description

[28] The VLF-triggered emission phenomenon has been modeled in a self-consistent manner with some success using particle-in-cell codes [Kato and Omura, 2006; Lampe *et al.*, 2006] and a Vlasov Hybrid code [Nunn, 1990, 1993]. As mentioned above, the particle-in-cell method can very accurately model the physics of the interaction, but have so far been either computationally expensive or limited to considering very simple hot electron distributions. The Vlasov Hybrid approach is attractive in that it allows the tracking of the evolution of the distribution function (more precisely, the evolution of the energy change of randomly sampled points in the input distribution function), from which we can develop insight into the physical processes involved. The model is limited, however, by the use of an artificial parameter to implement saturation as well as additional filtering of the wave fields to ensure system



**Figure 6.** Frequency-amplitude measurements for superposed epoch analysis of the event in Figure 1. Each long-period oscillation is considered an individual epoch, and the frequency measurements are averaged over the epochs, producing a single long-period oscillation. The measurement shows the clear development of sidebands associated with saturation, as well as the production of wideband waves. The frequency spread of the signal is exponential, with a decreasing falloff with increasing wave amplitude.

stability. Here we develop a similar model, tracking the distribution function as it evolves in time and space, and do not use any artificial parameters or assumptions concerning saturation.

[29] Because of the particular arrangement of the Siple wave-injection experiments, involving observations of the output of the magnetospheric interactions on the ground in the conjugate region, we can assume ducted propagation. That is, magnetospheric field-aligned cold plasma density perturbations act as guiding structures to keep the whistler mode wave normal vector closely aligned to the magnetic field, which is necessary for signal reception at the conjugate ground station. This assumption allows us to ignore variations transverse to the magnetic field, so that the model can be constructed in a single spatial dimension.

[30] We specify the wave magnetic field as

$$\tilde{B}_w = B_w \exp \sqrt{-1}(\phi + \Phi) \quad \Phi = \omega_0 t - \int k_0(\omega_0, z) dz \quad (5)$$

where  $B_w$  is the wave amplitude,  $\Phi$  is the rapid spatial and temporal variation of the wave field, and  $\phi$  is any remaining phase variation after the rapidly varying contribution  $\Phi$  is removed. Stated differently, the quantity  $\Phi$  is the expected variation of the wave phase without the influence of the interaction, and  $\phi$  represents any accumulated phase due to the interaction. Thus, the quantity  $\omega_0$  is the initial angular wave frequency,  $k_0$  is the initial value of the wave number, which varies spatially owing to the magnetic inhomogeneity and variation of the cold plasma density along the field line, and the sum  $\phi + \Phi$  is the instantaneous phase of the wave as a function of position and time. Here,  $z$  is the distance along the geomagnetic field line, and we have assumed that the wave is propagating northward in the  $+z$ -direction.

[31] Neglecting the displacement current, due to the slow-wave nature of whistler mode in the medium, and dividing the plasma conduction current into two components: a cold component, which controls the propagation characteristics of the wave, and a resonant component  $J_R$ , which is created by the resonant interaction between the wave and the energetic electrons, we can derive the following equations for the wave fields from Maxwell's equations and the linearized equation of motion for the cold electrons, assuming that the wave is monochromatic [Nunn, 1974; Katoh and Omura, 2006; Omura and Summers, 2006],

$$\frac{\partial B_w}{\partial t} + v_g \frac{\partial B_w}{\partial z} = -\frac{\mu_0}{2} v_g J_E \quad (6)$$

$$\frac{\partial \phi}{\partial t} + v_g \frac{\partial \phi}{\partial z} = -\frac{\mu_0}{2} v_g \frac{J_B}{B_w} \quad (7)$$

where  $J_E = -J_R \sin \Psi$  and  $J_B = J_R \cos \Psi$  are the components of the resonant current in the direction of the wave electric and magnetic field vectors, respectively, with  $\Psi$  being the angle between the wave magnetic field vector and the resonant current vector, and  $v_g$  is the wave group velocity, as defined by the rapid phase variation  $\Phi$  and the cold plasma dispersion relation.

[32] If the energetic electron distribution function,  $F(v_{\parallel}, v_{\perp}, \zeta)$ , is known, then the components of the resonant current are straightforward to calculate,

$$J_B = -e \int_{v_{\parallel 1}}^{v_{\parallel 2}} \int_0^{\infty} \int_0^{2\pi} v_{\perp}^2 F(v_{\parallel}, v_{\perp}, \zeta) \cos \zeta dv_{\parallel} dv_{\perp} d\zeta \quad (8)$$

$$J_E = e \int_{v_{\parallel 1}}^{v_{\parallel 2}} \int_0^{\infty} \int_0^{2\pi} v_{\perp}^2 F(v_{\parallel}, v_{\perp}, \zeta) \sin \zeta dv_{\parallel} dv_{\perp} d\zeta \quad (9)$$

where the interval  $[v_{\parallel 1}, v_{\parallel 2}]$  is the range of parallel velocities over which the resonant interaction with the wave significantly perturbs the distribution, and  $\zeta$  is the angle between the perpendicular velocity vector and the wave magnetic field vector.

[33] The key component of any simulation of VLF-triggered emissions lies in the manner by which the resonant currents are calculated, and here our approach differs from the technique used in the Vlasov Hybrid code. To calculate the resonant current, the Vlasov Hybrid code randomly samples the initial distribution function using a large number of "simulation particles." Then, the change in energy  $\delta W$  for each particle is tracked as the simulation progresses. At each time step, the value of  $\delta W$  as a function of position on the phase space grid is found by interpolating from the values of  $\delta W$  for the simulation particles in the phase space hypercube surrounding each grid point. The resonant current is then calculated by integrating over the values of  $\delta W$  that have been laid down on the phase space grid [Nunn, 1993].

[34] In contrast, the model developed here determines the resonant current directly from the evolution of the distribution function as a function of time and phase space under the influence of the whistler mode wave. Specifically, the distribution function at each time  $t + \Delta t$  is determined from the known distribution function at time  $t$  through the application of Liouville's theorem. That is,

$$F(t + \Delta t, v_{\parallel}, v_{\perp}, \zeta, z) = F(t, v_{\parallel} - \delta v_{\parallel}, v_{\perp} - \delta v_{\perp}, \zeta - \delta \zeta, z - \delta z) \quad (10)$$

where the parameters  $\delta z$ ,  $\delta v_{\parallel}$ ,  $\delta v_{\perp}$ , and  $\delta \zeta$  are the total changes in  $z$ ,  $v_{\parallel}$ ,  $v_{\perp}$ , and  $\zeta$ , respectively that were seen between  $t$  and  $t + \Delta t$  by the particle that is at the phase space position  $(v_{\parallel}, v_{\perp}, \zeta, z)$  at time  $t + \Delta t$  between  $t$  and  $t + \Delta t$ . Using the particle scattering equations [e.g., *Dysthe*, 1971], to first order

$$\delta z = -v_{\parallel} \Delta t \quad (11)$$

$$\delta v_{\parallel} = -\left(\frac{e}{m_e} v_{\perp} B_w \sin \zeta + \frac{v_{\perp}^2}{2\omega_c} \partial \omega_c\right) \Delta t \quad (12)$$

$$\delta v_{\perp} = \left(\frac{e}{m_e} (v_{p\parallel} + v_{\parallel}) B_w \sin \zeta + \frac{v_{\parallel} v_{\perp}}{2\omega_c} \frac{\partial \omega_c}{\partial z}\right) \Delta t \quad (13)$$

$$\delta \zeta = \left(\omega_c - k_{\parallel} (v_{p\parallel} + v_{\parallel}) + \frac{e}{m_e} \left(\frac{v_{p\parallel} + v_{\parallel}}{v_{\perp}}\right) B_w \cos \zeta\right) \Delta t \quad (14)$$

where  $v_{p\parallel}$  is the wave phase velocity parallel to the magnetic field, and  $\omega_c = 2\pi f_c$  is the angular electron gyrofrequency.

[35] Finally in the development of our model, we note that the wave parameters ( $v_{p\parallel}$ ,  $k_{\parallel}$ ) in equations (11)–(14) vary as a function of position along the field line. Furthermore, they must also vary as a function of time to produce changing frequency emissions. Thus, a method of tracking the evolution of these parameters during the simulation is required. If we again assume that the wave is monochromatic, only a single frequency component exists at each position  $z$  and the wave number  $k_{\parallel}$  describes the rate of variation of the total wave phase, i.e.,

$$k_{\parallel} = \frac{\partial}{\partial z}(\phi + \Phi) = \frac{\partial \phi}{\partial z} + k_0 \quad (15)$$

where we have used the definition of  $\Phi$  from equation (5). From equation (15), we can see that the additional variation in the wave frequency content relative to the initial value  $k_0$  is captured in the variation of  $\phi$ . Additionally, we can calculate  $v_{p\parallel}$  using  $k_{\parallel}$  and the cold plasma dispersion relation,

$$v_{p\parallel} = \frac{k_{\parallel} \omega_c}{k_{\parallel}^2 + \frac{\omega_c^2}{c^2}} \quad (16)$$

[36] We now have an algorithm for modeling of the VLF wave growth and triggering process. We define an initial distribution function  $F_{i,j,k,l}^0$  where the superscript 0 denotes that we are representing the value of distribution function at time  $t = 0$ , and the subscripts  $i, j, k, l$  denote that we are representing the value of the distribution function at the position in phase space ( $v_{\parallel} + i\Delta v_{\parallel}, j\Delta v_{\perp}, k\Delta\zeta, z_{\min} + l\Delta z$ ), where  $z_{\min}$  is the location of the wave input to the modeled space, and  $\Delta v_{\parallel}, \Delta v_{\perp}, \Delta\zeta$ , and  $\Delta z$  are the resolution parameters of the system. The wave field values  $B_w^0$  and  $\phi_l^0$  are initialized to zero,  $k_{\parallel l}^0$  is initialized to the wave number corresponding to the initial wave frequency (i.e.,  $k_0$ ), and the rapid phase variation  $\Phi$  is initialized as

$$\Phi_l^0 = \begin{cases} 0 & l = 0, \\ -\sum_0^{l-1} k_{\parallel l}^0 \Delta z & \text{otherwise.} \end{cases}$$

From here, the process becomes iterative:

[37] 1. For each ( $i, j, k, l$ ), determine the parameters  $\delta z$ ,  $\delta v_{\parallel}$ ,  $\delta v_{\perp}$ , and  $\delta\zeta$  using equations (11)–(14). Determine  $F_{i,j,k,l}^{n+1}$  by interpolating  $F^n$  to find the value corresponding to  $F^n(v_{\parallel} + i\Delta v_{\parallel} - \delta v_{\parallel}, j\Delta v_{\perp} - \delta v_{\perp}, k\Delta\zeta - \delta\zeta, z_{\min} + l\Delta z - \delta z)$ .

[38] 2. Calculate the resonant current using

$$J_{BI} = -e \sum_i \sum_j \sum_k v_{\perp}^2 F_{i,j,k,l}^{n+1} \cos \zeta \Delta v_{\parallel} \Delta v_{\perp} \Delta \zeta$$

$$J_{EI} = e \sum_i \sum_j \sum_k v_{\perp}^2 F_{i,j,k,l}^{n+1} \sin \zeta \Delta v_{\parallel} \Delta v_{\perp} \Delta \zeta$$

[39] 3. Update the field parameters  $B_w^n$  and  $\phi^n$  according to the wave update equations (6) and (7), respectively. Update the rapid variation of the wave phase according to

$$\Phi_l^{n+1} = \Phi_l^n + \omega_0 \Delta t$$

[40] 4. Calculate  $k_{\parallel l}^{n+1}$  and  $v_{p\parallel l}^{n+1}$  using equations (15) and (16), respectively.

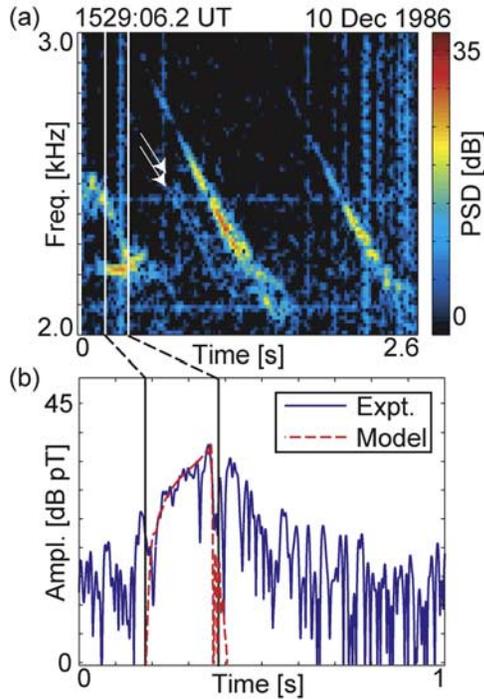
[41] 5. Iterate the above steps until the simulation is complete.

[42] We note that the growth term on the right-hand side of equation (7) is proportional to the inverse of the wave amplitude. Since the wave amplitude can become arbitrarily small (for example,  $B_w$  is initialized to zero), this term can advance the accumulated additional phase,  $\phi$ , at an equivalently high rate, producing frequency components that violate Nyquist considerations and affect system stability. Thus, at each time step, we limit the value of this term so that  $\phi$  cannot advance more than  $\pi$  radians in a single step. This limiting is equivalent to using an anti-aliasing filter on the frequency components generated in the course of the emission triggering process. Anti-aliasing is a natural consequence of the use of a discrete time step.

## 4. Results

[43] Before we apply our model to the analysis of the saturation process, we first execute several qualification runs to demonstrate that the model produces the expected behavior. We start by reproducing an experimental event, which is shown in Figure 7. We model a 200 ms constant-frequency pulse at 2.22 kHz which was transmitted at 1529:02.6 UT on 10 December 1986, shown in the bottom left of the spectrogram in Figure 7a. The narrowband amplitude in a 160-Hz bandwidth around the transmitted frequency is shown as the solid line in Figure 7b. This event is selected because the pulse is followed 200 ms later by two frequency ramps, sweeping at  $-1$  kHz/s from 2.96 kHz to 1.96 kHz. The group delay characteristics of the ramps allow us to determine the path of signal propagation ( $L \sim 4.69$ ) and equatorial cold electron density ( $\sim 94.1 \text{ cm}^{-3}$ ) using whistler diagnostics [Park, 1972; Sazhin *et al.*, 1992]. From the dispersion analysis, we find that the whistler mode delay for the 2.2 kHz pulse is 3.77 s, which corresponds to an arrival time of 0.17 s into the narrowband amplitude plot. The 200 ms pulse duration is marked by the vertical lines. The spectrogram and the narrowband plot both show wave energy arriving prior to this point as well as after the termination of the pulse. The early arrival is attributed to amplification on two weak ( $\sim 20$  dB less intense) paths with significantly shorter delays, highlighted by arrows in Figure 7a. Once the pulse arrives on the strong path, it dominates the received signal. Further extension of the wave energy beyond the 200 ms duration is attributed to a falling frequency emission triggered by a 200 ms pulse at 2.46 kHz transmitted immediately before the 2.22 kHz pulse.

[44] With the path properties, wave frequency, and pulse length determined, the only adjustable parameters left to choice are the input wave amplitude and ambient (i.e., initial) the hot electron distribution. Although ducted Siple signals were never observed in situ, several spacecraft observations of non-ducted obliquely propagating Siple transmissions have been observed which measured the wave magnetic field amplitude to be 0.02–0.05 pT [Kimura *et al.*, 1983; Rastani *et al.*, 1985; Sonwalkar and Inan, 1986]. For the cases considered here, where long-period saturation oscillations are active without significant triggering of free running plasma emissions, Siple transmitted at frequencies



**Figure 7.** Triggered emission event used to qualify the saturation model. (a) Time–frequency spectrogram of the event. The modeled pulse is at 2220 Hz and appears  $\sim 0.2$  s into the record. The 200 ms duration of the pulse is marked in the spectrogram using vertical lines. Dispersion analysis of the two following ramps shows a propagation path of  $L = 4.69$  and equatorial cold electron density of  $94.1 \text{ el/cm}^3$ . White arrows indicate two weak paths over which Siple transmissions also propagate. (b) Narrowband amplitude of the event in a 80 Hz bandwidth centered at 2220 Hz. The dashed line indicates the output of the model.

not exactly at the frequency to which the transmitter was tuned, resulting in lower radiated wave amplitudes. For example, during the event shown in Figure 1, the transmitter was tuned to 2460 Hz, and the measured power delivered to the antenna at 2700 Hz was 41% of the power delivered at 2460 Hz. In contrast, the cases where Siple transmissions were received on spaceborne receivers corresponded to transmissions that included the resonant frequency of the tuned transmitter. Therefore, we expect the input wave amplitudes to be reduced relative to those measured on spacecraft, and for modeling purposes we consider input wave amplitudes  $\leq 0.01$  pT.

[45] For the functional form of the initial hot electron distribution, we choose the distribution function from *Bell et al.* [2000],

$$F_{eq} \propto \frac{\sin^{25} \alpha}{v^4}$$

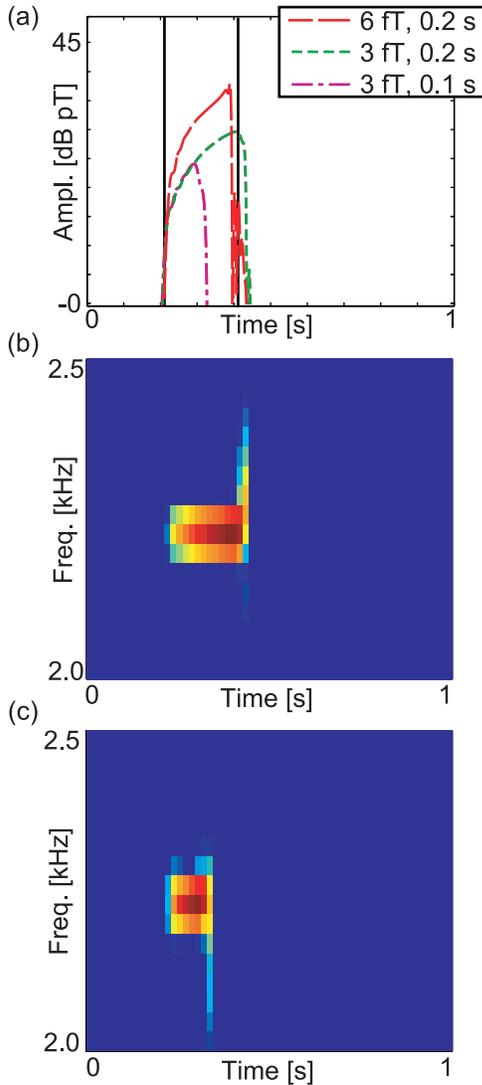
where  $F_{eq}$  is the equatorial value of the distribution function,  $\alpha$  is the electron pitch angle, and  $v$  is the electron velocity, which is a functional fit to energetic electron flux measurements aboard the POLAR spacecraft for energies less than 20 keV. To give the desired growth characteristics,

the total hot electron flux is reduced or enhanced as necessary. Consistent with *Nunn et al.* [2005], we find that the simulation results are not particularly sensitive to the choice of hot electron distribution, provided that it is sufficiently unstable (i.e., with high anisotropy and/or resonant electron flux) to amplify the small amplitude input waves to a level where the wave could strongly interact with the resonant particles within  $\sim 2000$  km of the magnetic equator. This amplified level is typically very close to the level required for the trapping of electrons within the wave potential, and requires a strongly unstable distribution corresponding to equatorial linear growth rates  $\geq 150$  dB/s.

[46] The various parameters controlling the resolution of the simulation were  $\Delta t = 0.5$  ms,  $\Delta z = 21147$  m,  $\Delta v_{\parallel} = 50000$  m/s,  $\Delta v_{\perp} = 1 \times 10^7$  m/s,  $\Delta \zeta = \pi/12$ . The parameters  $\Delta t$ ,  $\Delta v_{\parallel}$ ,  $\Delta v_{\perp}$ , and  $\Delta \zeta$  were selected by systematically halving their step sizes until the simulation result did not change between iterations. The spatial step  $\Delta z$  was selected to satisfy the Courant condition for the most limiting of  $v_{\parallel}$ ,  $v_g$ , and  $v_{p\parallel}$ , given  $\Delta t$  [Courant et al., 1967]. The parameter  $v_{\parallel 1}$  was varied at each  $z$ -position so that the range of  $v_{\parallel}$  values was a  $9 \times 10^6$  m/s range centered on the resonant velocity at each  $z$  corresponding to the input wave frequency. This range was also found by performing the simulation with successively larger values until the simulation result did not change between iterations. The perpendicular velocity was modeled over the range  $1 - 15 \times 10^7$  m/s, corresponding to an energy range of [284 eV, 63.9 keV], which is expected to cover the entire  $v_{\perp}$  range of interacting particles. The relative phase  $\zeta$  of the interacting distribution was modeled between  $[0, 2\pi]$ . The spatial extent of the model was  $[-2000 \text{ km}, 2000 \text{ km}]$  along the field line, centered around the magnetic equator. Modeling larger regions did not significantly affect the simulation results.

[47] The simulation was run with the above parameters, an input wave amplitude of 0.006 pT, and a hot electron distribution function with a calculated equatorial linear growth rate of 322 dB/s, producing the amplitude trace shown as a dashed line in Figure 7b. The wave arrives at the output of the modeled region with an amplitude of 10 pT, subsequently growing exponentially to a maximum amplitude of 90 pT at a growth rate of 61 dB/s. The growth rate and  $\sim 29$  dB of total growth fall well within the range of parameters reported for the Siple experiment [Helliwell and Katsufurakis, 1974]. Like the experimental result, shown as a solid line corresponding to a narrowband amplitude measurement with a 80 Hz bandwidth centered at 2.22 kHz, the modeled pulse reaches saturation and begins to decay before its termination, demonstrating that the model can reproduce saturation effects.

[48] To further qualify the model, we reduce the input wave amplitude to 0.003 pT and repeat the simulation with 200 ms and 100 ms pulses, expecting that the amplified pulse will not reach saturation prior to termination. As shown in Figure 8a, such is indeed the case. Here, the initial amplitude of the pulses as well as the observed growth rates are reduced relative to the 0.006 pT pulse. Spectrograms produced from the wave magnetic field at the output of the model (i.e.,  $B_{w_{\max}} \exp\sqrt{-1}(\phi_{l_{\max}}^n + \Phi_{l_{\max}}^n)$ ) for the two additional runs are shown in Figures 8b and 8c. The spectrogram in Figure 8b shows that the termination of



**Figure 8.** Qualification runs of the model. (a) Amplitude traces for the three qualification runs. The first, with an input wave amplitude of 6 fT and pulse length of 0.2 s, is intended to model the event shown in Figure 7. The other two show termination triggering of VLF emissions. (b) The first, with a 0.2 second pulse, triggers a rising emission. (c) The second, with a 0.1 second pulse, triggers a falling emission.

the 200 ms pulse triggers a rapidly rising emission. For the 100 ms pulse in Figure 8c, we see that the nature of the emission is changed to a faller. This effect of longer pulses triggering risers and shorter pulses triggering fallers was observed by *Helliwell and Katsufurakis* [1974], and serves to increase our confidence in the modeled behavior.

[49] The characteristic of the rising or falling nature emissions being controlled by pulse length is still an active area of research into the VLF-triggered emission instability. Some simple analytical relationships for the behavior have been proposed [*Helliwell*, 1967; *Roux and Pellat*, 1978], and other codes can reproduce the phenomenon [*Nunn et al.*, 2005], although these efforts have not been successful in providing an explanation for it. In addition to providing

confidence in the ability for our code to reproduce results consistent with experiment, the above result implies that the code can also be applied to the study of this characteristic of the instability. However, the study of this phenomenon is beyond the scope of this paper and is left for future work.

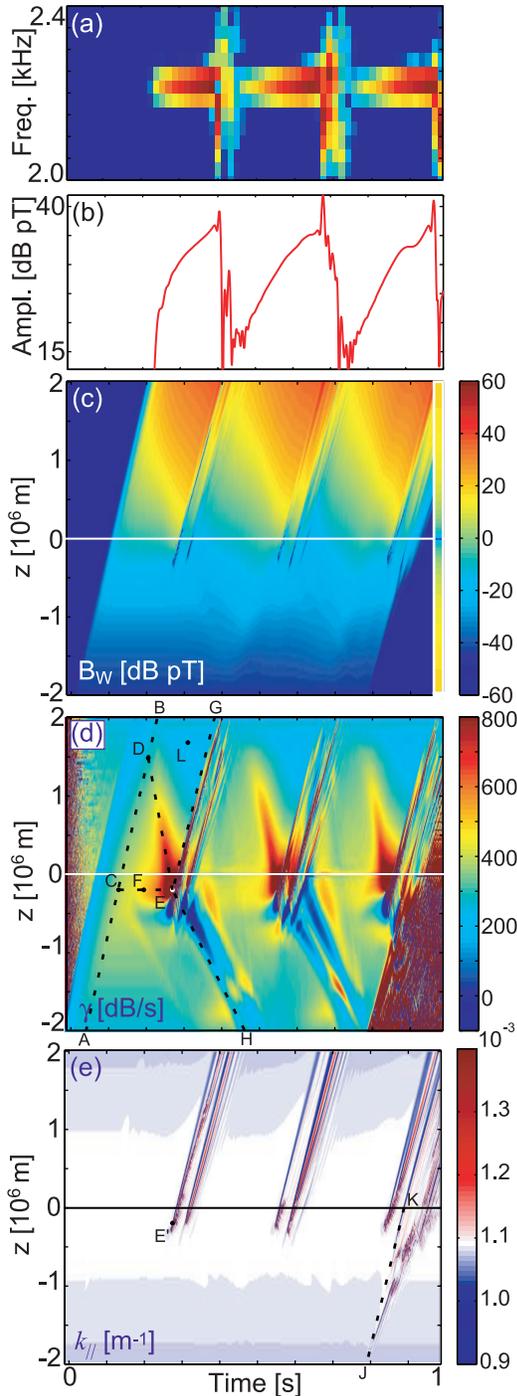
[50] Now studying the effects of saturation, we return the input wave amplitude to 0.006 pT and increase the pulse length to 800 ms. The results are shown in Figure 9. Figures 9a and 9b show the spectrogram and amplitude of the output wave, respectively, as a function of time. Figures 9c, 9d and 9e show the wave amplitude, growth rate, and wave number as a function of position along the field line  $z$  and time. Each column of pixels in the images is spaced 0.5 ms after the previous column. The last 0.25 s of Figure 9c represents the minimum wave amplitude required for particle trapping as a function of position. The equator is marked by a white horizontal line. All of the plots in Figure 9 share the same horizontal axis, with the simulation time running from 0 s to 1 s. The growth rate  $\gamma$  in Figure 9d is calculated from the resonant current  $J_E$  using [*Omura and Matsumoto*, 1982]

$$\gamma = \frac{1}{B_w} \frac{dB_w}{dt} = -\frac{\mu_0}{2} v_g \frac{J_E}{B_w}.$$

[51] From Figures 9a and 9b, we can see that the long-period oscillations, short-period oscillations, and wave frequency spread that we observed in section 2 are all reproduced by the model. We see three consecutive periods of growth followed by wave spreading and associated suppression of the wave below the initially received level. The short-period oscillations are most clearly developed during the second growth period. They start near the saturation level and continue through three oscillations as the wave amplitude drops to the suppressed level.

[52] The dynamics of the saturation process are captured in Figures 9c, 9d, and 9e. As the initial part of the pulse propagates toward the equator (along the dashed line segment  $\overline{AB}$  in Figure 9d), the growth rate smoothly varies from 200 to 320 dB/s, consistent with the growth rate predicted by linear theory. As the wave nears the equator, its amplitude ( $\sim 1$  pT) is sufficient to trap particles in a narrow region around the equator. Although the length of the trapping region is not enough to trap particles through several oscillations, the resonant interaction is strong enough to significantly perturb the resonant particle trajectories. Once the wave front passes the equator, the growth rate just south of the equator is enhanced owing to this resonant interaction effect (Point C). As the wave propagates northward from the equator, its amplitude falls below the trapping level as the trapping level increases owing to the effect of the inhomogeneity. The amplification of the wave continues in a linear manner such that the amplitude at the wave front crosses over the trapping level a second time at Point D, roughly 1500 km north of the equator. Once again, the growth rate south of the point where  $B_w \approx B_{tr}$  is increased; the resonant currents are enhanced as particles transition from a region where trapping can occur to one in which the magnetic inhomogeneity dominates.

[53] Once the wave energy has filled the simulation space, there are two regions of enhanced growth. The northern growth rate enhancement causes the wave ampli-



**Figure 9.** Output of the model showing saturation effects. The model is driven with a 6 fT pulse with pulse duration of 0.8 s. (a) Time frequency spectrogram of the model output, showing frequency spread at saturation and three long-period oscillations. (b) Amplitude trace of the model output, also showing short-period oscillations developing at saturation. (c) Wave amplitude  $B_w$  as a function of time and position within the modeled region. The magnetic equator is marked by a white line, and the equivalent color scale for the minimum trapping amplitude is shown at the right. (d) Measured growth rate  $\gamma$  as a function of time and position within the modeled region. (e) Measured wave number  $k_{\parallel}$  as a function of time and position within the modeled region.

tude just south of Point D to be increased above the local trapping level, which enhances the growth rate even further south, and so on, such that the enhanced growth rate is seen to move southward as time progresses, along the line segment  $\overline{DE}$ . In contrast, the enhanced growth rate just south of the equator remains stationary, since the trapping level increases south of the equator. As a result, the growth rate increases along the line segment  $\overline{CE}$ . When the two regions of enhanced growth rate merge, the growth south of the equator is strongly enhanced.

[54] As the maximum growth rate south of the equator (near  $z = -300$  km and Point F) increases above  $\sim 500$  dB/s, a depression in the growth rate just south of the maximum begins to appear. Further increasing growth causes the depression to become deeper, below  $-1300$  dB/s, at which point the wave is locally absorbed and saturation occurs (Point E). We see from the wave number plot, Figure 9e, that wave numbers corresponding to frequencies different from the transmitted frequency are developed coincident with saturation. At this point, the high-growth structure and off-frequency wave components propagate northward to the receiver, along line segment  $\overline{EG}$ , while the location of negative growth appears to stream southward along line segment  $\overline{EH}$ . This process then repeats for the other two long-period oscillations.

[55] It is clear by comparing Figures 9d and 9e that the generation of off-resonant wave frequencies is associated with the location of maximum wave growth; the off-resonant wave numbers appear immediately south of the location of maximum wave growth after saturation has occurred, Point E'. Furthermore, the termination of the pulse also generates off-resonant frequencies, even prior to the wave crossing the equator, as can be seen after 0.8 s in Figure 9e along line segment  $\overline{JK}$ . Thus, it appears that generation of additional wave frequencies is also associated with gradients in wave amplitude.

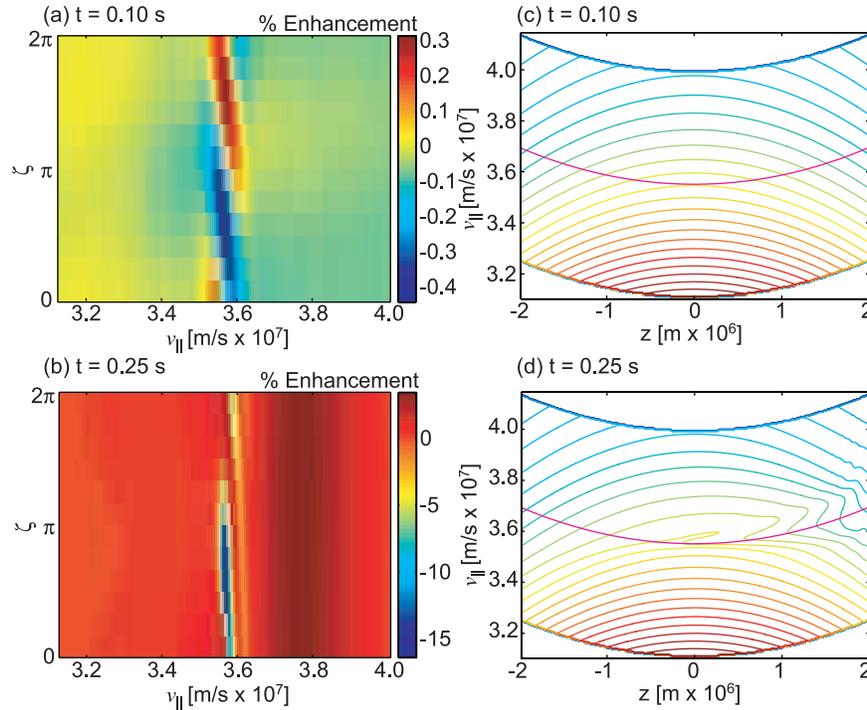
[56] The modification of the distribution function at two representative times is shown in Figure 10. Figures 10a and 10c represent the distribution function at  $t = 0.10$  s, when the leading edge of the pulse has just passed the equator. Figures 10b and 10d represent the distribution function at  $t = 0.25$  s, just prior to saturation, after a period significant nonlinear growth. Figures 10a and 10b represent the modification of the distribution function relative to its initial value plotted in the  $v_{\parallel} - \zeta$  plane, centered at the equatorial resonant velocity. That is, with

$$\delta F_{\text{eq}}^n = \sum_j F_{i,j,k,l=l_{\text{eq}}}^n$$

where  $l_{\text{eq}}$  is the spatial index at the equator, the plotted value is

$$\frac{\delta F_{\text{eq}}^n - \delta F_{\text{eq}}^0}{\delta F_{\text{eq}}^0}$$

Figures 10c and 10d show contours of constant phase space density averaged over  $v_{\perp}$  and  $\zeta$  in the  $v_{\parallel} - z$  plane. The values of  $v_{\parallel}$  are centered about the resonant velocity  $v_{\text{res}}$  for the initial wave frequency, shown as a curved line.



**Figure 10.** Modification of the energetic electron distribution. (a and b) Relative modification of the distribution function (integrated over  $v_{\perp}$ ) at the equator at times  $t = 0.10$  s and  $0.25$  s, respectively. (c and d) Contours of the distribution function (integrated over  $v_{\perp}$  and  $\zeta$ ) as a function of position along the magnetic field line at times  $t = 0.10$  s and  $0.25$  s, respectively.

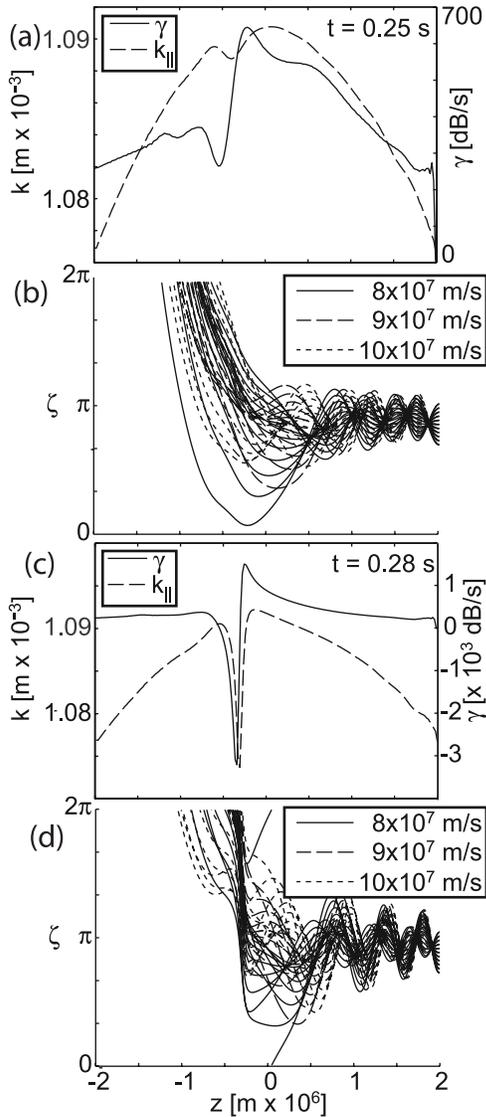
[57] At  $t = 0.10$  s (Figures 10a and 10c), the wave has filled slightly more than half of the modeled space, with the leading edge of the wave at  $z \sim 200$  km. Averaged over gyrophase, the distribution function in Figure 10c shows no appreciable change. In Figure 10a, the distribution function at the equator shows a bunching in phase similar to that predicted by linear theory, with a shape that is approximately symmetric about  $\zeta = \pi/2$ . From equation (8), it is easily shown that this configuration forms a resonant current with  $J_B \sim 0$ , giving little to no phase advance. Additionally, the magnitudes of enhancement and depression of the distribution function are roughly equal at  $\sim 3\%$ .

[58] At  $t = 0.25$  s (Figures 10b and 10d), just before saturation occurs, the wave has filled the entire modeled space. The distribution function in Figure 10d is significantly modified north of the equator owing to the presence of the large amplitude wave. The value of the distribution function is depressed near the resonant velocity and enhanced at parallel velocities just above the resonant value. From Figure 10b, we can see that this corresponds to a  $\sim 5\%$  enhancement just above the resonant velocity. However, this enhancement is roughly uniform in  $\zeta$  and contributes little to resonant current formation. Figure 10b is also significantly changed from Figure 10a. The distortion of the distribution function is no longer symmetric, displaying the presence of a phase space electron hole. The magnitudes of the enhancement and depression of the distribution function are significantly higher and no longer equal, with a  $\sim 15\%$  depression in the value of the distribution function associated with the phase space electron hole. Furthermore, the center of the hole is displaced from  $\zeta = \pi/2$  toward  $\zeta =$

$\pi$ , giving a significant resonant current component to drive the phase advance of the wave.

[59] This examination of the distribution function at  $t = 0.10$  s and  $t = 0.25$  s reveals that the growth process begins in a fashion consistent with linear theory and transitions to a condition with significant nonlinear effects, including particle trapping and the creation of a phase space electron hole. However, the examination of the whole distribution function gives little insight into the dynamics of the growth process. To understand these dynamics, we note that the fact that the wave grows as it propagates allows us to divide configuration space into two regions: one region where trapping cannot occur and a separate region where it can. In particular, a subset of initially-trapped particles streaming southward through the wave field will transition from the region where they are trapped to the region where they cannot be trapped. The effect of the transition from a trapped condition to an untrapped condition is examined in Figure 11, which shows test particle trajectories through the wave field developed at  $0.25$  s and  $0.28$  s into the simulation, respectively before and at saturation. Figure 11a shows the growth rate (solid line) and the wave number (dashed line) as a function of position. By  $0.25$  s, the growth rate just south of the equator has approximately doubled to  $\sim 650$  dB/s, and there is a sharp gradient in the growth rate around  $z = -400$  km. Between  $z = -400$  km and  $z = -600$  km, the growth rate is depressed below its initial level.

[60] Figure 11b shows test particle trajectories through the wave field at  $t = 0.25$  s. The initial  $v_{\parallel}$  and phase of the particles is chosen so that the particles are initially trapped in the high-amplitude portion of the wave. Three different



**Figure 11.** Evolution of the phase of trapped particle trajectories at two separate times during the simulation shown in Figure 9. (a and c) Wave number  $k_{\parallel}$  (solid line) and growth rate  $\gamma$  (dashed line) as a function of position at time  $t = 0.25$  s and  $0.28$  s, respectively. Both quantities show the effect of the detrapping of particles and the evolution of the phase space electron hole. (b and d) Particle trajectories for three perpendicular velocities. Particles are detrapped near the equator, leaving them with parallel velocities close to the equatorial resonant velocity. As a result, particle phase advances relative to the wave phase as particles stream south.

initial perpendicular velocities are shown, which sample the peak of the distribution function as a function of  $v_{\perp}$  at the resonant velocity. As these particles represent the densest part of the distribution function as well as the range of perpendicular velocities that are most deeply trapped by the wave potential well, they represent the most significant component of the distribution of the trapped particles, and are therefore representative of the effect of the particle trap on current formation and subsequent wave growth char-

acteristics. However, it should be noted that the untrapped, ambient electrons also contribute significantly to the formation of resonant currents. This, this discussion regards the effect of detrapping particles as an enhancement or depression of the current created by the resonant, untrapped particles, which is evidenced in the relative change of  $k$  and  $\gamma$  as the instability evolves.

[61] As the particles stream southward (right to left in the figure), their phase remains relatively close to their initial phase, as expected for trapped particles. After crossing the equator, they enter the region of rapidly changing wave amplitudes and the particles are detrapped. For some distance, between  $z = 500$  km and  $z = -500$  km, the particles remain relatively bunched in phase in a region centered between  $\zeta = \pi/2$  and  $\zeta = \pi$ , which causes the growth rate enhancement in this region. However, as the now-untrapped particles stream southward from the equator, the resonant velocity of the wave increases relative to the parallel velocities of the particles (which are decreasing owing to the effect of the inhomogeneity). This increase causes the relative phase  $\zeta$  of the particles to the wave phase to increase for every de-trapped particle. In turn, the phase space position of the electron hole swings from its formerly stable location, to some location  $\zeta > \pi$  before they begin to lose their bunched character. At phase space locations with  $\zeta > \pi$ , the resonant current  $J_E$  switches sign, causing a reduction in growth rates below their initial level.

[62] Initially, this transition of the resonant currents from growth to damping is slow, and the electron hole mixes with the ambient plasma before damping can occur. However, as the wave intensity and gradients increase, as shown in Figures 11c and 11d, which correspond to the wave field at  $t = 0.28$  s, the transition becomes more rapid. This enhanced pace of the transition is due to the influence of the resonant current  $J_B$  on the rate that the dissipating electron hole rotates through  $\zeta$ -space. Specifically, as the wave is amplified the size of the electron hole increases, which increases the magnitude of the resonant current. This increase leads to stronger deviations of  $\phi$  due to the effect of  $J_B$ , which drives the value of  $k_{\parallel}$  down (note that in the frame of the wave,  $\phi$  is seen to increase, as expected from (7)). This decreasing  $k_{\parallel}$  causes the resonant velocity of the wave to increase, causing the dissipating hole to rotate more rapidly in the wave reference frame. If the resonant currents are strong enough, then  $k_{\parallel}$  will be strongly reduced, and the strong de-trapping current transitions quickly from strong growth to damping.

[63] When the damping becomes strong enough, the wave is absorbed by the plasma, resulting in  $B_w \rightarrow 0$ . This reduction in wave intensity causes the term on the right-hand side of (7) to become large, causing rapid variations in  $\phi$  and a resulting spread in  $k_{\parallel}$ , which is observed at the output as the generation of wideband waves at saturation. Because of the rapid variations in  $\phi$ , the effect of  $J_B$  also varies rapidly, which is observed as rapid changes from damping to growth in the region of the detrapping currents, as seen in Figure 9d. The absorption of the wave also disrupts the amplification process and the whole spatial growth structure begins to propagate northward. As the region of rapid  $\phi$  variation and low  $B_w$  propagates away from the equator, the wave structure is reestablished and growth consequently resumes. This cycle of disruption and

resumption of growth is observed at the output as the short-period saturation oscillations.

[64] There are two additional features of the modeling results that deserve comment. First, in the region where the particles are strongly trapped by large wave amplitudes, the growth is depressed below the initial linear growth rate, as seen in Figure 9d, near Point L. This depression is a reflection of the fact that the energy from the non-trapped particle population is used to amplify the wave as well as accelerate the trapped electrons. The second feature worthy of note is the fact that a region of negative growth is generated at saturation which appears to propagate southward away from the equator, along line segment  $\overline{EH}$ . The velocity at which the negative growth region is approximately  $10^7$  m/s, which is significantly less than the equatorial resonance velocity of  $3.6 \times 10^7$  m/s, and therefore cannot be caused by a single group of depleted particles. This negative growth region is due to the destruction of anisotropy caused by the maximum amplitude portion of the wave packet. As the wave packet propagates northward from saturation, it leaves behind itself a depleted population that streams southward. The effect on the growth rate is not visible until the depleted population crosses over the resonant velocity south of the equator, which corresponds to positions in space and time along the line segment  $\overline{EH}$ .

[65] From the modeling results, we can make some generalizations about the mechanism of the instability. The nonlinear growth of the wave is not driven solely by the trapping of electrons. Indeed, in the region where trapping and the generation of the phase space electron hole are occurring, the growth rate of the wave is locally depressed. Instead, the enhancement of the growth rate occurs in regions where the wave grows from the condition of no trapping ( $B_w < B_{tr}$ ) to one in which trapping occurs ( $B_w > B_{tr}$ ). Before reaching this region, trapped electrons experience a deep wave potential and a separatrix that prevents them from mixing with the ambient electrons, forming the phase space electron hole. Upon reaching this region, the separatrix disappears, and the electron hole begins to mix with the ambient electrons. For some length, the electrons remain in resonance with the wave, enhancing the linear currents produced by the ambient electrons. If the wave amplitude gradients in the growth region are modest, and the transition from trapped to untrapped is smooth, resulting in an equivalently modest enhancement of growth. As the nonlinear growth continues and the amplitude gradients become sharper, the loss of the trap becomes much more abrupt, and the de-trapping electron hole stays more tightly bunched for a longer distance, enhancing resonant currents. However, if the hole stays bunched long enough that it starts to lose its resonance with the wave (owing to the effect of the inhomogeneity) before it dissipates, the resonant currents will start to affect the wave's phase in a manner to drive the loss of resonance faster, eventually causing saturation, as discussed above.

[66] *Dowden et al.* [1978], *Omura and Summers* [2006], and *Omura et al.* [2008] have previously identified that the phase space electron hole created by particle trapping in the presence of the magnetic inhomogeneity is responsible for the formation of resonant currents that cause the nonlinear wave growth associated with the instability. In agreement with these results, we have shown that the electron hole

indeed contributes to the formation of resonant currents. Additionally however, we have shown in the region where the electron hole exists, growth rates are reduced relative to the linear growth that occurs before the wave action is strong enough to trap electron. Furthermore, it is only the presence of wave amplitude gradients that allows the electron hole to dissipate and contribute to the strong enhancement of the wave growth associated with the nonlinear aspects of the instability.

## 5. Discussion and Conclusions

[67] We started by exploring the observational records of VLF-triggered emissions and cataloging phenomena that appear to be associated with the saturation of the exponential growth of the received wave. The phenomena that we found to be associated with saturation were long-period ( $\sim 1$  s) and short-period ( $\sim 20$ – $30$  ms) oscillation and a generation of incoherent wave energy extending in a significant ( $>100$  Hz) band around the transmitted frequency. We also characterized the oscillations in amplitude and frequency, noting that oscillations due to saturation effects could be the cause of the often reported sideband phenomenon [*Park and Chang*, 1978; *Park*, 1981; *Helliwell*, 1983, 1988; *Helliwell et al.*, 1986; *Sa and Helliwell*, 1988].

[68] It is worth noting that there is little distinction in the literature between two distinct forms of sidebands produced by whistler mode waves, and that some attempts are made to explain both with a single mechanism. The first kind of sideband is that which is generated at saturation, and the second is the generation of sidebands during the interaction between two waves closely spaced in frequency. The most notable difference between the two kinds is that sidebands produced at saturation occur after some period of growth, while sidebands produced during multi-wave interactions can occur immediately upon signal reception, without (yet) any notable growth. Generally, the sidebands produced at saturation appear sequentially, with the lowest-order sidebands appearing first and strongest [cf. *Park*, 1981], while the sidebands due to multiple waves show no such order [cf. *Park and Chang*, 1978].

[69] In addition to the observations of the VLF-triggered emission phenomenon at saturation, we have a model that can reproduce these observed long-period oscillations, short-period oscillations, and frequency spreading, as shown in Figures 9a and 9b. The model is also able to produce temporal amplification of the input signal that naturally leads to saturation. In particular, all of these phenomena can be explained from the modeling results by the formation of resonant currents and the influence of a detrapping particle population on the current formation process.

[70] Specifically, our model for the VLF-triggered emission phenomenon is based on resonant particle currents that are calculated directly from a modeled hot electron distribution function. The model is most similar to the Vlasov Hybrid Simulation, but does not require the use of an additional (artificial) saturation factor or repeated filtering of the wave field. The primary difference between our model and the Vlasov Hybrid Simulation [e.g., *Nunn*, 1990] is the way in which we track the evolution of the distribution function. The Vlasov Hybrid code randomly

seeds the modeled space with simulation particles and tracks the evolution of the individual particles, always referring back to the original distribution. In contrast, we seed the space at predetermined, regular intervals, referring back to the distribution from the previous time step. This method allows us to interpolate over a regular grid, greatly reducing the computational noise relative to interpolating from the irregularly spaced simulation particles.

[71] Nunn [1993] specifies a “fundamental density law” which ensures that any given grid point has at least one simulation particle available for use in interpolation (i.e., in the  $2^n$  hypercubes around the point, where  $n$  is the dimensionality of the simulation) with greater than 99.5% probability. However, with 1M grid points, the density law does not ensure coverage for some 5000 points, requiring additional interpolation over the grid to resolve these values [Nunn, 2005]. Furthermore, there are undoubtedly a number of points that rely on 1–2 particles to generate a value during the interpolation. In comparison, our scheme always has  $2^n$  values with which interpolation can be performed.

[72] The advantage that the Vlasov Hybrid has over our approach is that if the distribution has a particularly fine structure or if the dynamics of the triggered emission process are chaotic, the random seeding of the simulation particles provides the opportunity to sample that fine structure or chaotic interaction. In our case, the use of a regular phase space grid implies Nyquist limitations on the ability to sample any phase space structure, and any chaotic interactions would not be well explored.

[73] Additionally, our code has difficulty triggering the long-enduring, free-running plasma emissions which characterize both the experimental data and the Vlasov Hybrid results. Unlike the Vlasov Hybrid technique, we have no re-centering step in which the range of modeled parallel velocities is adjusted according to the distribution of  $k_{\parallel}$  values within the simulation space. This re-centering could serve to drive the frequency sweep of the simulation, but is necessary to maintain consistency with the assumption of monochromaticity inherent in the derivation of equations (6) and (7). Additionally, at saturation the spatial phase structure of the wave becomes very fine, and our choice of  $\Delta z$  may not be adequate to model the full frequency evolution of the wave, stunting the growth of off-frequency components. In the future, it would be appropriate to develop an adaptive mesh technique to adequately represent this post-saturation fine wave structure.

[74] One limitation that both codes share is the assumption of the monochromatic nature of the wave. At any point in the wave space, there is only one wave number defined. Thus, the model cannot accurately represent multi-wave interactions, such as entrainment, multi-wave sideband interactions, the coherence bandwidth effect, or the tendency to trigger only risers on long pulses. To treat such problems would require another derivation of the wave field equations and may be better treated in the frequency domain, which is left for future work.

[75] **Acknowledgments.** We would like to acknowledge D. Musetescu for his work in the reduction and archiving of the digitized Siple data. Further thanks are also due to D. L. Carpenter for his insightful comments on this work and M. A. Golkowski and A. M. Gibby for many fruitful discussions regarding the theoretical and modeling aspects of this work.

This work was supported by the Office of Naval Research via contract N00014-05-C-0525.

[76] Amitava Bhattacharjee thanks Anatoly V. Streltsov and another reviewer for their assistance in evaluating this paper.

## References

- Bell, T. F. (1985), High-amplitude VLF transmitter signals and associated sidebands observed near the magnetic equatorial plane on the ISEE 1 satellite, *J. Geophys. Res.*, *90*, 2792.
- Bell, T. F., and R. A. Helliwell (1971), Pulsation phenomena observed in long-duration VLF whistler mode signals, *J. Geophys. Res.*, *76*, 8414.
- Bell, T. F., and U. S. Inan (1981), Transient nonlinear pitch angle scattering of energetic electrons by coherent VLF wave packets in the magnetosphere, *J. Geophys. Res.*, *86*, 9047.
- Bell, T. F., U. S. Inan, R. A. Helliwell, and J. D. Scudder (2000), Simultaneous triggered VLF emissions and energetic electron distributions observed on POLAR with PWI and HYDRA, *Geophys. Res. Lett.*, *27*, 165.
- Brinca, A. L. (1972), Whistler side-band growth due to nonlinear wave-particle interaction, *J. Geophys. Res.*, *77*, 3508.
- Chang, D. C. D., and R. A. Helliwell (1979), Emission triggering in the magnetosphere by controlled interruption of coherent VLF signals, *J. Geophys. Res.*, *84*, 7170.
- Cornilleau-Wehrin, N., and R. Gendrin (1979), VLF transmitter-induced quiet bands: A quantitative interpretation, *J. Geophys. Res.*, *84*, 882.
- Courant, R., K. O. Friedrichs, and H. Lewy (1967), On the partial differential equations of mathematical physics, *IBM J. Res. Dev.*, *11*, 215.
- Das, A. C. (1968), A mechanism for VLF emissions, *J. Geophys. Res.*, *73*, 7457.
- Dowden, R. L., A. D. McKay, L. E. S. Amon, H. C. Koons, and M. H. Dazey (1978), Linear and nonlinear amplification in the magnetosphere during a 6.6-kHz transmission, *J. Geophys. Res.*, *83*, 169.
- Dysthe, K. B. (1971), Some studies of triggered whistler emissions, *J. Geophys. Res.*, *76*, 6915.
- Helliwell, R. A. (1967), A theory of discrete VLF emissions from the magnetosphere, *J. Geophys. Res.*, *72*, 4773.
- Helliwell, R. A. (1983), Controlled stimulation of VLF emissions from Siple Station, Antarctica, *Radio Sci.*, *18*, 801.
- Helliwell, R. A. (1988), VLF wave stimulation experiments in the magnetosphere from Siple Station, Antarctica, *Rev. Geophys.*, *26*, 551.
- Helliwell, R. A., and J. P. Katsufraakis (1974), VLF wave injection into the magnetosphere from Siple Station, Antarctica, *J. Geophys. Res.*, *79*, 2511.
- Helliwell, R. A., D. L. Carpenter, and T. R. Miller (1980), Power threshold for growth of coherent VLF signals in the magnetosphere, *J. Geophys. Res.*, *85*, 3360.
- Helliwell, R. A., U. S. Inan, J. P. Katsufraakis, and D. L. Carpenter (1986), Beat excitation of whistler mode sidebands using the Siple VLF transmitter, *J. Geophys. Res.*, *91*, 143.
- Kennel, C. F., and H. E. Petschek (1966), Limit on stably trapped particle fluxes, *J. Geophys. Res.*, *71*, 1.
- Katoh, Y., and Y. Omura (2006), A study of generation mechanism of VLF triggered emission by self-consistent particle code, *J. Geophys. Res.*, *111*, A12207, doi:10.1029/2006JA011704.
- Kimura, I., H. Matsumoto, T. Mukai, K. Hashimoto, T. F. Bell, U. S. Inan, R. A. Helliwell, and J. P. Katsufraakis (1983), EXOS-B/Siple Station VLF wave-particle interaction experiments: 1. General description and wave-particle correlations, *J. Geophys. Res.*, *88*, 282.
- Lampe, M., G. Joyce, W. M. Manheimer, A. Steltsov, and G. Ganguli (2006), Quasineutral particle simulation technique for whistlers, *J. Comput. Phys.*, *214*, 284.
- Liemohn, H. B. (1967), Cyclotron-resonance amplification of VLF and ULF whistlers, *J. Geophys. Res.*, *72*, 39.
- Matsumoto, H., and I. Kimura (1971), Linear and nonlinear instabilities and VLF emissions in the magnetosphere, *Planet. Space Sci.*, *19*, 567.
- Matsumoto, H., K. Hashimoto, and I. Kimura (1980), Dependence of coherent nonlinear whistler interaction on wave amplitude, *J. Geophys. Res.*, *85*, 644.
- Nunn, D. (1974), A self-consistent theory of triggered VLF emissions, *Planet. Space Sci.*, *22*, 349.
- Nunn, D. (1990), The numerical simulation of VLF nonlinear wave-particle interactions in collision-free plasmas using the Vlasov hybrid simulation technique, *Comput. Phys. Commun.*, *60*, 1.
- Nunn, D. (1993), A novel technique for the numerical simulation of hot collision-free plasma: Vlasov Hybrid Simulation, *J. Comput. Phys.*, *108*, 180.
- Nunn, D. (2005), Vlasov Hybrid Simulation—An efficient and stable algorithm for the numerical simulation of collision-free plasma, *Transp. Theory Stat. Phys.*, *34*, 151.
- Nunn, D., Y. Omura, H. Matsumoto, I. Nagano, and S. Yagitani (1997), The numerical simulation of VLF chorus and discrete emissions observed on the Geotail satellite using a Vlasov code, *J. Geophys. Res.*, *102*, 27,083.

- Nunn, D., A. Demekhov, V. Trakhtengerts, and M. J. Rycroft (2003), VLF emission triggering by a highly anisotropic energetic electron plasma, *Ann. Geophys.*, *21*, 481.
- Nunn, D., M. Rycroft, and V. Trakhtengerts (2005), A parametric study of the numerical simulations of VLF triggered emissions, *Ann. Geophys.*, *23*, 1.
- Omura, Y., and H. Matsumoto (1982), Computer simulations of pasic processes of coherent whistler mode wave-particle interactions in the magnetosphere, *J. Geophys. Res.*, *87*, 4435.
- Omura, Y., and D. Summers (2004), Computer simulations of relativistic whistler-mode wave-particle interactions, *Phys. Plasmas*, *11*, 3530.
- Omura, Y., and D. Summers (2006), Dynamics of high-energy electrons interacting with whistler mode chorus emissions in the magnetosphere, *J. Geophys. Res.*, *111*, A09222, doi:10.1029/2006JA011600.
- Omura, Y., D. Nunn, H. Matsumoto, and M. J. Rycroft (1991), A review of observational, theoretical and numerical studies of VLF triggered emissions, *J. Atmos. Terr. Phys.*, *53*, 351.
- Omura, Y., Y. Katoh, and D. Summers (2008), Theory and simulation of the generation of whistler-mode chorus, *J. Geophys. Res.*, *113*, A04223, doi:10.1029/2007JA012622.
- Park, C. G. (1972), Methods of determining electron concentrations in the magnetosphere from nose whistlers, *Tech. Rep. 3454-1*, Stanford Electron. Lab., Stanford, Calif.
- Park, C. G. (1981), Generation of whistler-mode sidebands in the magnetosphere, *J. Geophys. Res.*, *86*, 2286.
- Park, C. G., and D. C. D. Chang (1978), Transmitter simulation of power line radiation effects in the magnetosphere, *Geophys. Res. Lett.*, *5*, 861.
- Raghuram, R., T. F. Bell, R. A. Helliwell, and J. P. Katsufakis (1977), Echo-induced suppression of VLF transmitter signals in the magnetosphere, *J. Geophys. Res.*, *82*, 2787.
- Rastani, K., U. S. Inan, and R. A. Helliwell (1985), DE 1 observations of Siple transmitter signals and associated sidebands, *J. Geophys. Res.*, *90*, 4128.
- Roux, A., and R. Pellat (1978), A theory of triggered emissions, *J. Geophys. Res.*, *83*, 1433.
- Sa, L. A. D. (1990), A wave-particle-wave interaction mechanism as a cause of VLF triggered emissions, *J. Geophys. Res.*, *95*, 12,277.
- Sa, L. A. D., and R. A. Helliwell (1988), Structure of VLF whistler mode sideband waves in the magnetosphere, *J. Geophys. Res.*, *93*, 1987.
- Sazhin, S. S., M. Hayakawa, and K. Bullough (1992), Whistler diagnostics of magnetospheric parameters: A review, *Ann. Geophys.*, *10*, 293.
- Sonwalkar, V. S., and U. S. Inan (1986), Measurements of Siple transmitter signals on the DE 1 satellite: Wave normal direction and antenna effective length, *J. Geophys. Res.*, *91*, 154.
- Smith, A. J., and D. Nunn (1998), Numerical simulation of VLF risers, fallers, and hooks observed in Antarctica, *J. Geophys. Res.*, *103*, 6771.
- Stiles, G. S., and R. A. Helliwell (1977), Stimulated growth of coherent VLF waves in the magnetosphere, *J. Geophys. Res.*, *82*, 523.

---

T. F. Bell, A. R. Gibby, and U. S. Inan, Space, Telecommunications, and Radioscience Laboratory, Stanford University, 308 Packard Building, 350 Serra Mall, Stanford, CA 94305, USA. (argibby@stanford.edu)