SATURATION EFFECTS IN VLF TRIGGERED EMISSIONS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF
ELECTRICAL ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Andrew Ryan Gibby
May 2008

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____

(Umran S. Inan)   Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____

(Timothy F. Bell)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____

(Roger T. Howe)

Approved for the University Committee on Graduate Studies.

.

This dissertation is dedicated to my parents,
who have shown me what dedication truly is.

Abstract

The phenomenon of artificially triggered VLF emissions was first reported in 1964. Caused by a plasma instability in the Earth's radiations belts, VLF triggered emissions occur when an externally imposed VLF wave, propagating through the magnetosphere in the whistler mode, engages in a cyclotron resonance interaction with energetic electrons trapped in the Earth's magnetic field. Salient characteristics of VLF triggered emissions include exponential temporal growth of the triggering wave and generation of free-running plasma emissions, whose frequency can differ significantly from that of the input signal. The explanation of this phenomenon has proven to be analytically intractable, primarily due to its nonlinear nature and the complicating inhomogeneity of the Earth's magnetic field.

This study focuses specifically on the saturation characteristic of the phenomenon, that is, the phase of the instability associated with the termination of exponential growth. From the analysis of data from the Siple Station, Antarctica, VLF wave injection experiment recorded in 1986, three characteristics of the instability have been related to saturation:

1. Long-period oscillations, characterized by a cycle of exponential growth of the input wave to saturation, followed by suppression of the input wave, then followed by renewed exponential growth,

2. Short-period oscillations, previously associated with sidebands, and

3. Generation of incoherent off-frequency wave energy.

Exploration of this instability requires the use of a numerical simulation. To that end, a simple model is developed from first-principle considerations and applied to the

problem. Given realistic inputs, the model reproduces the observed characteristics of the instability during the growth and saturation phases. An analysis of the modeling results reveals that the nonlinear effects of the instability are driven by wave amplitude gradients, specifically those gradients that reflect a transition from a condition where the wave can trap electrons in its potential well in the presence of the magnetic inhomogeneity to a condition where such trapping cannot occur.

Acknowledgements

It is very rare that scientific work is done in a vacuum, and this work is no exception to that rule. Thus, I would like to take this opportunity to thank those individuals who made this effort possible.

First, I wish to express sincere thanks to my principal adviser Professor Umran Inan for allowing me complete freedom in the direction of my research. There were many times that I became painfully aware that I had more than enough rope to hang myself, but Umran's enthusiasm, persistence, and encouragement have turned what could have been a mediocre thesis into something of which I am proud.

Next, I would like to thank the members of my defense committee: Dr. Timothy Bell, Professor Roger Howe and Professor Piero Pianetta. I am grateful to Dr. Bell for his insight into the problem at hand and into the progress of science. To the third reader of my dissertation, Professor Roger Howe, I thank for entertaining the notion of reading a STAR lab thesis, as well as for his own insights into my future. I am also thankful to Professor Piero Pianetta for chairing my oral defense session.

Professor Don Carpenter has been extremely helpful in all of my work, providing historical perspective and anecdote that breathed life into a 35-year-old experiment. Don is an example of what every young scientist should aspire to be, displaying the value of meticulous hard work and the careful characterization of success.

My success at Stanford would not have been possible without the experience that I gained during my service in the United States Navy. The environment at Naval Reactors, encouraging personal discipline and intellectual maturity, shaped me early in my professional career. I am particularly indebted to Brian McNamara, who taught me the meaning of completed staff work, and Joe Demko, who taught me many lessons

on management, hockey, and the value of unflappability in times of crisis.

I also would like to thank all past and present VLF group members, particularly Marek Golkowski, Prajwal Kulkarni, and Morris Cohen, with whom I had many discussions about the practical and theoretical aspects of wave-injection experiments. It was through the efforts of trying to explain my results to this intelligent, knowledgeable, and very skeptical audience that I was forced to refine my ideas until they became retrospectively obvious. The expertise and perspective provided by Ev Paschal was also very helpful in the development of my understanding of the history and facilities of Siple Station.

I would like to give special thanks to my parents Michael and Barbara. For every success in my life, I am and always will be indebted to them. More than these words, though, I hope that my every action acknowledges my gratitude to them. My brother Alex and his wife Kathy have also been a wellspring of support through this entire process, as evidenced by the phalanx of retriever-mangled plush toys that line my bookshelf.

I have been surrounded by leaders all of my life, but I have never met one more capable or inspiring than my brother Aaron; he is truly a Leader of Men. I have always been the beneficiary of his leadership and counsel, and this work is no exception.

To my beautiful Elisha, I would like to express my love and gratitude for your endless patience and unwavering support through this entire process. In little ways, you remind me every day why I fell in love with you.

<div align="right">ANDREW R. GIBBY</div>

Stanford, California
May 2008

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The purpose of this dissertation is to expand the scientific understanding of the mechanisms that drive the Very Low Frequency (VLF) triggered emission instability, a phenomenon associated with the earliest observations of VLF waves propagating in near-Earth space. In particular, this work focuses on understanding the amplification aspects of the instability by studying the termination of that amplification, that is, the characteristics of the instability associated with saturation. This work is primarily motivated by the pursuit of pure science and the understanding of the physical world and the near-Earth space environment. However, aspects of this work also have significance in historical, conceptual, and engineering application contexts.

This dissertation is intended to stand alone; a reader new to this topic should be able to understand the conclusions of this work with only an introductory understanding of electromagnetics and plasma physics, without detailed knowledge of the Siple Station wave injection experiments or nuanced understanding of the VLF triggered emission instability. The intent is to provide the unfamiliar reader with enough background to be conversant in both topics.

1.1 The Inner Magnetosphere

The region of space near the Earth within which electrodynamic processes are dominated by the influence of the Earth's magnetic field is known as the "Magnetosphere."

The magnetic field of the Earth is effectively that of a dipole (with an axis tilted and off-center relative to the rotation axis of the Earth), distorted by the action of the solar wind, so that the Magnetosphere has an asymmetric shape along the axis pointing into the solar wind. As a result, the magnetosphere extends approximately 10-12 Earth radii toward the Sun, where one Earth radius equals 6370 km, and more than 100 Earth radii away from the sun. The region where the Earth's magnetic field is approximately that of a dipole (say, less than 5 Earth radii from the center of the dipole) is known as the inner magnetosphere, which is the region of space considered in this study.

Space in the inner magnetosphere is composed entirely of plasma. The plasma in this region is magnetized by the influence of the geomagnetic field, fully ionized, and tenuous enough that it can be regarded as collisionless. Furthermore, the dipole-like geometry of the geomagnetic field traps the charged particles of the plasma as they gyrate about the magnetic field. The plasma in the inner magnetosphere is typically divided into two populations, as shown schematically in Figure 1.1. The majority of the plasma is cold, composed of particles with negligible thermal motion ($T \sim$ 5000 K or $\sim$0.5 eV [*Bezrukikh et al.*, 2003]), circulating through the magnetosphere relatively slowly, driven by the rotation of the Earth and the dynamo action of the solar wind on the boundary of the magnetosphere. One characteristic of this cold plasma population is that it frequently exhibits a sharp decrease in density near $L=4$ (where L is the McIlwain L-parameter, described in Section 1.1.1), which is known as the "plasmapause" [*Carpenter*, 1963, 1966]. The region of the inner magnetosphere within the plasmapause is known as the "plasmasphere."

In addition to the cold population, a small fraction (i.e., $\lesssim$1%) of the plasma in the inner magnetosphere is composed of energetic particles. Since the velocity of these particles in the direction of the magnetic field line is not, in general, zero, the particles are able to travel along the direction of the magnetic field line. Because of the near-dipole nature of the Earth's magnetic field, the magnetic field strength increases along a magnetic field line away from the geomagnetic equator. Accordingly as the gyrating electrons move towards the planet, their rate of gyration increases, at the expense of reduced parallel velocities, until they reach a point at which they
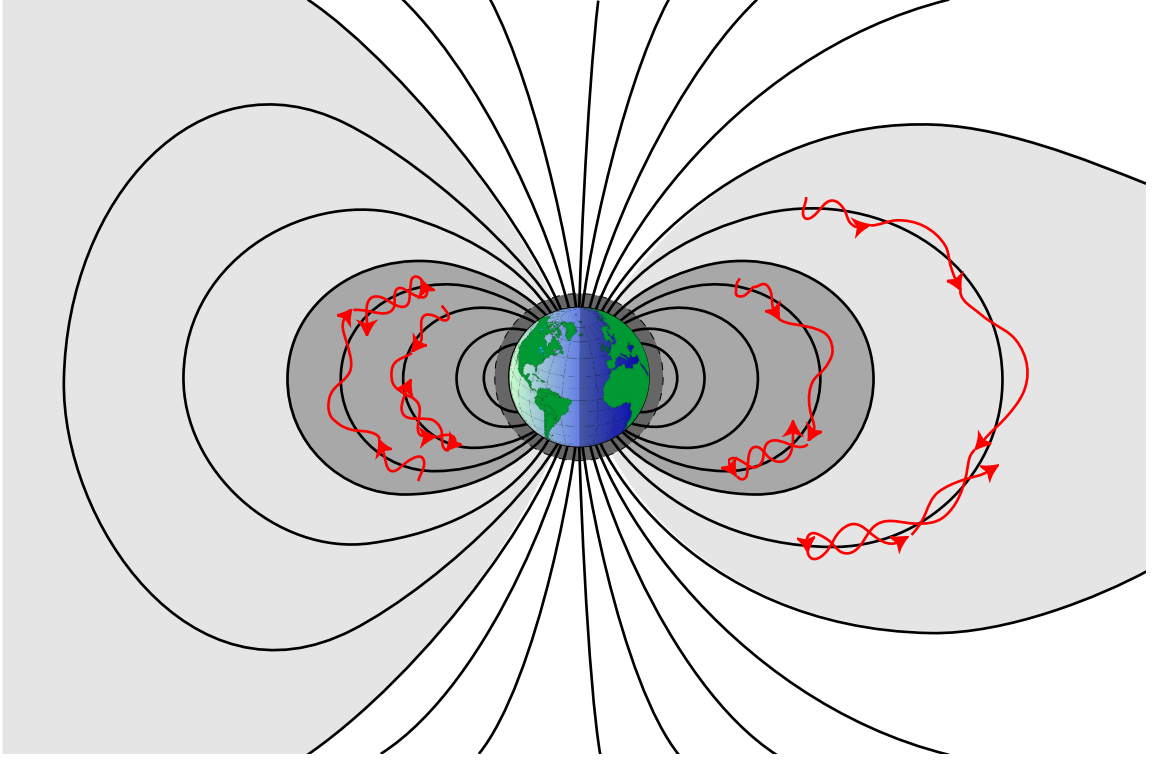
Figure 1.1: A schematic representation of the inner magnetosphere. The black lines represent magnetic field lines from the Earth; in this region, the geomagnetic field is close to that of a dipole. The asymmetry in the field line geometry is due to the action of the solar wind, where the sun is to the left in the figure. The gray scale colors represent the cold plasma density within and outside of the plasmasphere. The red lines represent the trajectories of trapped, radiation belt electrons. Adapted from *Inan* [1977].

"mirror" back to the other hemisphere. As a result, these particles bounce back and forth between opposite hemispheres as they gyrate about the magnetic field, forming the so-called trapped radiation belts of energetic particles.

The majority cold plasma population in the inner magnetosphere dominantly determines the propagation properties (i.e., wavelengths and phase velocities) of electromagnetic waves. For conditions typical in the inner magnetosphere, the only wave mode that can propagate in the VLF frequency range (technically 3-30 kHz, but effectively 1-20 kHz for the purposes of this work) is known as the "whistler mode."

More accurately (and throughout much of the plasmasphere), the whistler mode corresponds to waves with frequencies below the electron gyrofrequency (that is, the frequency at which the electrons gyrate about the ambient magnetic field) and above the proton gyrofrequency. Whistler mode waves propagating with k-vectors in the direction of the ambient magnetic field are right-hand circularly polarized, have electric and magnetic fields that are transverse to the direction of propagation, and propagate with a velocity that is significantly less than the speed of light in free space [*Helliwell*, 1965, Sec. 3.3].

The inner magnetosphere regularly contains field-aligned density enhancements which can efficiently guide whistler mode waves [*Smith*, 1961; *Angerami*, 1970]. These density enhancements are commonly referred to as "ducts" [*Helliwell*, 1965, Sec. 3.6]. The effect of the presence of these ducts is shown schematically in Figure 1.2, in which a terrestrial VLF source launches waves that enter the magnetosphere, propagating in the whistler mode. The direction of wave propagation is shown by two possible ray paths. The blue ray path represents a wave that is not ducted; the direction of propagation is controlled by the natural variation of the index of refraction in the inner magnetosphere, which is a function the local magnetic field strength, electron density, angle of propagation relative to the magnetic field, and other factors. Such non-ducted waves generally propagate with wave vectors at large angles with respect to the ambient magnetic field, and are not observable again on the ground. In contrast, the red ray path represents a wave that is coupled into a duct. In this case, the direction of propagation remains close to the direction of the ambient magnetic field. The wave propagates out to the geomagnetic equator, and then back to the ionosphere in the opposite hemisphere.

The ionosphere is a region of the Earth's upper atmosphere that is ionized by radiation from the sun. Relative to the overlying regions, the ionosphere is dense and collisional with a large neutral population. The bottom of the ionosphere exhibits a sharp gradient in the charged particle density, which represents an abrupt change in the index of refraction for whistler mode waves and gives rise to total internal reflection of most whistler mode waves that are incident from above. Ducted waves, however, arrive at this sharp boundary with near-vertical k-vectors, so that a portion
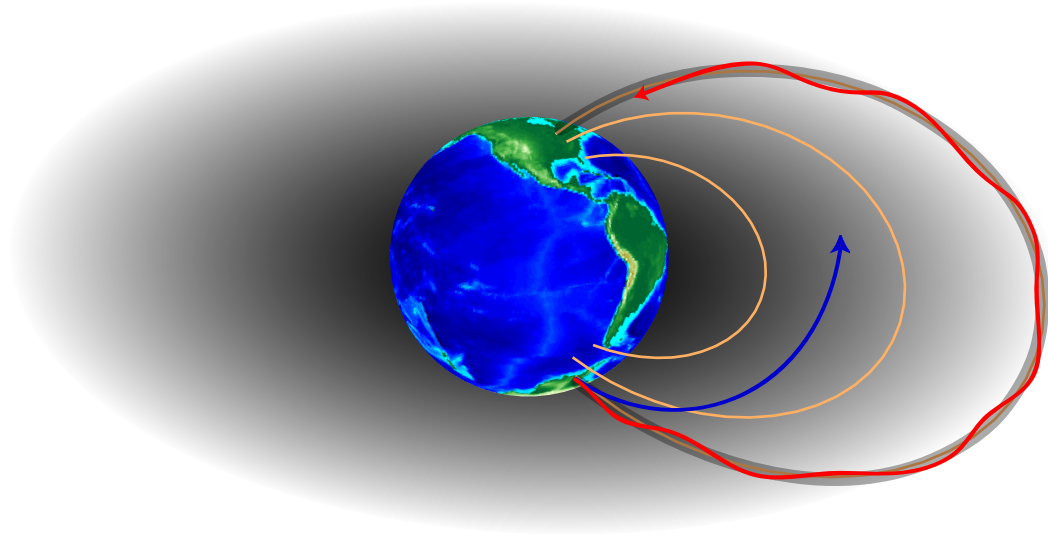
Figure 1.2: A schematic representation of the effects of ducts on the propagation of whistler mode waves. The blue and red lines represent ray paths for two signals from a terrestrial VLF source that injects whistler mode waves into the magnetosphere. The blue ray path is not ducted, and propagates on a path that does not return to the ionosphere in such a manner that it can penetrate the ionosphere. The red ray path is ducted, and follows the magnetic field line to be incident on the ionosphere in the northern hemisphere. Because of the near-vertical incidence of the ducted wave, it will be transmitted through the ionosphere.

of the wave energy can be transmitted through the ionosphere, and terrestrial VLF receivers may receive the signal that has propagated on a magnetospheric path.

While the cold electrons dominantly control the propagation characteristics and ray paths for whistler mode waves, the energetic radiation belt electrons can cause the wave to be damped or amplified. Indeed, the hot electrons provide the free energy necessary to drive the instability that is the focus of this study. In particular, the present study is focused on the physical processes by which a whistler wave modifies the distribution of the energetic electrons, and how those electrons, in turn, modify the wave or generate additional wave components.

1.1.1 The L-parameter

A convenient coordinate for describing magnetospheric positions in the frame of the geomagnetic field is the McIlwain L-parameter. The L-parameter is used to specify the surfaces along which particles trapped in the geomagnetic field move. The L value of a given drift shell in the real geomagnetic field corresponds to the equatorial crossing, in Earth radii, of the equivalent drift shell (i.e., the particles have the same adiabatic invariants) in a tilted, off-centered dipole model of the Earth's magnetic field. Thus, if a magnetic field line or position in the magnetosphere is specified using its L value, it corresponds roughly to a position on a dipole field line that crosses the magnetic equator at L Earth radii [*Walt*, 1994, Ch. 4].

1.2 Wave Injection Experiments

As discussed above, magnetospherically propagating whistler mode waves can couple into field-aligned guiding structures and correspondingly penetrate the ionosphere at the other hemisphere to be observed on the ground. This property allows these signals to be used for diagnostics of the condition of the plasma in the inner magnetosphere. For example, whistler mode waves generated by natural or man-made VLF sources on the ground can be used to determine the cold plasma density along a ducted path [*Sazhin et al.*, 1992]. Thus, the artificial generation of whistler mode waves can be used to study the magnetospheric plasma in a controlled manner.

The concept of such an experiment is shown in Figure 1.3. A terrestrial VLF transmitter, shown here in Antarctica, produces VLF waves that enter the magnetosphere as whistler mode waves. The whistler mode waves couple into a magnetospheric duct, if one is available, then propagate up the magnetic field line to a region near the magnetic equator. In the region of the radiation belts, energetic electrons are available to interact in cyclotron resonance with the whistler mode wave. Near the geomagnetic equator, the interaction between the whistler mode wave and the energetic electrons is strongest, and the wave and the electrons exchange energy and momentum, resulting in the modification of the wave (e.g., amplification of the wave
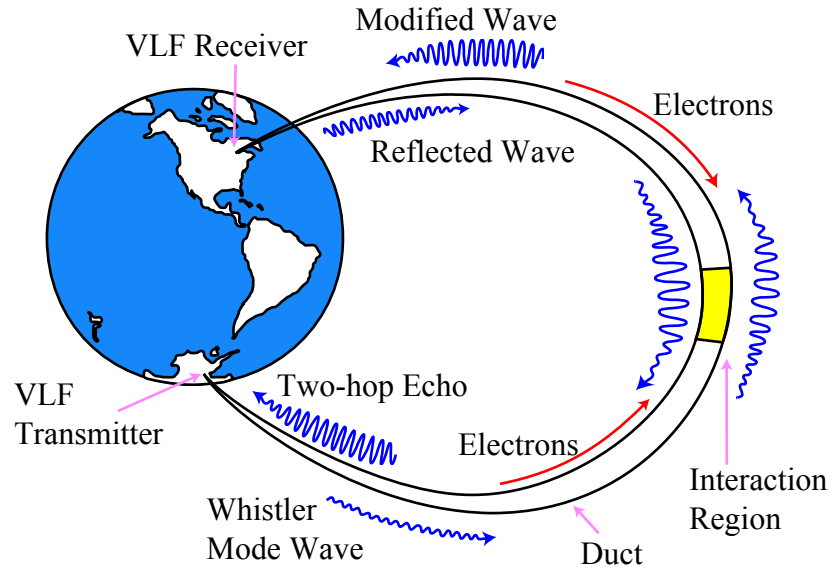
Figure 1.3: A schematic representation of a wave injection experiment. A VLF transmitter is located on the ground, shown here in Antarctica, which produces VLF waves that propagate into the magnetosphere, interact with energetic particles near the magnetic equator, and are received on the ground by a VLF receiver in the northern hemisphere.

and triggering of plasma emissions) and in the modification of the distribution of the energetic electrons. Following the interaction, the wave propagates down the field line to the northern hemisphere, where it penetrates through the ionosphere to be received by a ground-based VLF receiver.

At the ionosphere in the northern hemisphere, a portion of the wave energy is reflected back into the magnetosphere. That energy can couple back into the same duct, or other available ducts, propagating back through the equatorial region, interacting with the energetic particles a second time. After the second interaction, the wave continues down the field line to the southern hemisphere, where it can once again penetrate through the ionosphere. Such a signal is received as a "two-hop echo" of the original transmission, delayed by several seconds due to the slow group velocity of whistler mode waves under magnetospheric conditions.

These types of wave injection experiments serve as controlled methods for studying both the energetic electron population of the radiation belts, as well as the underlying

cold plasma of the inner magnetosphere. The energetic electrons serve to drive the unstable interaction with the whistler mode wave, so that the measurement of the amplification characteristics of the interaction, coupled with a detailed theoretical understanding of the phenomenon, can in principle reveal the state of the radiation belt electrons. On the other hand, the cold plasma controls the propagation and dispersion characteristics of the whistler mode waves. Consequently, measurement of the group delay of whistler mode signals can be used to determine the condition of the cold plasma, as has been well demonstrated by the use of whistler diagnostics [*Park*, 1972]. Thus, measurements of whistler mode wave phenomena can provide information about both plasma populations.

1.3 Studies of VLF Triggered Emissions

The unstable interaction between the energetic electrons of the radiations belts and ducted whistler mode waves has been observed for more than 40 years [*Helliwell*, 1965, Ch. 7]. At times, it has been labeled by the result of the instability most visibly apparent on a time-frequency spectrogram, i.e., "VLF triggered emissions," or by the apparent preference of the instability for the amplification of coherent waves, i.e., the "coherent wave instability." The salient aspects of the phenomenon are the temporal amplification of the whistler mode wave from its originally received level, as measured by a stationary receiver, and the subsequent triggering of free-running plasma emissions, where additional wave frequency components are generated by the instability. These new frequency components are considered free-running in that they do not correspond to frequency components of the transmitted whistler signal, with the exception that the emissions begin at or close to the frequency of the transmitted wave

Several wave injection experiments have been successful at activating the VLF triggered emission instability. The primary source of experimental data on this phenomenon comes from the Siple experiment, in which a VLF transmitter was operated at Siple Station, Antarctica, between 1973 and 1989. In addition to the results from Siple, a transportable VLF transmitter with a 1200 m balloon-suspended antenna was

used to successfully transmit signals at 6.6 kHz from Port Heiden, Alaska to Dunedin, New Zealand on magnetospheric paths, where the VLF triggered emission instability was observed [*Koons and Dazey*, 1974; *McPherson et al.*, 1974; *Dowden et al.*, 1978]. Most recently, VLF wave injection experiments have been performed using the High Frequency Active Auroral Research Program (HAARP), where VLF waves were generated by modulating the auroral electrojet using high frequency ionospheric heaters [*Inan et al.*, 2004b; *Platino et al.*, 2004]. As a result of these experiments, a large body of observational data has been collected on the VLF triggered emission instability.

Despite the large body of work characterizing the instability, much of the physical mechanism of the instability remains unexplained, but not for lack of trying. In the 1960s and 1970s, the theoretical efforts to understand this instability focused on analytical studies, that led to mostly phenomenological descriptions of results, the complete study of the phenomenon not being analytically tractable, because of the need to simultaneously solve Maxwell's equations and the Lorentz force (or Vlasov equation) in the presence of the inhomogeneous (i.e. spatially varying) geomagnetic field. [*Bell and Buneman*, 1964; *Das*, 1968; *Matsumoto and Kimura*, 1971; *Sudan and Ott*, 1971; *Dysthe*, 1971; *Brinca*, 1972, 1973; *Helliwell and Crystal*, 1973; *Nunn*, 1974; *Karpman et al.*, 1974; *Roux and Pellat*, 1978; *Cornilleau-Wehrlin and Gendrin*, 1979]

In addition to analytical work, a number of numerical modeling efforts have been used in attempts to reproduce the experimental manifestation of the instability. In the 1970s and 1980s, the modeling efforts were centered around the use of test-particle codes, which suffer from the constraint that they are not self-consistent in that only the motion of the particles, as described by the action of given wave fields, is considered [*Inan*, 1977; *Inan et al.*, 1978; *Bell and Inan*, 1981; *Serra*, 1984; *Inan*, 1987]. Self-consistent particle codes began to be used in the 1980s but were limited in terms of the size of the problem (i.e., either the dimensionality, the size of the interaction region, or the number and distribution of energetic particles) that could be considered [*Matsumoto et al.*, 1980; *Omura and Matsumoto*, 1982]. The first self-consistent code to consider the full problem was the Vlasov Hybrid Simulation, formulated by *Nunn*

[1990]. The Vlasov Hybrid formulation produced results consistent with experiment, but required the imposition of a number of non-physical elements in order to do so [*Nunn*, 1990, 1993; *Nunn et al.*, 2005]. Recently, computing power has become sufficient to address the instability using full particle models, and a number of particle-in-cell codes have been applied to the problem [*Omura and Summers*, 2004; *Katoh and Omura*, 2006; *Omura and Summers*, 2006; *Lampe et al.*, 2006]. These codes reproduce aspects of the phenomenon, but not in a manner that agrees well with experimental results, most likely because the codes have been driven using unrealistic initial conditions. Furthermore, particle-in-cell codes produce results specifying the state of a large number of individual particles. Consequently, the measurement of the plasma distribution function using these codes is cumbersome and seldom reported. Thus, physical understanding of the mechanism involved through the evolution of the distribution function during the interaction is infrequently gained from the application of these codes.

1.4 Thesis Organization

Chapter 1, the present chapter, introduces the motivation for this work and the relevant background for understanding the contributions of this work.

Chapter 2 begins with a history of the observational aspects of the VLF triggered emission instability, with particular attention paid to the history of the Siple wave injection experiment. In addition to the history, a concise review of the observed characteristics of the VLF triggered emission instability is provided.

In Chapter 3, the theoretical aspects of the VLF triggered emission instability are presented, starting from first principles, and focusing on a conceptual understanding of the wave-particle interaction that controls the instability.

Chapter 4 presents a new observational analysis of the characteristics of the VLF triggered emission instability specific to saturation. This presentation includes detailed consideration of the amplitude, frequency, and phase characteristics of the observed signals.

Chapter 5 first develops a new model for the study of the VLF triggered emission

instability. The necessary inputs for the model are specified so that they are consistent with experimental observations. Several results from the model are presented, showing that the model generates results that match experimental results, and that the model reproduces the observed characteristics of saturation. Sensitivity studies are also presented, which demonstrate the variation of the instability output as the input parameters are varied individually. The results of the model are also explained in the context of the theoretical discussion of Chapter 3.

Chapter 6 develops a new heuristic model for the VLF triggered emission instability using quasi-linear diffusion as a basis for the wave-particle interaction of the instability. The results of the diffusion model are compared to the previous, more general model and the advantages and deficiencies of a diffusion approach are discussed.

Chapter 7 summarizes the results of this work and concludes with a discussion of future extensions of the work presented in this dissertation.

1.5 Scientific Contributions

The contributions to knowledge presented in this dissertation are as follows:

- Analyzed and identified the characteristics of the VLF triggered emission instability specific to the saturation of the temporal growth of the signal.

- Developed two new models for the simulation of the VLF triggered emission instability.

- Applied the models to reproduce the saturation characteristics of the VLF triggered emission instability.

- Provided a physical explanation for the dynamics of the temporal growth and saturation process of the VLF triggered emission instability.

Chapter 2

VLF Triggered Emissions

2.1 The Discovery of VLF Triggered Emissions

The first observation of a transmitted VLF wave triggering a free-running plasma emission was recorded in November 1959 by the Stanford International Geophysical Year VLF receiver located in Wellington, New Zealand [*Helliwell et al.*, 1964]. In this case, the US Navy VLF transmitter NPG, located in Jim Creek, Washington, and transmitting at 18.6 kHz, triggered weak, rising frequency tones. The result was not noted until 1964 when the data record was being examined for another purpose. Examination of other records revealed that a case of triggering by the NAA transmitter, located in Cutler, Maine, and transmitting at 14.7 kHz, had been recorded by the VLF receiver aboard USNS *Eltanin* in 1961 [*Helliwell*, 1965, sec. 7.2]. On the *Eltanin* record, the triggering event was much richer, showing tones that varied both up and down in frequency over a much longer record. In addition, the record revealed that emissions were triggered almost exclusively by the 150 ms Morse dashes and only rarely by the 50 ms Morse dots. This remarkable phenomenon was dubbed the, "dot-dash anomaly."

These observations of VLF transmitters were not the first observations of VLF signals on magnetospheric paths triggering free-running emissions. Natural, discrete, varying frequency emissions known as "chorus" had been observed for years, with characteristics similar to those of emissions observed to be triggered by transmitters.

In addition, emissions had also been observed to be triggered by whistlers, magneto-spherically dispersed signals initiated by lightning.

The discovery of artificially stimulated emissions hinted at the possiblitly of controlled experiments on the magnetospheric plasma and the desire to do so led to the establishment of a VLF transmitter at a site known as Byrd Longwire, which was 18 km from Byrd Station, Antarctica (80.02° S, 119.53° W; Section 2.2.1 explains the rationale for the use of an Antarctic station as the site for a VLF transmitter) which operated between 1966 and 1969. The long wire of the antenna was laid out on the ice and was soon covered by accumulated snow. The ice lowered the Q (resonator quality factor) of the antenna due to near field losses, and movement of the ice created frequent breaks of the antenna wire, making regular operation impossible. Despite these problems, the transmitter was successfully used to sound the D and E regions of the ionosphere with VLF waves [*Helms et al.*, 1968], and signals from the transmitter were received aboard the OGO 4 spacecraft. However, signals from the Byrd transmitter were never received on the ground in the northern hemisphere, due to the large distance that the signals had to propagate in the Earth-ionosphere waveguide before reaching locations near the plasmapause where whistler ducts were available. Whistlers received at Byrd Station were typically observed to have propagated along paths near $L \sim 4$, while Byrd was located near $L \sim 7.25$, which corresponded to a $\sim$1000 km propagation distance in the Earth-ionosphere waveguide [*Helliwell and Katsufrakis*, 1974; *Helliwell*, 1977; *Helliwell and Katsufrakis*, 1978].

2.2 Siple Station

2.2.1 History

Although the experiment at Byrd produced mixed results, interest in the phenomenon of VLF triggered emissions continued, in light of accumulating evidence of natural triggering events. In 1969, an effort was undertaken to find the ideal site for a VLF wave injection experiment. A site was finally selected at 75.93° S, 84.25° W, in Ellsworth Land, 2381 km east of McMurdo Station, near the Sentinel range of the

Ellsworth Mountains. The site was found to be optimum, according to a number of selection criteria. Being in Antarctica, it was well removed from the major sources of electromagnetic noise in the VLF band, most notably man made interference and thunderstorm activity. The site was furthermore close to Eights Station, which had been identified specifically as an ideal location for monitoring natural VLF signals by Neil Brice, during an Antarctic traverse in December 1962, and had additionally proved optimum for monitoring US Navy VLF transmitters. At $L \sim 4.3$, the site offered access to high magnetic latitudes, the plasmapause, and abundant and complex natural VLF emissions. Additionally, the site was atop a 2 km thick sheet of ice, thus elevated above the conducting ground by a significant fraction of a wavelength, allowing reasonable radiation efficiencies [*Raghuram et al.*, 1974]. Finally, the magnetic conjugate point of the site was accessibly located near the city of Roberval in Quebec, Canada, making the establishment of a conjugate monitoring station straightforward [*Helliwell*, 1970; *Helliwell and Katsufrakis*, 1974; *Helliwell*, 1985, 1988a; *Paschal*, 2008b].

In the austral summer of 1969/70, Siple Station was established, named in honor of Paul Siple, the Antarctic pioneer, scientist, and Eagle Scout, who had passed away in 1968. Scientific operations began at Siple even before construction of the station was completed, including a balloon x-ray campaign in January 1971 that linked energetic particle precipitation from the magnetosphere to natural VLF noise bursts [*Rosenberg et al.*, 1971; Holmes and Narver, Inc.].

Station construction was completed in the austral summer of 1972/73 by Naval Mobile Construction Battalion Unit 71, making Siple the smallest and most remote American Antarctic station, with a winter-over complement of four personnel. The station was supplied completely by airlift, an operation that was frequently complicated by low visibility and winds up to 70 knots. Because Siple was supplied only by air, the crash of the ski-equipped LC-130 number 917 at South Pole on 28 January 1973 resulted in the loss of the aircraft and forced an early start to the station's first winter season. The first winter-over crew consisted of Jay Klink, station mechanic; Evans "Ev" Paschal, VLF transmitter operator; Russell Threlkeld, MD, station doctor; and William "Bill" Trabucco, station leader. [Holmes and Narver,

Inc.; *Trabucco*, c.1990].

Figure 2.1: Artist's conception of Siple Station in its original configuration, with a single, 21.2 km dipole antenna. Shown also is a ski-equipped LC-130 transferring fuel to the station.

The station was occupied in a winter-over state continuously from 1973 to 1980 with the exception of 1976, when a two-aircraft crash at Carrefour before the station could be supplied for the year resulted in the evacuation of the station. In the 1980s, the station was wintered over in 1982, 1983, and 1986 [*Helliwell*, 1986]

The station in its original configuration (known as Siple I) consisted primarily of a single 12 gauge corrugated metal arch, 13.5 m wide and 66 m long. Inside the arch were living and working areas cobbled together out of modules on metal skids. The modules were originally used at the Byrd Longwire site with an additional T-5

(prefabricated-panel) building from McMurdo Station [*Paschal*, 2008b]. Electrical power was provided by three diesel generators, one of which, rated to 300 kW, provided power for the scientific experiments, while the other two, rated to 90 kW each, provided power for the station. Fuel was also stored within the arch in two 50,000 gallon bladders. (The proximity between the fuel storage and living spaces incidentally led to an aroma of "eau de DFA [Diesel Fuel-Arctic]" for food that was not properly stowed in reserved food storage areas [*Trabucco*, 1987b].)

In addition to the facilities located within the arch, fuel and supply caches were located on the surface of the ice for emergency use or for storage of hazardous material. In the summer, a "summer camp" was created on the ice to house the additional construction and maintenance personnel present for station upgrade or upkeep. In addition to additional maintenance activity, the summer also served as a time for additional science campaigns, most notably those associated with Arcus and Nike Tomahawk rockets in the summers of 1977/78 and 1980/81 [Holmes and Narver, Inc.; *Kintner et al.*, 1983; *Trabucco*, c.1990].

The annual snow and ice accumulation at Siple was approximately 1.6 m per year, on average. The result of this accumulation was the slow burial of surface caches and the arch structure itself. The caches were periodically cleared out and moved to maintain accessibility, while the station was effectively allowed to become an underground facility. Furthermore, the rising ice level also required the transmitting antenna to be raised periodically. The build up of ice above the station led to the deformation of the arch, and in the austral summer of 1977/78, the construction of a second, 80 m long arch (dubbed Siple II) began. The two were joined by a 2 m metal passageway, with the Siple I arch continuing in service as a fuel storage facility. To support the upgrade to the Jupiter transmitter (see Section 2.2.2), an additional 25,000 gallons of fuel were stored in the Siple II arch [Holmes and Narver, Inc.; *Trabucco*, 1987b].

The VLF receiving station in the northern hemisphere was located near Roberval, Quebec, and was established in 1970. As the experiment proceeded over the next two decades, development in the Roberval area, in particular an increase in activity at the Aluminum Company of Canada bauxite reduction plants, led to an increasing

amount of background noise from power line harmonic radiation. By 1981, the station at Roberval was divided into two parts, a main station and a quiet station 20 km away. Eventually, the installation of two new 735 kV transmission lines, coupled with the secular variation of the Earth's magnetic field and corresponding northwesterly drift of the conjugate point of Siple Station, led to the establishment of a new VLF site in 1985 at Lake Mistissini (in French: Lac Mistassini), on a Cree Indian reservation 300 km away [*Helliwell*, 1986, 1987a].

In the final years of operation (the summers of 1986/87, 1987/88, and 1988/89), Siple Station was manned on a summer-only basis, which reduced the overall maintenance requirements on the station, slowed the degradation of the Siple II arch, and reduced fuel consumption, thus (most importantly) reducing the required number of supply flights to the station. Finally, in January 1989, the station was closed due to funding reductions and a reduction in the availability of aircraft for supply flights. As part of the station closing, the longwire antenna was cut every kilometer, to reduce any potential inductive effect that could occur if a new transmitter were to be established at Siple at some later time [*Helliwell*, 1986; *Chiang*, 1987; *Lynch*, 1987; *Trabucco*, 1987a; *Helliwell*, 1988b,c].

2.2.2 Facilities

The primary components of the Siple wave injection experiment were the VLF transmitter and accompanying dipole antenna. The initial requirements on the transmitter were simple: it had to be flexible enough to reproduce the dot-dash anomaly. The requirements on the antenna were similarly straightforward, in that it only had to radiate with sufficient efficiency to illuminate the magnetosphere. Previous observations of the Forrest Port, New York, Omega navigation transmitter had shown that 100 W transmitters were capable of triggering emissions [*Kimura*, 1968], setting an antenna efficiency requirement on the order of a few percent.

In the austral summer of 1969/70, a test dipole was erected at Siple. Impedance measurements on the test antenna led to the design of a 21.2 km antenna, suspended above the ice on 6 m aluminum channel supports every 61 m, giving the antenna

a resonant frequency of approximately 5 kHz. The antenna was installed along the magnetic east-west direction, and was appropriately dubbed the "longwire" antenna to differentiate it from the various other antennas at the station, particularly the 9 m high antenna for the Siple VLF receiver [*Helliwell*, 1983b; *Trabucco*, c.1990].

The original transmitter at Siple, named "Zeus," was the same solid-state switching transmitter that had been used at Byrd, providing up to 80 kW. The transmitter was capable of transmitting frequencies between 1.5 kHz and 20 kHz and could change frequencies every 10 ms using a computer controlled audio frequency synthesizer, whose accuracy was determined by a rubidium frequency standard. The transmitter was limited in that it could only provide constant amplitude, phase-continuous signals, and the high Q of the pulsed resonators of the transmitter gave the system a bandwidth of approximately 500 Hz, requiring frequent retuning to achieve operational frequency changes. Also, the nature of the frequency synthesizer in the Zeus transmitter further restricted the transmission of frequencies above the tuned frequency. Fixed-duration pulses were formed using frequency-shift-keying, where the transmitter provided power at the desired frequency for the length of the pulse (say, 200 ms at 5 kHz) and then shifted to another "idler" frequency to which the transmitter was not tuned for the desired gap before the next transmission (say, 200 ms at 4.5 kHz). The idler pulses served to keep a constant load on the generators [*Helliwell and Katsufrakis*, 1974, 1978; *Chang et al.*, 1980; *Paschal*, 2008b].

Soon into the experiment, maximum power out of the transmitter was identified as desirable; increasing the radiated power offered a number of experimental advantages. These advantages included: increased scattering of trapped electrons (and increased precipitation of trapped radiation as a result); increased range of L-shells which signals could reach; stronger signals received in the northern hemisphere, even when amplification was not present; and increased occurrence of emission triggering. To that end, a surplus transmitter from the US Navy's Omega navigation network was obtained in the austral summer of 1973/74 and later installed in the austral summer of 1978/79. At 150 kW, this transmitter, named "Jupiter," offered a significant gain over Zeus. Additionally, Jupiter was significantly more flexible than

Zeus and was capable of operating at two different frequencies, provided by two independent computer controlled audio frequency synthesizers, the outputs of which could be either summed (for independent frequency transmissions) or multiplied (for modulated transmissions). Also, Jupiter was capable of changing frequency and/or amplitude every millisecond. Finally, the tuning network between the transmitter and antenna was redesigned so that Jupiter could provide even more power at lower frequencies [*Helliwell*, 1981a; *Carpenter*, 1982a; *Carpenter and Bao*, 1983; *Helliwell*, 1983b, 1988a].

To encourage propagation at higher magnetic latitudes (see Section 2.3.5), the resonant frequency of the antenna was reduced in the austral summer of 1982/83 to 2.5 kHz by extending the longwire antenna to 42 km [*Helliwell*, 1984]. Knife switches were installed at points approximately 10.5 km from the station, giving the capability to disconnect the new portion of the antenna if a higher resonant frequency for the system was desired [*Helliwell*, 1983b].

Starting in 1980, plans were made for adding an additional 42 km dipole to the antenna. There was some debate over whether or not the second dipole should be added parallel to the existing longwire antenna or perpendicular to it, in a "cross" configuration. Both configurations offered a 3 dB gain over the existing antenna. The cross offered the gain by enabling the transmission of right-hand circularly polarized waves, which would more directly convert to the whistler mode; left-hand polarized waves were reflected or absorbed by the ionosphere, resulting in only 50% of the power entering the ionosphere from the existing linearly polarized waves. The parallel configuration offered a similar gain by increasing the radiation resistance of the antenna system. In the end, the cross was selected since it was easier to drive, with the feed points for both dipoles at the station. In comparison, the parallel dipole required an additional ∼10 km transmission line to provide power. One drawback for the cross was that it had to be either buried under the aircraft skiway or disconnected during flight operations [*Helliwell*, 1981d].

The cross was installed in two stages. In the 1984/85 and 1985/86 austral summers, a 21.2 km section was installed in the magnetic north-south direction, with a portion that ran over the skiway. This resulted in the disconnection of the second

21 km of the east-west longwire so that the resonant frequency of the transmitter was near 5 kHz for the majority of the 1986 operating year. In October of 1986, the north-south cross was disconnected to permit flight operations, and the east-west antenna was reconfigured to be 42 km in length, reducing the resonant frequency of the antenna back to ~2.5 kHz. In the following summer, the remainder of the cross was installed, and the section crossing the skiway was buried [*Helliwell*, 1983b; *Wolf*, 1986; *Helliwell*, 1989b; *Trabucco*, c.1990].

In addition to the VLF transmitter at Siple and the VLF receivers at Siple and the conjugate ground stations in Quebec, additional scientific equipment was installed at both locations which was not directly related to the wave injection experiment. The equipment included fluxgate magnetometers, pulsation magnetometers, and riometers, and at times HF radars, photometers, and digital ionosondes [*Lanzerotti*, 1978; *Rosenberg and Barcus*, 1978; *Helliwell*, 1981a; *Paschal*, 2008b].

2.2.3 Station Life

Operations at Siple Station were focused around two activities: science and life support. Due to the short manpower, particularly during winter-over periods, duties were shared between scientific and operational staff. Additionally, any visitor to the station was expected to assist in the day-to-day upkeep activities [Holmes and Narver, Inc.].

The VLF experiments were operated by two Stanford employees specifically hired for the task, who usually shared the duties of running the transmitter and the passive (i.e., recording) laboratory. As post-graduate engineers, the operators were expected to perform engineering tasks, such as programming and upgrading the transmitter, as well as routine maintenance. During twelve-hour shifts, the operators were expected to make the VLF transmissions, acquire data from the passive experiments at the station, and observe and record the weather every six hours, as well as carry out additional community duties [*Helliwell*, 1983a,b; *Logan*, 1986a]. *Paschal* [2008a] offered this recollection of station life during the 1973 winter-over:

Bill [Trabucco] and I were employed by Stanford. Jay [Klink] and Dr.

> [Russell] Threlkeld were employed by Holmes and Narver, who had the Antarctic support contract with the NSF at the time. Bill, Jay, and I all shared cooking on successive days, though Bill was the only baker among us. Threlkeld spent most of his time in his room writing a novel, and ate there on his own for the most part. Bill, Jay and I...had a wonderful, very busy, winter. We had all wintered-over before and expected lots of free time. We were surprised how much work we needed to do [to] get the station finished and all the experiments working. It was probably the workload that made our time so satisfying. Bill had been at McMurdo [Station] in 1969 as the Cosray operator for Bartol. Jay had been at Palmer [Station] in 1970 as the mechanic when he was in the Navy. I had been the station science (civilian, as opposed to Navy) leader at Byrd [Station] during 1970 for Stanford. We remain good friends. Threlkeld had little work to do and had a tough winter. None of us have heard from him since he left at the beginning of the next summer.

Recreation often involved climbing up and down ladders; outdoor activities, when safe, were encouraged for mental and physical well-being, and the aurora hut, sporting the Station's only window, was a popular location to relax. In addition, a station library, movies, ham radio, games, hobbies, and crafts were also available. As a curious final note, the manufacture of alcohol was explicitly forbidden, but bottled beer, wine, and spirits were available in recreational quantities [Holmes and Narver, Inc.; *Trabucco*, 1987b; *Paschal*, 2008b].

2.3   The Siple Wave Injection Experiment

2.3.1   Operations

The operation of the Siple Experiment was coordinated between three distinct locations: Stanford, Siple, and Roberval (or Lake Mistissini). Stanford provided overall direction for the experiment, while Siple and Roberval were responsible for the generation, reception, and recording of the VLF signals corresponding to the triggered emission instability.

Time-frequency-amplitude formats for the transmitter were designed at Stanford. The formats were guided by previous experiments to answer specific questions. For

example, a format was designed to reproduce and study the dot-dash anomaly in the first operating year of the transmitter. The formats were designed by a variety of individuals, including students and staff, and approved for transmission by the senior Stanford researchers. Approved formats were conveyed to the Siple operator, who programed them into the transmitter [*Helliwell and Katsufrakis*, 1974; *Helliwell*, 1981a].

Individual formats were bundled into "packages," so that each package was designed to study a variety of effects [*Helliwell*, 1988a]. This use of packages was an experimental design decision, trading off between the repetition of one format (and the potential for the repetition of results for a single aspect of the instability) and results from many formats (and the potential to reveal new aspects of the instability). This design decision came from the character of signal reception in the northern hemisphere; although the reception of signals from Siple was not uncommon (see Section 2.3.2), periods of exceptional reception, from which the detailed behavior of the instability could be teased out of the data, were rare. Thus, the decision was made to study a variety of effects rather than individual effects in depth.

At the receiving stations in the northern hemisphere, continuous VLF recordings were made on reel-to-reel magnetic tapes during Siple operations. When the transmitter was not operating, "synoptic" recordings were made, during which one minute of data was recorded every 5, 15, 30, or 60 minutes. Approximately every month, the data from Roberval (or Lake Mistissini) were packaged and shipped to Stanford, where 35 mm spectrogram records were made. The spectrograms were then reviewed and interesting phenomena were identified and cataloged. On an as-needed basis, further processing was performed on selected records, which included the production of narrowband amplitude plots, specific spectrograms with different frequency resolution characteristics, and quantitative measurements of spectral characteristics [*Helliwell*, 1981a].

The VLF receiver at Siple was operated only in the synoptic manner, due to limitations on the quantity of magnetic tape available for data storage. An unintended side-effect of the installation of the north-south leg of the cross antenna (see Section 2.2.2) was to couple noise from the high voltage exciter into the VLF receiver.

As a result, the operator was required to manually interrupt the high voltage for 30 s during half of the synoptic minutes so that the natural VLF background and echoes of previous transmissions could be observed.

During transmitter operations, ATS-3 satellite communications were established between Siple and Roberval so that the character of signal reception could be fed back to the Siple operator [*Helliwell*, 1981a]. During transmissions to the DE-1 and EXOS-B satellites, the satellite receiver data were telemetered in real time to Stanford from the NASA DSN tracking station in Goldstone, California, and operational parameters were fed back to the Siple operator through the ATS-3 link [*Carpenter*, 1982b; *Bell et al.*, 1983a]. During the summer-only operation of the Siple transmitter after 1986, additional research personnel were on-hand at Siple to direct the experiment [*Helliwell*, 1987b].

In 1973, the transmitter was operated at fixed frequency set points determined beforehand. In 1974, a real time spectrum analyzer was available to the Siple operator, and frequency setpoints could be selected in real time based on observations of natural activity or echoes of transmitted signals [*Helliwell and Katsufrakis*, 1974; *Carpenter and Miller*, 1976; *Helliwell*, 1981b]. This became the general mode in which the experiment was performed. For example, the 1986 operator used the following procedure for the selection of the transmitter tuned frequency, f_{set} [*Carpenter*, 1986a; *Logan*, 1986b]:

1. f_{set} is selected from the following frequencies:

   - 4980-5040 Hz - The frequency range corresponded to the resonance of the 21 km-by-21 km cross antenna (see Section 2.2.2).

   - 4020 Hz - A frequency that did not interact with the station science generator, which was operated at 58 Hz.

   - 3360-3600 Hz - The frequency range that tended to produce the strongest reception in the northern hemisphere.

   - 3000 Hz - Another frequency that did not interact with the station science generator.

- 2700 Hz - The lowest tunable frequency.

2. Check for echoes by transmitting at each of the above frequencies. Set f_{set} to the frequency of strongest echoes.

3. If no echoes are present, but a natural band of VLF noise is present, set f_{set} 500 Hz below the maximum frequency of the noise band.

4. If no echoes or band of noise is present, select frequency at random from the above frequencies.

From the above procedure, one can tell that the frequency selection process late in the experiment had become quite complex, based on previous experience. Some aspects, such as the placement of f_{set} near the top of a VLF noise band, were based solely on experience and have not yet been quantified.

2.3.2   Reception Statistics

As mentioned above, reception of Siple signals in the northern hemisphere is not an uncommon occurrence. For example, in 1986 data is available for 140 days of transmitter operation. Of those days, signals are recorded in the northern hemisphere on 90 days, which corresponded to reception on 61% of the operating days. On a minute-by-minute basis, approximately 38% of individual transmissions are received during this period. In contrast to mere reception of the signals, less than 10% of the operating days yield "excellent" conditions, wherein the signal to noise ratio in the northern hemisphere is sufficiently high to yield measurements sufficient for scientific purposes [*Helliwell*, 1988a].

Table 2.1 shows the reception percentage given by the ratio:

$$\frac{\text{Number of days transmissions were received in the northern hemisphere}}{\text{Number of days Siple transmitted}}$$

for each year of transmitter operation. Statistics have not yet determined for the 1979 and 1983 operating years. The effect of the upgrade from Zeus to Jupiter is clearly seen in the numbers, where the daily reception percentage jumps from ~25% to ~40%.

	Percent of Days in which
Year	Transmissions Were Received
1973	24
1974	16
1975	22
1977	25
1978	20
1979	-
1980	28
1982	58
1983	-
1986	61

Table 2.1: Siple yearly reception statistics. Reception percentages are given by the ratio, (Number of days received in the northern hemisphere)/(Number of days transmitted). [*Carpenter and Miller*, 1976; *Miller*, 1979; *Carpenter*, 1982a; *Carpenter and Bao*, 1983]

This effect is attributed to the increase in power between the two transmitters, as well as the ability for Jupiter to provide even greater power at lower frequencies. The difference from year to year is attributed primarily to changes in geomagnetic activity. Particularly, fewer transmissions are succesfully received when geomagnetic conditions are highly disturbed [*Carpenter*, 1982a; *Carpenter and Bao*, 1983].

Signal reception in the conjugate region (at Roberval or Lake Mistissini) also follows a diurnal pattern. Signals are detected at all local times, but the overall patterns show a night side minimum, a maximum in reception near local dawn, a decreasing trend through the day, and a secondary peak in activity near 1900 magnetic local time [*Helliwell and Katsufrakis*, 1974; *Carpenter and Miller*, 1976; *Carpenter*, 1982a; *Carpenter and Bao*, 1983]. This trend appears consistently in the periods analyzed (1979, 1980, 1982, and 1986). On a seasonal basis, Siple transmissions tend to be more successful during austral winter than summer, due to reduced ionospheric absorption above the transmitter [*Carpenter and Miller*, 1976; *Miller*, 1979; *Carpenter*, 1982a; *Carpenter and Bao*, 1983].

As an exception to the trends, there was one period in the austral summer of 1980/81 (during the Nike/Argus rocket campaign) which showed very successful reception at unexpected times, perhaps due to the abnormal campaign operation of
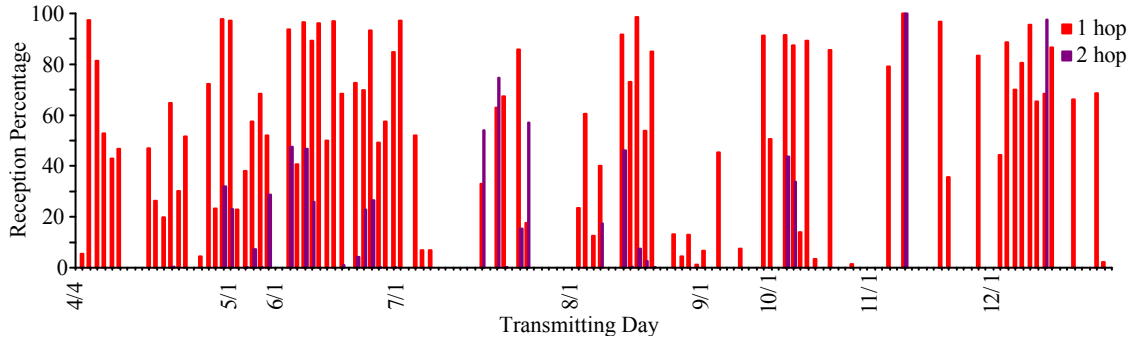
Figure 2.2: Reception percentages for every operating day in 1986 in which the VLF receiver at Lake Mistissini was operating. Red bars indicate reception percentages for "1 hop" signals propagating to the northern hemisphere and purple bars indicate reception percentages for "2 hop" signals received by the Siple VLF receiver.

the transmitter. During the period between 1 December 1980 and 14 January 1981, 16 days of transmissions were attempted. Of the 17 distinct periods in which the transmitter was operated, signals were received in the northern hemisphere in 14 of those periods. These highly successful transmission periods were all concentrated near local noon [*Carpenter*, 1981, 1982a].

Reception of two-hop echoes at Siple are significantly less likely than reception of one-hop signals in the northern hemisphere. In 1986, echoes are received on 21 days, giving a daily reception percentage of only 15%, as opposed to a 61% reception percentage in the northern hemisphere. The minute-by-minute reception is similarly reduced to ∼7% instead of 38% reception in the northern hemisphere . As shown in Figure 2.2, on occasion the reception of two-hop echoes is more likely than the reception of signals in the northern hemisphere. This behavior is interpreted as evidence of a "screening effect" in which signals arriving in the northern hemisphere could not propagate through the ionosphere [*Carpenter*, 1983]. There is some evidence in ionosonde data from previous years that a sporadic E layer in the ionosphere above the receiver may have been responsible for this effect, but this relationship was never quantified [*Carpenter*, 1983].

Another interesting trend to note in the data from Figure 2.2 is the lack of apparent preference for reception in the austral winter. In particular, the activity is located in

individual periods, while reception during July, August, and September is particularly sparse. This exception to the seasonal affect is also observed in 1980, when reception is spaced throughout the year in 21 separate periods, including particularly good reception in the austral summer, as previously mentioned [*Carpenter*, 1981].

Reception of signals in the north is also related to geomagnetic conditions [*Carpenter and Miller*, 1976]. In general, conditions for reception become worse as geomagnetic activity increases. Furthermore, the days of exceptional activity tend to occur on the geomagnetically quietest days. The difference in success rates in the 1980 and 1982 operating years is attributed to a significant decrease in activity in 1982 over 1980. In particular, 1980 had 62 days in which the sum of the planetary geomagnetic index, $\sum K_p$, was greater than 15, indicating strong geomagnetic disturbances. In contrast, 1982 had only 20 days with this level of disturbance [*Carpenter and Bao*, 1983]. This trend is representative of a dependence of signal reception on solar activity, with 1979/1980 representing a solar maximum period.

In the presence of natural VLF activity, transmissions tend to be more successful in the presence of lightning-generated whistlers and in the presence of band-limited VLF noise [*Carpenter*, 1987]. Once again, however, this relationship has not been not quantified, and warrants further study.

Polarization experiments performed in 1986 and later [*Mielke et al.*, 1992] show that the transmission of right-hand polarized waves increases the reception percentage relative to the transmission of linearly polarized waves, as expected. Similarly, left-hand polarized waves are much less likely to be received in the northern hemisphere than linearly polarized waves. Any reception of left-hand polarized signals is attributed to mode conversion at the lower ionosphere due to oblique incidence of the transmitted wave. Left-hand polarized waves are, on average, received at levels approximately 15 dB below the levels at which right-hand polarized waves are received.

Finally, the statistics available (before 1983) show that the vast majority of signal receptions occurred when f_{set} satisfies the condition $0.2 f_{c,\text{eq}} < f_{\text{set}} < 0.5 f_{c,\text{eq}}$, where $f_{c,\text{eq}}$ is the electron gyrofrequency at the equator along the path of propagation. During this period, the most successful frequencies of transmission are at 5 kHz, which

corresponds to the resonant frequency of the antenna, and approximately 3 kHz. The apparent success at 3 kHz is due to the range of frequencies corresponding to the propagation path with its foot directly over Siple, representing the path with the minimum subionospheric propagation loss on average [*Carpenter and Bao*, 1983].

2.3.3   Propagation Path

The cold plasma in the magnetosphere regularly contains field-aligned density irregularities that efficiently guide whistler mode signals. These density irregularities, called "ducts," tend to occur frequently in the inner magnetosphere widely distributed over a range of $L > 2$. The ducts act to locally increase the index of refraction for whistler-mode waves, thus serving as guiding structures. The presence of ducts allows the signals generated by Siple to propagate through the magnetosphere on field-aligned paths, with wave normals closely aligned to the ambient magnetic field, and although Siple signals may exit the duct as they approach the northern hemisphere, their entry point into the Earth-ionosphere waveguide typically stays within 100 km of the latitude (at ionospheric heights) of the field line along which they propagate [*Leavitt et al.*, 1978; *Carpenter*, 1982a]. More importantly, their wave normal angles stay nearly field aligned and in fact tend closer to vertical, thus being in the narrow transmission cone for penetration through the sharp lower ionospheric boundary.

One characteristic of ducts is the fact that the guiding of whistler mode waves within them over the entire magnetospheric path is only possible when the wave frequency is lower than half the local electron gyrofrequency, i.e. $\omega < \omega_c/2$. Although ducting at frequencies above $\omega_c/2$ may occur within ducts composed of density depletions, unducting occurs for all types of ducts when the wave frequency equals half the local electron gyrofrequency [*Helliwell*, 1965, Sec. 3.6]. Since the geomagnetic field strength along the path is minimum at the equator, the unducting condition $\omega = \omega_c/2$ is satisfied for wave frequencies above the equatorial value of this quantity (i.e., $\omega_{c,eq}/2$) at points along the propagation path. This characteristic explains the reduction in ground-based whistler-mode signal reception at frequencies above $\omega_{c,eq}/2$. [*Carpenter*, 1968; *Angerami*, 1970; *Carpenter and Miller*, 1976; *Carlson et al.*, 1985].

The "average" transmission path has its endpoint centered on the magnetic field line above Siple at approximately 100 km altitude. As mentioned above, this path represents the most likely path due to a minimum of spreading loss in the Earth-ionosphere wave guide. In addition, Siple, at $L \sim 4.3$, is situated such that the plasmapause is regularly in close proximity to the station. As a result, Siple transmissions typically propagate within the plasmasphere or in density gradients close to it, such that electron densities along the transmission path are usually within a factor of two of typical plasmaspheric densities [*Helliwell*, 1987c].

Propagation along paths outside the plasmapause occur only infrequently, usually during anomalous quieting periods during moderate geomagnetic storms. The reception of signals propagating outside the plasmapause is associated with short travel times ($\sim$1 s), emissions with high frequency sweep rates (up to 4 times typical rates), and amplitudes much higher than those of the triggering signal [*Carpenter and Miller*, 1983].

Naturally, signals from the Siple transmitter can excite multiple ducts when they are available. Thus, signals in the northern hemisphere can typically be received over multiple paths and a corresponding number of delay times. For constant frequency signals, the reception of signals from multiple paths corresponds to an apparent lengthening of the received pulse relative to the transmitted pulse. For signals that change frequency with time, the signals are seen to arrive at discrete, separate times. The magnetospheric responses on multiple paths vary greatly, with different amounts of growth and triggering seen on each path. Generally, though, the response from one duct dominates, with amplitudes $\sim$5 dB greater than the amplitudes received from other paths. However, the strongest signal path can shift from one duct to another within minutes, then shift back at some later time [*Carpenter*, 1980; *Sonwalkar et al.*, 1997].

Because of the ducted nature of the propagation path, there are a large number of factors that affect the amplitude and reception of signals in the northern hemisphere. These effects include [*Sonwalkar et al.*, 1997]:

1. Spreading in the Earth-ionosphere waveguide to the magnetospheric injection point.

2. Transmission and absorption through the ionosphere. Rocket measurements noted that wave power dropped by 17 dB during trans-ionospheric propagation [*Kintner et al.*, 1983].

3. Coupling into the duct. Of the wave energy entering the magnetosphere, only ~10% was estimated to enter the ducted mode [*Bell and Inan*, 1975; *Helliwell and Katsufrakis*, 1978].

4. Spreading in the duct to the region where the wave can interact with resonant electrons. Ducts were inferred to be approximately 200-400 km wide at the equator, corresponding to a 15-30 km width at an altitude of 300 km. [*Angerami*, 1970]

5. Plasma interactions. These included the nonlinear interaction of the triggered emission instability as well as linear growth and Landau damping.

6. Propagation from the interaction region to the ionosphere, including spreading from the duct exit point.

7. Transmission and absorption through the ionosphere. Echoes of Siple transmissions indicated that a significant fraction of the incident wave was reflected at the ionosphere in the northern hemisphere.

8. Spreading in the Earth-ionosphere waveguide to the receiver location.

2.3.4 Energetic Electron Distribution

The energetic electron distribution (see Section 3.1.3) with which the Siple signals interact is one of the chief unknowns of the experiment. The rapid switching of activity from one duct to another as well as temporal changes in the growth parameters of the instability indicate that the interacting distribution had some level of spatial structure that is made evident as the energetic electrons drift through the spatial duct structure defined by the cold electron population [*Carpenter*, 1980]. Growth parameters are nearly constant on a 30 second time scale, but can change drastically on time scales

of 15 to 100 minutes [*Sonwalkar et al.*, 1997]. Due to this near-term repeatability of the response, the modification of the distribution function for a single transmission is generally considered to be modest. However, repeatable changes to the detailed response can appear on the order of minutes [*Golkowski et al.*, 2008].

Several *in situ* measurements of the energetic electron distribution function are available, associated with reception of Siple signals or triggered emissions from other VLF transmitters. *Kimura et al.* [1983] received Siple signals aboard the EXOS-B spacecraft coincident with highly anisotropic distribution functions which became more anisotropic as the spacecraft approached the plasmapause. In five orbits, the spacecraft received Siple signals with additional triggered emissions. In these cases, the distribution function was observed to be high in flux, but not highly anisotropic at pitch angles above 40 degrees, where the pitch angle is defined as the angle between the electron velocity vector and the ambient magnetic field (measurements below 40 degrees were not made).

While no other associated energetic particle measurements are available from periods during which the Siple experiments were carried out, *Bell et al.* [2000] report the observation of triggering of emissions from an Omega VLF transmitter, onboard the POLAR spacecraft, also in association with a highly anisotropic electron distribution. In this case, a functional fit to the observed data for 1-20 keV electrons is found to be

$$F_{\text{eq}} \propto \frac{\sin^{25} \alpha}{v^4} \qquad \tan \alpha = \frac{v_\perp}{v_\parallel}$$

where α is the particle's pitch angle. Functions of the form

$$F_{\text{eq}} \propto \frac{\sin^m \alpha}{v^n}$$

are frequently used to model the energetic electron distribution, particularly because they offer a simple relation between m and the linear growth rate for whistler mode waves (see Section 3.1.4) [*Gurnett and Bhattacharjee*, 2005, sec. 9.3].

Although the two cases mentioned above are the only observations of electron distribution function in conjunction with triggering VLF signals, there is additional evidence of highly anisotropic electron distributions near the plasmapause. *Burke*

et al. [1995] observe aboard the CRRES spacecraft that the distribution function of 10-30 keV electrons near the plasmapause is significantly peaked at pitch angles above 50 degrees. Furthermore, the structure of the observed distribution indicates multiple "clouds" of anisotropic electrons limited in longitudinal extent. In contrast to these observations within the plasmapause, observations of the electron distribution outside the plasmapause during the same period contain a significant population of electrons all the way down to the loss cone.

Thus, the likely form of the interacting energetic electron distribution in the triggered emission instability is highly anisotropic with a high energetic electron flux. Some theoretical work [*Pasmanik et al.*, 2002] indicates that if the distribution function contains a step discontinuity in velocity space, the growth of the wave can be enhanced by up to a factor of 60. This type of distribution function is not observed aboard the spacecraft operated during the Siple experiment; however, it would have been difficult to observe such a distribution due to the limited resolution of the particle detectors used.

2.3.5 Evolution of the Goals of the Siple Experiments

The goals of the Siple experiment evolved as the experiment matured. Specifically, the goals shifted as the capability of the transmitter was upgraded and the operational team gained experience with the experiment. As a result, the experimental aims became more nuanced as the experiment progressed, adapting to the knowledge gained during earlier operations.

In its original configuration, one of the stated goals of Siple was simply to demonstrate that the experiment could be performed [*Helliwell*, 1970]. Earlier experience with a much shorter antenna, laid out on the ice at Byrd Station (see Section 2.2.1), was largely unsuccessful, and it was unclear that Siple could even transmit into the magnetosphere. Also unknown were the actual radiation efficiency of the transmitter and the conversion efficiency into the whistler mode in the magnetosphere.

Once the feasibility of the experiment was demonstrated, the experimental goals shifted to investigations of the phenomenon of triggered emissions, and energy was

devoted to detailed measurements of the various effects of the instability and classification of the spectral morphology of the triggered emissions [*Helliwell*, 1981a,c, 1983b, 1985, 1986; *Carpenter*, 1986b; *Helliwell*, 1989a].

In addition to the wave-particle interactions associated with the triggered emission instability, efforts were made to induce particle scattering using Siple signals and detect the resulting precipitation. The upgrade to the higher power Jupiter transmitter, the lengthening of the antenna to 42 km to increase the system efficiency at lower frequencies, and the installation of the cross antenna to transmit a polarization preferential to ionospheric penetration were all, in part, intended to increase the likelihood of precipitating particles. This increased precipitation was to be brought about either by increasing the power injected into the magnetosphere, or by allowing lower frequency operation, which corresponded to interactions with higher energy electrons, the ionospheric effects of which would have been easier to detect. Lower frequency operation also encouraged propagation at higher *L*-values, and hence, outside the plasmapause, where precipitation was expected to be stronger and/or possibly more detectable, since the precipitation would involve higher energy electrons [*Helliwell*, 1981a; *Matthews*, 1981; *Helliwell*, 1985, 1983b, 1986, 1989b].

2.3.6 Experimental Critiques

The Siple Experiment was largely funded by the National Science Foundation (NSF) under the Division of Polar Programs with some analysis funding also provided by the Atmospheric Sciences Division. As a result, the grant proposals for the experiment were subject to anonymous peer review. These reviews offered an insight into the reception of the experiment by the contemporary scientific community. At times, the reviews were complementary, calling it, "one of the safest scientific investments the NSF can make," [*Anonymous*, 1983b] and, "among the most interesting of the past decade" [*Anonymous*, 1987]. Others [*Anonymous*, 1989] were less complimentary, categorizing the experiment as, "phenomenology and bird-watching," or, "recondite, jargon-filled, and increasingly unclear in its purpose."

Fundamentally, the critiques [*Anonymous*, 1980, 1983a,b, 1987, 1989] were centered around one common theme: the emphasis in the experiment was too phenomenological without enough formal treatment of the plasma physics involved in the instability. Reports of experimental results involved recitations of the characteristics of received signals with little explanation in theoretical terms. The mechanisms for the instability proposed by Stanford, while creative and intuitive, failed to provide testable predictions.

Furthermore, the experiment suffered as a result of this lack of a unifying theory that could be tested. The complexity of the formats (i.e. the individual experiments) increased with time according to the increasing capability of the transmitter. This increased complexity, in turn led to the discovery of increasingly complex, apparently "new" phenomena. However, without a unifying framework under which the results could be understood, no differentiation could be made between new phenomena and phenomena which may have just been more complex manifestations of previously observed effects. As a result, researchers from the rest of the community, not being deeply immersed in the nuance of the experiment, became steadily less receptive toward its continuation.

In its defense, however, the Siple experiment was disadvantaged in that it was conducted about 20 years too soon. The majority of the study of the instability was not analytically tractable (see Chapter 3); the last purely analytical paper studying the instability was authored by *Karpman, et al.* in 1974, offering little understanding of the physical mechanism for the instability. Since then, numerical analyses have been exclusively applied to the problem, growing slowly in sophistication and realism as computational ability has increased, to the point where the models can be driven with realistic inputs to yield realistic outputs based on fundamental physical principles.

2.4 Characteristics of the Instability

2.4.1 Growth and Triggering

The two most striking aspects of the VLF triggered emission instability are the exponential temporal growth of the signal from its initially received value and the triggering of free running plasma emissions. The growth experienced by a whistler mode wave packet as it interacts with the energetic electrons can be conceptually divided into two categories. The first is spatial growth that occurs uniformly at each position along the wave packet, which would not be measurable by a stationary VLF receiver, neglecting initial rise-time effects. Because of the fact that this type of spatial growth is that which is predicted by linear plasma theory, it is termed "linear" growth (see Section 3.1). The second form of growth involves some feedback process such that later portions of the wave packet are amplified more than initial portions. It is this second form of growth that the terrestrial VLF receivers at Roberval and Lake Mistissini measure. Due to its contrast to the character of linear growth, this form of growth is called "nonlinear" growth.

Linear Growth

Due to the spatial character of linear growth, it is not observable by terrestrial VLF receivers, neglecting initial rise-time effects. Therefore, direct measurements of the amount of linear growth experienced by Siple signals are difficult to make, requiring observations in which multiple, closely spaced spacecraft with VLF receivers fly through the region in which this growth is occurring. Because of this difficulty, no direct measurements of spatially growing signals have been reported. However, approximately 1% of the signals received in the northern hemisphere have no visible temporal growth, indicating that linear growth is the only active process, if any growth occurs at all. Under these conditions, an estimate of the path losses gives an estimate of the total linear growth to be on the order of 20 dB [*Dowden et al.*, 1978; *Carlson et al.*, 1985]. However, given the number of unknowns in the condition of a given propagation path (see Section 2.3.3), this estimate is highly uncertain.

Additionally, there are other indications that linear growth is active during the instability. Notably, the initial level (i.e., before any temporal growth) at which Siple signals is received varies on 5-15 minute time scales, which is faster than the variation that is expected for other path parameters, such as duct availability or ionospheric absorption [*McPherson et al.*, 1974; *Sonwalkar et al.*, 1997]. In particular, the total growth along the path (and the corresponding initially received level) due to linear effects is independent of the frequency content of the wave. In one case [*Carpenter et al.*, 1997], the initial reception amplitude for a single frequency transmission at 2150 Hz is the same as for a two-frequency transmission, with waves at 2140 Hz and 2160 Hz. Despite having the same initial level, the subsequent temporal growth characteristics for the two-frequency transmission differ significantly from the temporal growth characteristics of the single frequency transmission. This result indicates that the initial level of the signal is controlled by linear growth, while the subsequent temporal growth is controlled by nonlinear factors, including the presence of other whistler mode signals closely spaced in frequency (see Section 2.4.3).

Nonlinear Growth

In contrast, the nonlinear growth driven by the instability is characterized by temporal exponential growth, as observed at a fixed location. A particularly clear example of the growth process is shown in Figure 2.3, in which the characteristics of the process are clearly visible, including the triggering of a free running emission. One remarkable aspect of the nonlinear growth process is that it is often active when no natural VLF activity or emissions are observable on the ground, which indicates that the instability registers a preference toward the amplification of coherent signals (i.e., with bandwidths less than 10 Hz) [*Stiles and Helliwell*, 1977; *Helliwell and Katsufrakis*, 1978]. Because of this, it is often called the "coherent wave instability."

The nonlinear growth process is highly selective in frequency [*Helliwell and Katsufrakis*, 1978]. That is, the bandwidth of frequencies over which the nonlinear growth process is active can be as low as 200 Hz [*Bell et al.*, 1983a]. Specifically, the selection of transmission frequency, f_{set}, is more important in activating the instability than any other factor.
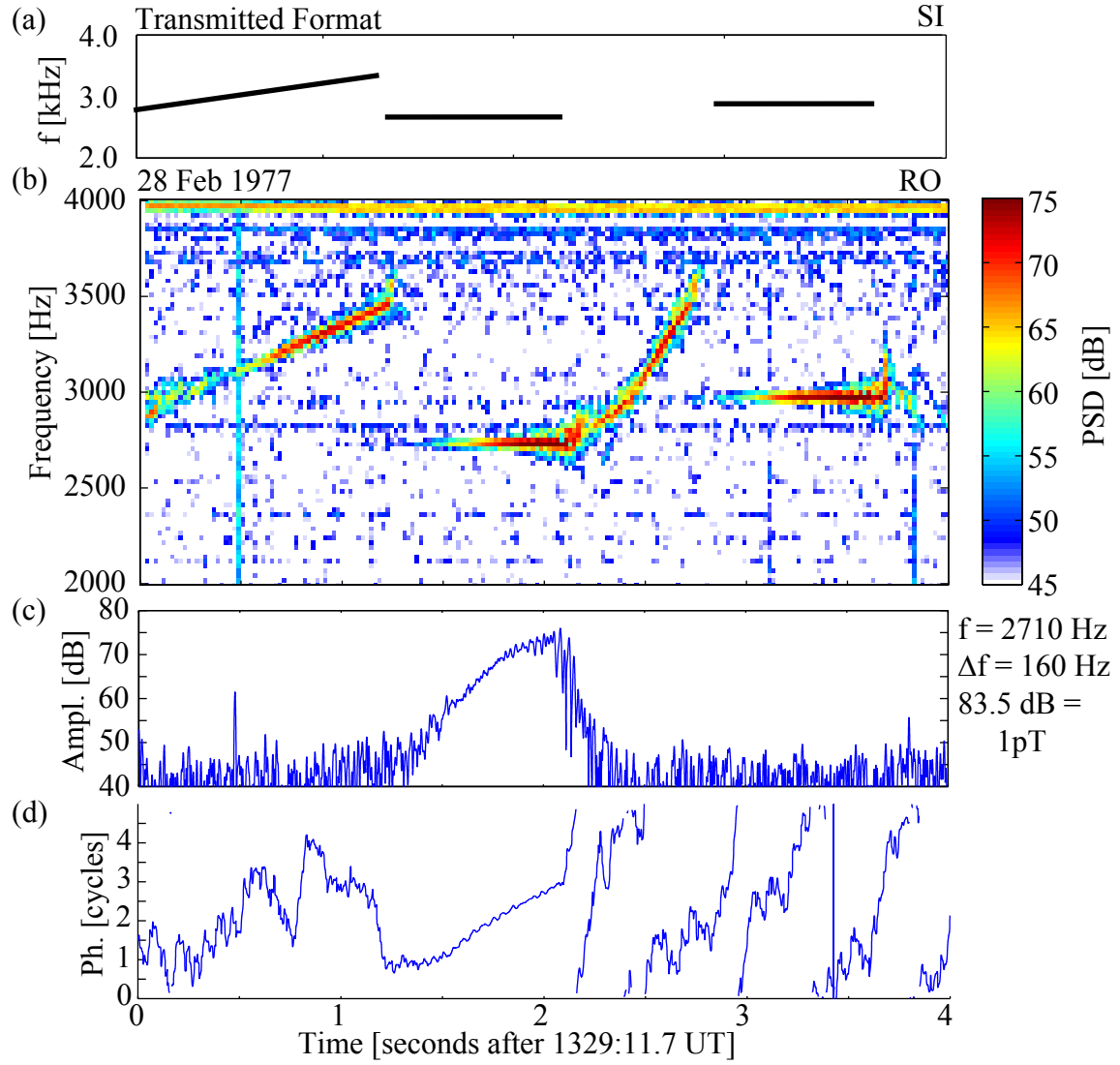
Figure 2.3: Growth and triggering characteristics of the instability. The characteristics are illustrated using a particularly clear example from a 2.71 kHz pulse transmitted on 20 February 1977. (a) The time frequency format transmitted from Siple. (b) The time-frequency spectrogram of the VLF data received at Roberval (c) A narrowband amplitude plot, centered on 2.71 kHz, with a 160 Hz bandwidth. (d) The phase of the narrowband data in (c).

The nonlinear growth process yields between 20 dB and 35 dB of total growth, with 30 dB being an often quoted value. The exponential growth rate varies widely, between 25 dB/s and 250 dB/s, on 5-15 minute time scales [*Helliwell and Katsufrakis*, 1974; *Helliwell*, 1977; *Stiles and Helliwell*, 1977; *Helliwell and Katsufrakis*, 1978; *Helliwell et al.*, 1980; *Helliwell*, 1983a,b; *Carpenter et al.*, 1997]. These growth parameters are considered to be typical, but the variation and statistical characteristics of these parameters have yet to be quantified. In addition, the growth rate is uncorrelated with the transmission frequency (over the band of frequencies in the active band), initial level, or length of the transmission [*Carpenter et al.*, 1997]. The only factors that appear to affect the growth rate are the presence of other signals close in frequency to the transmission frequency, or the use of an input signal that changes in frequency, such as the time-frequency ramp in Figure 2.3 [*Stiles and Helliwell*, 1977; *Carlson et al.*, 1985].

The initial phase of the transmitted wave has no effect on the nonlinear growth characteristics of the instability [*Chang et al.*, 1980]. During growth, the received signal shows a small, but always positive, change in frequency, with phase measurements indicating a regular phase advance at a rate of 1-10 Hz. In the cases where the growth is non-monotonic, with variations in the slope of growth or even suppression of the wave, the periods of growth disruption are associated with rapid changes in phase [*Helliwell*, 1983c; *Paschal and Helliwell*, 1984].

Threshold for Amplification

One oft-cited characteristic of the temporal growth is that it does not occur unless the triggering signal is above a certain amplitude threshold. Specifically, transmissions with slowly increasing amplitude experience a rapid transition from a period of no evident temporal growth, during which the received power is approximately proportional to the transmitted power, to a period with strong temporal growth. Unsurprisingly, the level of transmitted power at which this occurs varies widely with time (up to 30 dB) and with individual ducts [*Helliwell et al.*, 1980; *Helliwell*, 1983a,b, 1986, 1988a].

Saturation

Assuming that the transmitted pulse is long enough in duration, the temporal growth inevitably leads to saturation. At saturation, the amplitude character of the signal becomes irregular and oscillatory, with no well-defined level [*Stiles and Helliwell*, 1977; *Dowden et al.*, 1978]. Indeed, the amplitude in many cases reaches a second saturation level 5-10 dB higher than the initial saturation level [*Dowden et al.*, 1978]. Under conditions of a single, constant propagation path, the saturation level remains constant over a period of approximately one hour, with small ($\sim$10 dB) fluctuations on 5-15 minute time scales [*Carlson et al.*, 1985; *Sonwalkar et al.*, 1997]. Saturated signals remain at the saturation level until a rising frequency emission is triggered, at which point the signals are suppressed back to a level near or below the initial signal level [*Stiles and Helliwell*, 1977; *Chang and Helliwell*, 1979]. The signals remain suppressed until the rising emission separates from the triggering frequency by approximately 70 Hz. Additionally, at saturation the phase of the signal occasionally jumps ahead one or more cycles before becoming phase locked again with the saturated signal; this behavior coincides with the oscillations that occur at saturation [*Dowden et al.*, 1978; *Paschal and Helliwell*, 1984]. Chapter 4 provides a thorough discussion of the characteristics of the instability at saturation, along with examples and analysis of experimental data.

Triggering

The triggering of free running emissions is a natural consequence of the growth process, in which the generated wave frequencies differ significantly from the frequency of the triggering signal while remaining relatively narrow in bandwidth. The frequency of the triggered emission starts at the frequency of the triggering signal (within the frequency resolution of the measurement), and is associated with a significant increase in the frequency content (i.e. a broadening in spectral content) around that frequency. In general, the frequency of the signal initially increases with time [*Stiles and Helliwell*, 1975].

Two types of triggering are evident (see Figure 2.4): triggering at saturation and

triggering at pulse termination [*Helliwell and Katsufrakis*, 1978; *Helliwell*, 1986]. If triggering occurs at saturation, then only rising frequency emissions are produced [*Helliwell*, 1983a]. Following the triggering of the rising emission, the signal grows again to the saturation level at which point another emission is triggered. This pattern continues until the pulse terminates [*Helliwell*, 1988a]. As the signal growth rate increases, the period of this pattern decreases due to the more rapid return to saturation. Triggering also occurs at pulse termination (see Figure 2.4). If the pulse terminates below, but near, the saturation level, the emission continues to grow up to the saturation level. If, however, the pulse terminates significantly below the saturation level, no emission is triggered [*Stiles and Helliwell*, 1977]. This lack of triggering at low signal levels is responsible for the dot-dash anomaly, as shorter pulses experience less total amplification before termination and therefore are less likely to trigger on termination.

Termination triggering produces both rising frequency emissions and falling frequency emissions [*Helliwell*, 1988a]. Rising frequency emissions are triggered during periods of higher growth rate, while falling emissions are triggered during periods of lower growth rate, indicating that the time-frequency character of termination is also related to the signal amplitude at termination [*Lasch*, 1969; *Stiles and Helliwell*, 1977; *Helliwell*, 1977]. Similarly, shorter pulses, which have experienced less amplification, are more likely to trigger fallers than longer pulses [*Helliwell and Katsufrakis*, 1974].

Generally, rising frequency emissions are associated with oscillating amplitude variations as the free running emission progresses, while falling frequency emissions show less variation [*Carlson et al.*, 1985]. Falling frequency emissions also tend to show a more repeatable time-frequency character with a narrower instantaneous bandwidth [*Stiles and Helliwell*, 1975].

When multiple paths are available, triggering can occur on some or all of the paths, with the triggering on the various paths often being similar in character. Emissions from multiple paths are able to cross (i.e. have the same frequency at the same time, as observed at the receiver) without apparent interaction.
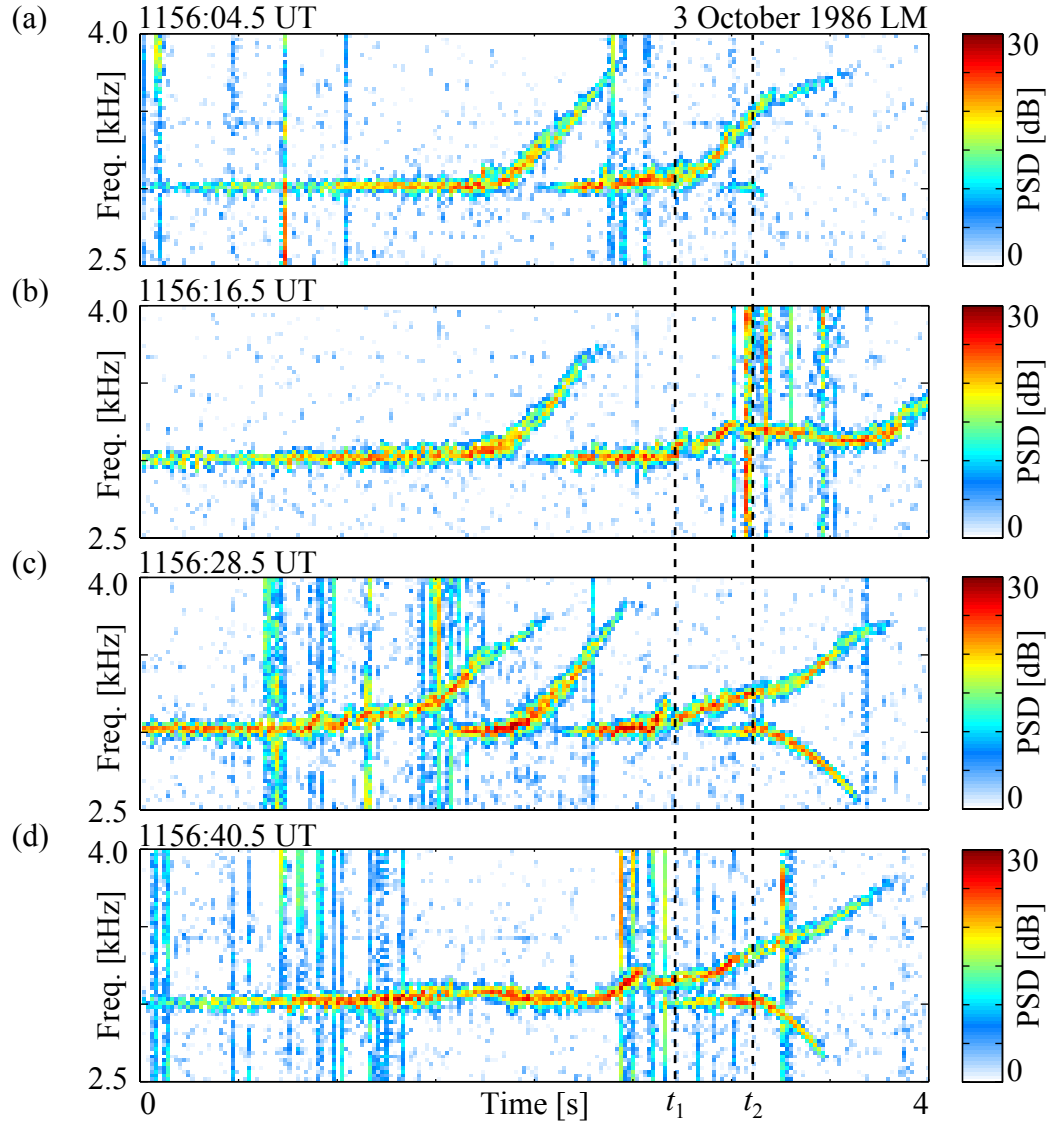
Figure 2.4: Triggering characteristics. Four examples of triggering are shown, from data taken at Lake Mistissini at 1156 UT on 3 October 1986, displayed as time-frequency spectrograms. Each 4-second record follows the previous record by 12 seconds, showing the last several seconds of a 5 second pulse. Two whistler ducts are active, and the termination times of the pulses are marked by dashed vertical lines. Specifically, the termination times of the signals propagating through the shorter delay duct and longer delay duct occur at times t_1 and t_2, respectively. The shorter delay duct is more active, triggering rising, pre-termination emissions, as well as rising emission on termination; the longer delay duct is less active, triggering falling emissions on termination.

Band Limited Impulse

At any time after saturation, an impulsive emission of 50-500 Hz in bandwidth can accompany the signal [*Helliwell*, 1988a]. This "band limited impulse" often occurs at the termination or interruption of constant frequency signals (as seen in the second constant-frequency pulse in Figure 2.3), but can also occur during the evolution of a free running emission. Often, the frequency band of the impulse is above the frequency of the signal that spawns it. In addition, the band limited impulse frequency often triggers free running emissions [*Helliwell*, 2000].

2.4.2 Sidebands

There are two distinct sideband effects that occur during the VLF triggered emission instability. The first is associated with a transition from growth to triggering in which the sidebands appear after some growth has occurred and do not appear to affect the characteristics of the growth. As shown in Chapter 4, this sideband effect is associated with saturation of the instability. The second sideband effect is associated with two-wave modulation, and is discussed in more detail in Section 2.4.3. The difference between the two effects is shown in Figure 2.5, where the single-wave sideband effect is shown on the right, and the two-wave sideband effect is shown on the left. The distinction between the two (other than the formats of the transmitted signals) is that the appearance of sidebands occurs immediately for the two-wave effect, while sidebands for the single-wave effect appear after some growth. This difference is most clearly recognizable in the narrowband amplitude plot of Figure 2.5(c), in which the beating associated with the two-wave sidebands is visible during the growth phase of the instability, while there is no such amplitude fluctuations during the growth phase of the single-wave transmission.

Single-wave sidebands are most easily observed on signals with slow growth rates, when the wave remains near the saturation level for some time before triggering an emission. Sidebands can develop either symmetrically about the transmitted wave frequency, or asymmetrically, with a frequency spacing ranging from a fraction of a Hertz to tens of Hertz. In the case of asymmetric sidebands, the sidebands at

frequencies above the transmitted frequency tend to have higher amplitudes. High order sidebands can occur, up to the 7th order, extending up to 100 Hz above and below the frequency of the transmitted wave [*Park and Chang*, 1978; *Chang et al.*, 1980; *Park*, 1981; *Helliwell*, 1988a; *Sa and Helliwell*, 1988].

2.4.3   Two-wave experiments

Throughout the Siple wave injection experiment, formats were transmitted that contained multiple frequency components. With the constant-amplitude limitation of the Zeus transmitter, multiple frequencies were transmitted by rapid frequency shift keying, such that the modulation of the signal created fundamental and sideband components of the transmission. The Jupiter transmitter, in contrast, had two independent frequency synthesizers which could transmit two waves with arbitrary frequency modulation. From these experiments, many aspects of the coherence of the interaction were observed.

Two waves with a frequency spacing of $\sim$5 Hz or less behave like a single wave, and waves with frequency spacings greater than approximately 120 Hz generate independent magnetospheric responses. Between these frequency spacing values, the response is suppressed relative to the independent response, with a minimum response occurring at a frequency spacing of approximately 20 Hz. This suppression, occurring as soon as the second frequency component is added to the transmitted signal, is up to 15 dB, with the suppression of the lower frequency typically greater than the suppression of the higher frequency signal. In addition, transmissions containing simulated noise have a similar suppression effect on the response to coherent signals. This suppression is believed to be due to a disruption of the coherent nature of the instability and as a result is called the "coherence bandwidth" effect [*Helliwell et al.*, 1986; *Helliwell*, 1988a; *Carpenter et al.*, 1997]. In the literature [e.g.,*Helliwell*, 1988; *Mielke and Helliwell*, 1992], the suppression of the magnetospheric response is exclusively attributed to an increase in the threshold for amplification (see Section 2.4.1), that is, the threshold is believed to increase when two waves are present. However the observations can also be explained by a reduction in the growth rate and/or
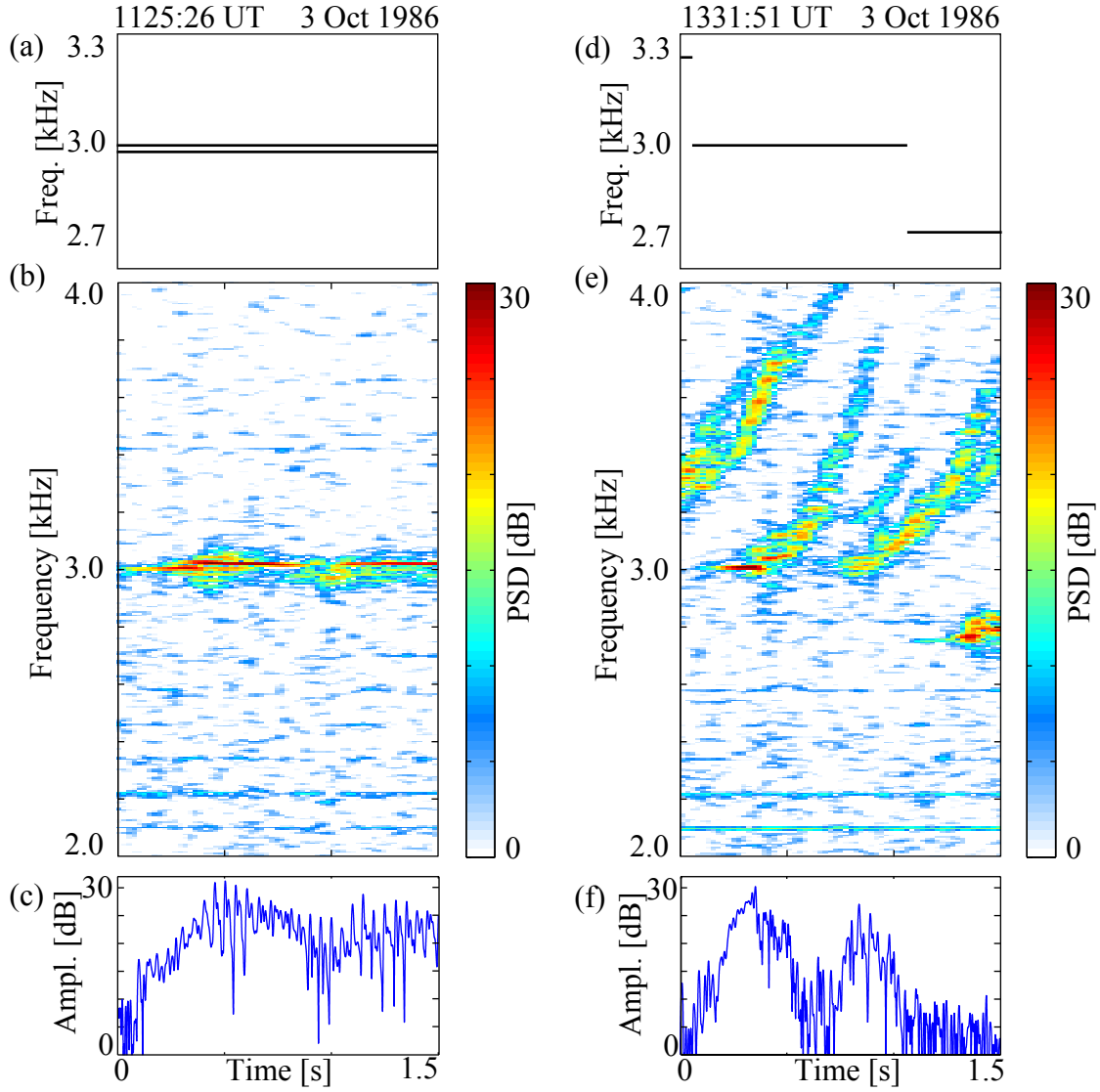
Figure 2.5: Sideband effects in the instability. Two distinct 1.5 s segments from data received at Lake Mistissini on 3 October 1986 are displayed. The left segment, (a-c), received at 1125:26 UT, represents the reception of a transmission using two closely spaced frequencies at 2.98 and 3.0 kHz, while the right segment, (d-f), received at 1331:51 UT, represents the reception of a single frequency transmission at 3.0 kHz. (a,d) The time frequency formats transmitted from Siple. (b,e) The time-frequency spectrograms of the VLF data received at Lake Mistissini (c,f) Narrowband amplitude plots, centered on 3.0 kHz, with a 160 Hz bandwidth.

saturation level of the instability.

The modulation of the input wave naturally creates sideband frequencies that alters the magnetospheric response. In this case, additional sidebands (i.e., other than those present in the transmitted signal) appear immediately on the received signal, without additional amplification. In effect, the instability alters the modulation of the input signal at the beginning of the growth phase of the instability. At saturation, these additional sidebands can trigger emissions of their own. As shown on the left side of Figure 2.5, two waves of the same input amplitude can be amplified by different amounts. In this example, in fact, the strongest received component is at 3.02 kHz, which is a sideband of the original signal. Another characteristic of this effect shown in Figure 2.5(c) is the beat pattern observed in the narrowband amplitude plot. There, two beats are received at the beat frequency of the transmitted waves (20 Hz), but the first beat is preferentially amplified over the second beat, followed by a deep null in the received signal. In some cases, this pattern can trigger weak emissions at the beat frequency [*Park and Chang*, 1978; *Helliwell*, 1979a; *Chang et al.*, 1980].

2.4.4 Echo Suppression

A closely related phenomenon to suppression by multiple input waves at closely spaced frequencies is one of echo suppression, an example of which is shown in Figure 2.6. In this case, the response to a transmitted signal is suppressed by the echo of a previous Siple transmission. Specifically, a transmitted signal reflects off the ionosphere in the northern hemisphere, propagates back to the southern hemisphere, then reflects a second time. At that point, the signal propagates back to the interaction region. If a second transmission is propagating through the interaction region at that time, the response to the second signal is suppressed. In the example shown in Figure 2.6, both the second signal and the echoing signal are part of the same 20 s constant frequency transmission.

Echo suppression occurs if the two signals are within 100 Hz of one another with transmission times separated by less than the two-hop propagation time. The reduction in response of the second transmission is typically 3 dB down, but can be as
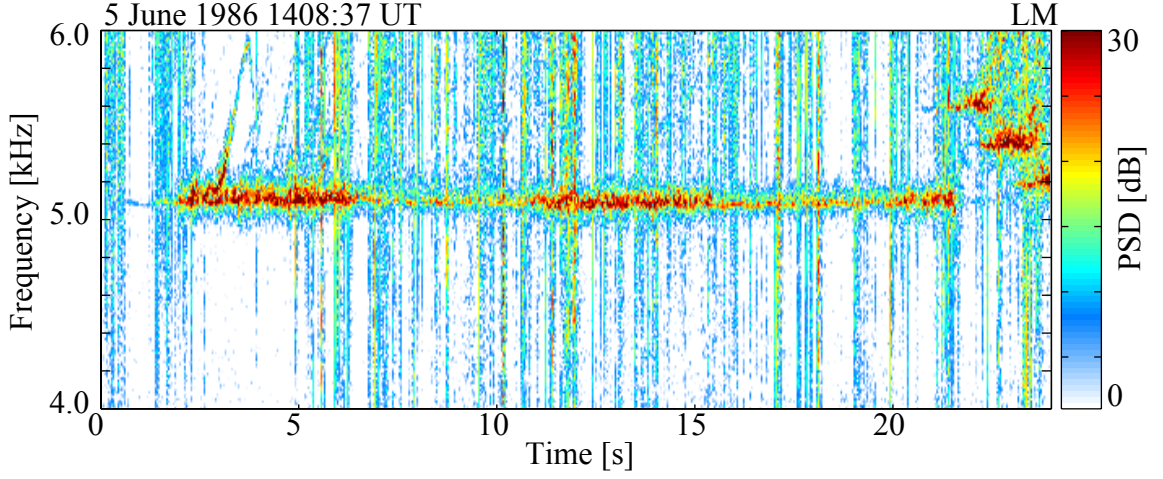
Figure 2.6: Echo Suppression. Echo suppression is clearly visible in this reception at Lake Mistissini of a 20 s key-down transmission at 5040 Hz on 5 June 1985 at 1408:32 UT. The measured two-hop delay from nearby whistlers is 4.2 s, which corresponds to the onset delay of the suppression effect.

large as 20 dB. Furthermore, when the second transmission terminates, the presence of the echoing transmission is not always observed, implying that the echoing signal is much weaker than the second signal. This effect is highly repeatable, occurring on approximately 30% of days when signals are detected, and coincides with conditions of strong echoing. As a result, transmission formats in later years of Siple operation regularly contain a 7 s gap before returning to a specific transmitted frequency [*Raghuram et al.*, 1977b; *Helliwell*, 1977; *Helliwell and Katsufrakis*, 1978; *Helliwell*, 1988a].

2.4.5 Entrainment

Another striking effect of the coherent wave instability is that of entrainment, shown in Figure 2.7. In this effect, a free-running emission is captured by a coherent transmission, such that the time-frequency character of the emission is subsequently controlled by the coherent wave. The entraining signal can be much less intense than the free running emission, while the initial triggering signal, triggered emission, and entrained emission all have approximately the same amplitude. The process occurs very
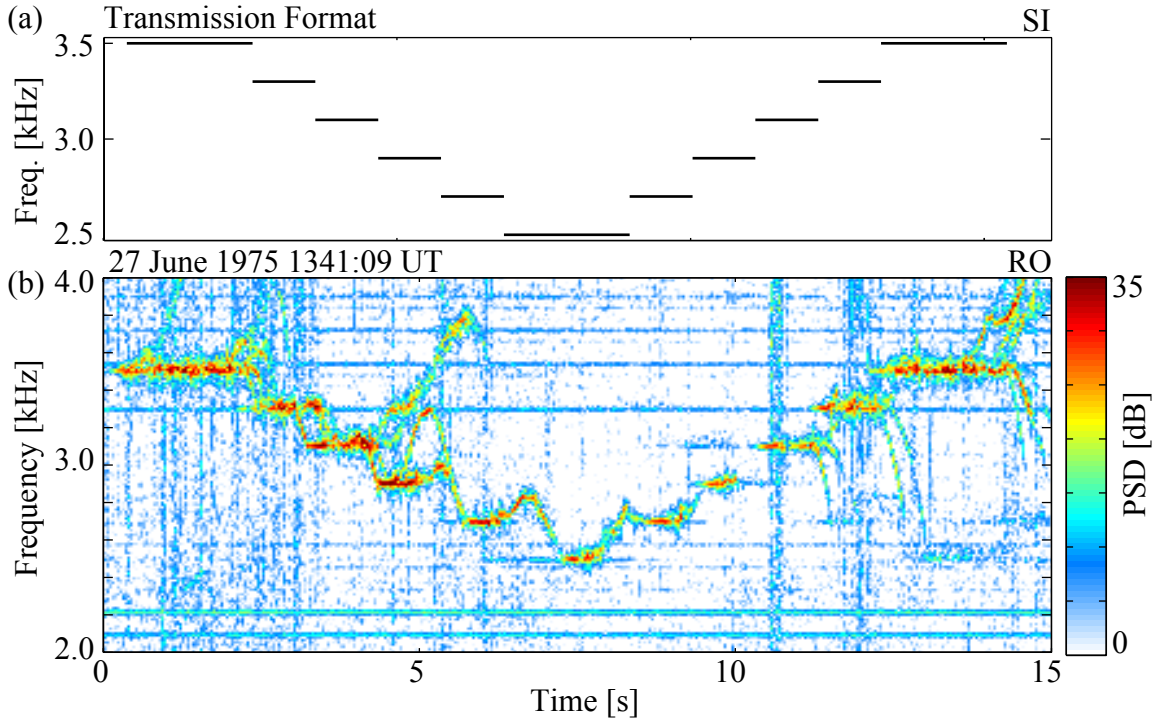
Figure 2.7: Entrainment. A case of entrainment from Roberval on 27 June 1975 at 1341:09 UT, in which a "stair-step" transmission from Siple starting at 3500 Hz triggered a free-running plasma plasma emission whose time-frequency characteristics are controlled by the subsequent constant frequency transmissions. (a) Time-frequency format transmitted by Siple. (b) Received signal at Roberval, showing characteristics of entrainment.

rapidly and is often associated with a band limited impulse upon entrainment. The triggering signal and the following entraining signal are always observed to have delay characteristics indicating that they both propagated along the same duct [*Helliwell et al.*, 1975; *Helliwell and Katsufrakis*, 1978; *Helliwell*, 1979b, 1988a].

The entrainment effect tends to prefer emissions that decrease in frequency with time. In fact, strong rising frequency emissions often cut-off the growth of constant frequency signals if their traces coincide in time and frequency [*Helliwell*, 1977, 1983a]. *Helliwell* [1979b] contains a tabulation of "allowed" spectral forms for entrainment. However, examples of entrainment in the literature are sparse, and no thorough study of the phenomenon has been conducted such that allowed (or disallowed) spectral forms can be stated conclusively.

2.4.6   Relation to Natural Activity

Due to the magnetospheric nature of many of the natural emissions in the VLF range, Siple transmissions frequently interact with these natural signals. Conditions with strong, multihop, lightning-generated whistlers often coincide with successful Siple transmissions to the northern hemisphere. In addition, whistlers are sometimes received at higher amplitudes in a frequency range above the range in which the nonlinear amplification associated with the instability is active. Under these conditions, whistlers can be received at the frequencies below the active band with amplitudes approximately 20 dB down compared to amplitudes within the active band. For example, if the active frequency range for the instability in a given case is 3.0 kHz to 3.5 kHz, the amplitude of a whistler in the frequency range greater than 3.5 kHz could be 20 dB higher than its amplitude in the frequency range less than 3.0 kHz [*Helliwell and Katsufrakis*, 1974; *Dowden et al.*, 1978].

In addition to serving as an indication of the natural condition at the time of signal reception, whistlers also interact with the nonlinearly amplified VLF signals in a more active manner. Similar to the effect of rising frequency emissions (see Section 2.4.5), a coincident whistler can terminate the process of nonlinear growth, returning the received signal to its initially received level [*Dowden et al.*, 1978]. Furthermore, whistler precursors (i.e. perturbations of the amplitude or phase of a magnetospherically propagating signal prior to two hop whistler reception) are occasionally observed on Siple signals [*Paschal*, 1990].

Siple signals are also observed on rare occasion to interact with bandlimited VLF noise, or hiss bands, in such a manner as to suppress the noise level in a band 50-200 Hz below the frequency of the transmitted signal. The bandwidth of this suppression tends to increase with the amplitude of the received signal at the transmitted frequency, developing 5-25 s after the initial reception of the Siple transmission. The intensity of the noise just below this "quiet band" is generally enhanced by this process. This phenomenon requires conditions of good echoing during periods of deep geomagnetic quieting, and lasts between 15 and 60 minutes [*Raghuram et al.*, 1977a; *Cornilleau-Wehrlin and Gendrin*, 1979].

Finally, reception of natural, discrete chorus emissions tends to reduce the likelihood of Siple reception in the northern hemisphere. Siple signals, when received concurrent with chorus, often show little or no nonlinear growth [*Helliwell*, 1977; *Helliwell and Katsufrakis*, 1978; *Helliwell*, 1983a].

2.4.7 *In Situ* Measurements of Siple Transmissions

Throughout the course of the Siple wave injection experiment, repeated attempts were made to observe Siple transmissions *in situ*, that is, to receive Siple signals using spaceborne VLF receivers. In fact, the first observations of Siple transmissions were recorded aboard the Explorer 45 spacecraft in June 1973. In addition to this case, Siple attempted transmissions to a number of spacecraft over the course of the experiment, including the DE-1 (US), DE-2 (US), ISEE-1 (US), ISIS-1 (Canada), ISIS-2 (Canada), S8l-I (US), ARCAD-1 (France/USSR), VIKING (US), EXOS-B (Japan), and Explorer 45 (US) spacecraft [*Carpenter*, 1982b; *Helliwell*, 1983b, 1986, 1988a].

Reception aboard a spacecraft provides a number of advantages, in that signals are observed in the midst of the amplification medium, potentially providing insight into the linear versus nonlinear contribution of the growth processes as well as confirming the spatial extent over which the instability is active (full disambiguation of the growth process would require multiple, closely spaced satellites to observe transmissions in the same duct). Unfortunately, no observations of Siple signals during ducted amplification exist. However, many successful receptions of Siple and other VLF transmitters in a non-ducted mode do exist, from various spacecraft receivers.

Signals are received onboard spacecraft over a wide range of L-values and latitudes and mostly within 30 degrees of the Siple meridian, with the strongest signal reception near the Siple meridian. Siple signals are also observed both inside and outside the plasmapause, always on non-ducted paths. Signals propagate to spacecraft along direct, non-ducted paths; along "hybrid" non-ducted paths, in which the Siple signal first propagates through a duct, reflects in the northern hemisphere, and is subsequently received by the spacecraft; and as signals believed to be leaking from

ducts [*Bell and Inan*, 1975; *Inan and Helliwell*, 1982; *Bell et al.*, 1983a; *Sonwalkar et al.*, 1984].

Received signal intensities vary from 0.01 pT up to 0.5 pT, with the strongest signals received outside the plasmapause and/or after crossing the magnetic equatorial plane. Direct signal intensities prior to crossing the magnetic equator are limited to amplitudes between 0.01 and 0.05 pT [*Inan et al.*, 1977; *Kimura et al.*, 1983; *Rastani et al.*, 1985; *Sonwalkar and Inan*, 1986].

Spectral Broadening

One of the effects noted on the ISIS-2 spacecraft above Siple is a significant spread ($\sim$10% of the transmitted frequency) in frequency of signals as they propagate upward. This "spectral broadening" effect occurs in the altitude range between 700 and 3800 km, with L-values between 2.3 and 11. This condition is coincident with lower hybrid noise as well as a band of VLF noise with an irregular cutoff frequency. Both of these effects may be associated with the precipitation of low energy (less than 1 keV) electrons. Spectral broadening is apparently caused by quasi-electrostatic waves generated in a linear mode conversion from upgoing whistler mode waves as the waves scatter from ionospheric density irregularities that are possibly caused by this precipitation. The mode conversion results in a large spread of k-vectors, with the resulting doppler shifted waves appearing as noise bands around the transmitted frequency, often in a characteristic "chevron" shape [*Bell et al.*, 1983b; *Helliwell*, 1985; *Inan and Bell*, 1985; *Bell and Ngo*, 1990].

Chapter 3

VLF Triggered Emission Theory

The VLF triggered emission instability is believed to be caused by ducted whistler mode waves that interact in an unstable manner with energetic electrons trapped in the Earth's radiation belts. The primary mode of interaction is one of cyclotron resonance, where the waves, traveling at speeds much lower than the free space speed of light, are able to coherently modify the energy and momentum of electrons that are streaming along the geomagnetic field line counter to the direction of wave propagation.

The predominant portion of the interaction between the waves and counter streaming electrons occurs near the geomagnetic equator, for two reasons. At the equator, the rate of change of the spatially varying, inhomogeneous geomagnetic field is minimized, and therefore the length of resonant interaction between the waves and electrons is maximized. Additionally, the inhomogeneity of the Earth's magnetic field causes the trajectories of all trapped electrons to pass through the magnetic equator, resulting in a higher density of electrons available for the interaction in the region around the equator, especially at higher pitch angles. This higher density of energetic electrons maximizes the availability of free energy to drive the instability [*Ashour-Abdalla*, 1972].

Because of the existence of guiding structures, or ducts, in the inner magnetosphere, the analysis of the interaction can be simplified significantly. Namely, the whistler mode wave can be assumed to be propagating only along the magnetic field,

which is defined to be the z direction. Therefore, all of the parameters of the interaction are varying only in the z direction, which allows the problem to be considered in only one spatial dimension.

Despite that simplifying assumption, the instability exhibits a number of nonlinear effects, such as the generation of off-frequency wave components and the preferential amplification of later portions of a wave packet. The nonlinear effects imply that the state of the energetic electrons is being modified as the instability progresses [*Das*, 1968], and that such modification is localized in phase space to the point where a full analytical treatment of the problem is impossible [*Matsumoto and Kimura*, 1971]. The analysis of the instability is additionally complicated by the presence of the inhomogeneity. Thus, any solution to the problem must necessarily address the evolution of the whistler simultaneously with the modification of the energetic electrons in an inhomogeneous medium.

The purpose of this chapter is to provide a detailed discussion of the previous analytical theory developed to treat the VLF triggered emission instability. The discussion starts from first principles, assuming some familiarity with electromagnetics and plasma physics. The discussion first covers the characteristics of the whistler mode waves that activate and are produced by the instability. The chapter then explains the concepts necessary for an understanding of the mechanism by which the waves modify the state of the energetic electrons and, in turn, how the state of the energetic electrons modifies the properties of the waves, including an illustrative discussion of nonlinear effects that cannot be analytically treated. The chapter concludes with a summary of the potential mechanisms for saturation of the instability-driven growth that have been proposed by others, or are apparent from the preceding theoretical discussion.

3.1   Linear Plasma Theory and the Whistler Mode Instability

Linear plasma theory predicts an instability in the whistler mode. Specifically under certain conditions, whistler mode waves can experience an unstable interaction with energetic electrons such that the waves are amplified exponentially. Here, the properties of this interaction, as considered by linear plasma theory, are derived mathematically, along with a discussion of the physical interaction involved.

3.1.1   Parallel Propagating Whistler Mode Waves

Whistler mode waves exist in cold, magnetized plasmas, where spatial variations can be neglected on the scale of a wavelength. In particular, the angular frequency of the wave ω is less than the angular electron cyclotron frequency, $\omega_{\mathrm{c}} = qB_0/m_{\mathrm{e}}$, where q is the fundamental charge, m_{e} is the rest mass of the electron, and B_0 is the strength of the static magnetic field. Additionally, the wave frequency is high enough above the ion cyclotron frequency that the motion of the ions in the plasma may be neglected.

Under these conditions, with a ducted wave that is propagating along the magnetic field line, the dispersion relation for the whistler mode wave, which is a right-hand circularly polarized wave (see Appendix A), is given by

$$\frac{c^2 k^2}{\omega^2} = 1 + \frac{\omega_{\mathrm{p}}^2}{\omega\left(\omega_{\mathrm{c}} - \omega\right)} \simeq \frac{\omega_{\mathrm{p}}^2}{\omega\left(\omega_{\mathrm{c}} - \omega\right)} \tag{3.1}$$

where c is the free space speed of light, k is the wavenumber, and

$$\omega_{\mathrm{p}} = \sqrt{\frac{q^2 n_{\mathrm{c}}}{m_e \epsilon_0}}$$

is the plasma frequency, where n_{c} is the number density of cold electrons and ϵ_0 is the permittivity of free space.

The left hand side of Equation (3.1) is the square of the index of refraction of the medium, which represents the ratio of the free space speed of light to the phase

velocity of the wave. The phase and group velocities of the wave are given by

$$v_{\mathrm{p}} = \frac{\omega}{k} = c\frac{\omega^{1/2}\left(\omega_{\mathrm{c}} - \omega\right)^{1/2}}{\omega_{\mathrm{p}}} = \frac{k\omega_{\mathrm{c}}}{k^2 + \dfrac{\omega_{\mathrm{p}}^2}{c^2}} \tag{3.2}$$

$$v_{\mathrm{g}} = \frac{\partial\omega}{\partial k} = 2c\frac{\omega^{1/2}\left(\omega_{\mathrm{c}} - \omega\right)^{3/2}}{\omega_{\mathrm{c}}\omega_{\mathrm{p}}} = \frac{2k\omega_{\mathrm{c}}\dfrac{\omega_{\mathrm{p}}^2}{c^2}}{\left(k^2 + \dfrac{\omega_{\mathrm{p}}^2}{c^2}\right)^2} = \frac{2k\left(\omega_{\mathrm{c}} - \omega\right)}{k^2 + \dfrac{\omega_{\mathrm{p}}^2}{c^2}} \tag{3.3}$$

respectively, where expressions have been provided for the velocities in multiple forms for convenience. Typically for conditions in the inner magnetosphere, the index of refraction of the medium is much greater than unity. Correspondingly, the propagation velocities for whistler mode waves are significantly slower than the speed of light, so that whistler mode waves are "slow" in nature.

The dispersion relation in Equation (3.1) is derived assuming that the plasma is composed of particles with no thermal energy. The condition that causes the VLF triggered emission instability involves the interaction between the whistler mode waves and energetic radiation belt electrons. The remainder of this discussion focuses on this interaction.

3.1.2   Cyclotron Resonance

In general for energetic electrons, the component of the electron velocity along the magnetic field line is non-zero. That is, $v_{\parallel} \neq 0$. Additionally, the sense of rotation for the electron gyromotion is the same as the sense of rotation of the electric and magnetic fields of the right-handed whistler mode wave. In the condition shown schematically in Figure 3.1, where the energetic electrons stream along the magnetic field in one direction and a whistler mode wave propagates along the field line in the opposite direction, there exists a parallel electron velocity v_{res}, such that the wave frequency doppler-shifted into the frame of the electron is the same as the electron

gyrofrequency. That is, the condition

$$\omega_\mathrm{c} - kv_{||} \simeq \omega$$

is satisfied when $v_{||} = v_\mathrm{res}$, or

$$v_\mathrm{res} = \frac{\omega_\mathrm{c} - \omega}{k}.$$

This condition is known as cyclotron resonance. Under cyclotron resonance, the wave fields and the electron velocity vector are stationary in gyrophase with respect to one another. In this configuration, the electron experiences the action of the wave field over a long period of time (i.e., over the period that the resonance condition is satisfied), and its velocity can be significantly modified. When this occurs with some resonant fraction of the energetic electron population, the energies and momenta of the resonant electrons are changed. Specifically, currents are created in the plasma such that it becomes damping or amplifying to the whistler mode wave. If this interaction causes a significant change in the state of the plasma over large spatial distances or long periods of time, nonlinear effects, such as those involved in the VLF triggered emission instability can develop.

3.1.3   The Energetic Electron Distribution Function

To study the interaction between the energetic electrons and a parallel propagating whistler mode wave, a method of specifying the state of the energetic electrons is required. If the interaction with only one electron is considered, the specification of the electron state is straightforward, as shown in Figure 3.2(a). In this case, only the electron position $\mathbf{r}$ and the electron velocity $\mathbf{v}$ must be specified. As the interaction with the whistler mode wave evolves, the state of the electron can be fully tracked by considering how $\mathbf{r}$ and $\mathbf{v}$ change as a result of the interaction.

In actuality, however, many energetic particles are involved in the interaction with a single wave packet, making the specification of their individual position and velocity vectors burdensome. In that case, the state of all of the energetic electrons is specified

Direction of
Wave Propagation

Direction of
Electron Motion

Sense of Wave Field
Rotation to a
Stationary Observer

Sense of
Electron
Gyration

B₀

z
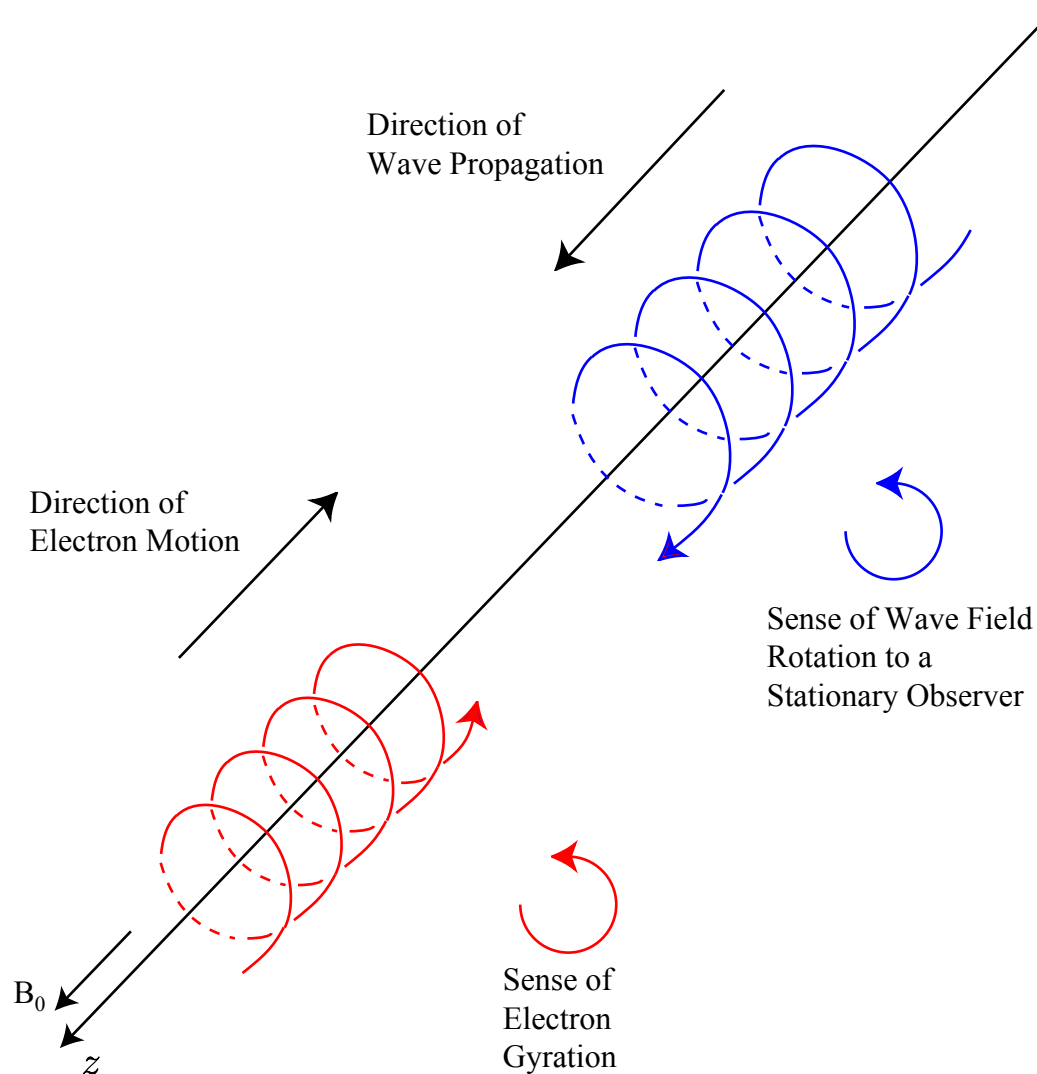
Figure 3.1: A schematic sketch of cyclotron resonance, in which a counter-streaming electron interacts with a whistler-mode wave propagating along the magnetic field direction. The sense of electron gyration and wave field rotation is the same. Adapted from [*Walt*, 1994, fig. 7.4].
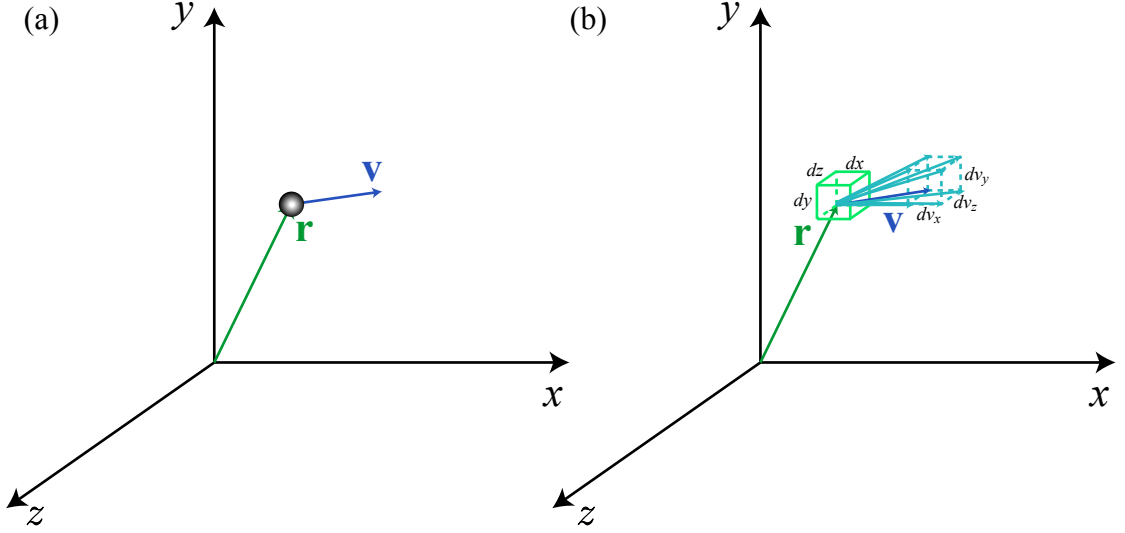
Figure 3.2: The concept of the energetic electron distribution function. (a) The specification of a single electron, with one position vector $\mathbf{r}$ and one velocity vector $\mathbf{v}$. (b) The specification of many electrons as a density of electrons in a small volume around $\mathbf{r}$ and a small range of velocities about $\mathbf{v}$.

as a density of electrons, in a small volume around the position described by $\mathbf{r}$ and a small range of velocities about the velocity vector $\mathbf{v}$, as shown schematically in Figure 3.2(b) for a single $(\mathbf{r}, \mathbf{v})$ combination. This density of electrons as a function of position and velocity is called the energetic electron distribution function, $F(\mathbf{r}, \mathbf{v})$ [*Gurnett and Bhattacharjee*, 2005, Sec. 5.1].

In the ducted case considered here, the whistler mode wave propagates only in the direction of the magnetic field z, so the variation in the electron distribution function and other parameters of the problem are only along the z direction. Correspondingly, the electron distribution function needs to be specified only as a function of position z, rather than with the full position vector $\mathbf{r}$, i.e., $F(\mathbf{r}, \mathbf{v}) \rightarrow F(z, \mathbf{v})$.

There are many ways of ascribing certain properties to the distribution function, but attention is called specifically to the condition of anisotropy because of its role in the whistler mode wave-particle interaction. Anisotropy and its opposite condition, isotropy, are illustrated in Figure 3.3. In the case of an anisotropic distribution function, the state of the energetic electrons is such that a certain direction in velocity
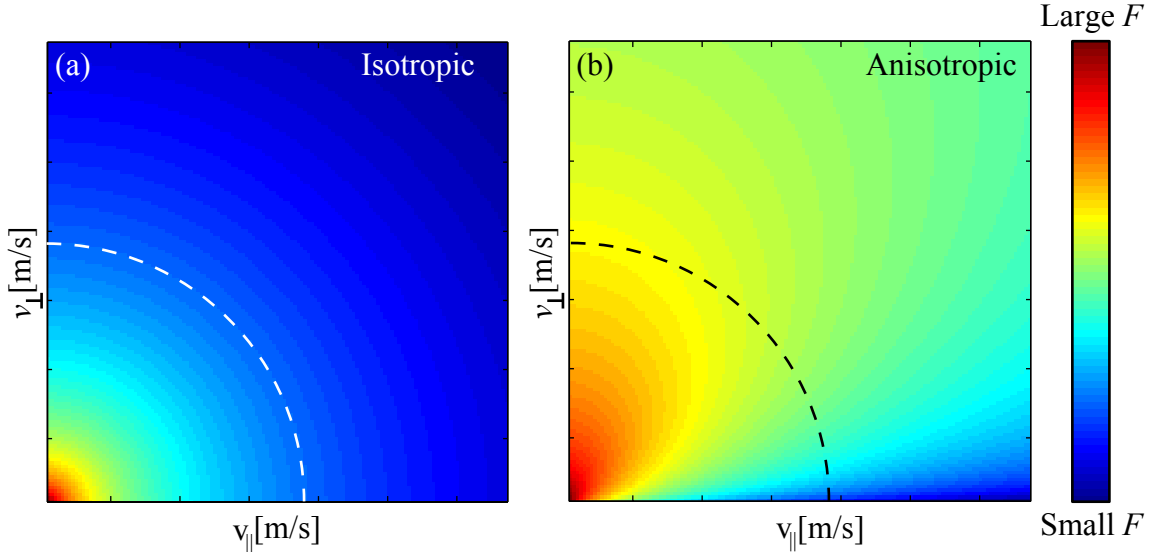
Figure 3.3: The concept of an anisotropic distribution function. (a) An isotropic distribution function, where the value of the distribution function does not vary along a constant velocity contour. (b) An anisotropic distribution function, where the value of the distribution function varies along a constant velocity contour; that is, the value of the distribution function depends on the direction of the electron velocity. The distribution function is peaked in the direction perpendicular to the magnetic field direction, which corresponds to a value of $A > 0$ in Equation (3.10).

space is preferred. In contrast, an isotropic distribution function is uniform as a function of the direction of electron motion. To illustrate, Figure 3.3(a) and Figure 3.3(b) both show a contour of constant velocity laid down in velocity space, represented by a dashed line. In the case of an isotropic distribution function, Figure 3.3(a), the value of the distribution function does not change along the constant velocity contour. Specifically, the value of the distribution function in the figure is represented by the color of the image; at all points along the contour in Figure 3.3(a) the color in the image (and corresponding value of the distribution function) is the same. In contrast, the value of an anisotropic distribution function, shown by the color of the image in Figure 3.3(b), changes as a function of position along the contour.

In the magnetosphere, the energetic electron distribution is often anisotropic, a condition which is aided by the configuration of the Earth's magnetic field. Specifically, energetic electrons that are trapped by the geomagnetic field execute a bounce

motion as they stream along the magnetic field line. That is, the increasing strength of the Earth's magnetic field seen by a electron as it travels toward the Earth along a field line (see Figure 1.1) causes the electron to slow and eventually reverse its direction of motion such that it travels back toward the magnetic equator. The altitude at which this "mirroring" occurs is dependent on the angle between the electron velocity vector and the magnetic field line, called the "pitch angle" α where

$$\alpha = \tan^{-1} \frac{v_\perp}{v_\|}. \tag{3.4}$$

Electrons with velocity vectors perfectly aligned with the magnetic field ($\alpha = 0$) never mirror, while electrons at the magnetic equator with velocity vectors entirely perpendicular to the magnetic field ($\alpha = \pi/2$) always remain at the equator. However, electrons that mirror at low altitudes are more likely to interact with the neutral atmosphere and not return to the magnetosphere. Thus, there exists an intermediate pitch angle α_{lc} that corresponds to the minimum angle that electrons at the equator must have such that their mirror heights are above the dense atmosphere. The cone of trajectories described by $\alpha < \alpha_{lc}$ is known as the "loss cone." As a result of this effect, the density of energetic electrons is often very low in the loss cone, and the distribution function is equivalently anisotropic.

In principle, the presence of such a loss cone may simply result in a distribution which is devoid of elelectrons for $\alpha < \alpha_{lc}$, but which is nevertheless constant for $\alpha > \alpha_{lc}$. While such a distribution is anisotropic, the anisotropy is at relatively small pitch angles (e.g., $\alpha_{lc} \simeq 7°$ at $L=4$), for which the prependicular energies of electrons are relatively small and thus may not contribute significantly to wave growth (see Section 3.5). However, because the distribution function is under equilibrium driven by a constant diffusion of electrons from higher to lower pitch angles [*Kennel and Petschek*, 1966], the actual distribution function is significnatly more anisotropic, as described in Section 2.3.4, with density increasing with increasing pitch angle, for $\alpha > \alpha_{lc}$.

3.1.4 The Linear Growth Rate

As shown in Appendix A, the addition of a plasma component with thermal energy introduces an imaginary part to the wave dispersion relation. In general, the consequence of this result is that the plasma becomes either damping or amplifying. To mathematically quantify the amount of damping or amplification, the wave frequency is considered to have real and imaginary components, i.e., $\omega \to \omega_r + i\gamma$, where ω_r and γ are purely real quantities. If the small growth rate approximation is made, $|\gamma| \ll |\omega_r|$, a Taylor series expansion of the dispersion relation, $D(k, \omega_r + i\gamma)$, can be made about ω_r, yielding,

$$D(k, \omega_r + i\gamma) = D(k, \omega_r) + i\left[\frac{\partial D(k, \omega)}{\partial \omega_r}\right]\gamma = 0 \tag{3.5}$$

Expressing the first term in Equation (3.5) in terms of its real and imaginary parts,

$$D(k, \omega) = D_r(k, \omega) + iD_i(k, \omega),$$

and neglecting the small imaginary part of the second term, the real and imaginary parts of Equation (3.5) give the relations

$$D_r = 0 \tag{3.6}$$

$$\gamma = \frac{-D_i}{\partial D_r/\partial \omega} \tag{3.7}$$

Equation (3.6) gives the expression for the real component of the dispersion relation, which controls the relationship between the wave frequency ω and wavenumber k and hence the corresponding propagation characteristics of the wave. Equation (3.7), on the other hand, gives an expression for the exponential temporal growth rate. If the value of γ in Equation (3.7) is negative, then the wave is damped by the plasma; if it is positive, then the wave grows exponentially.

From Equation (A.16), the real and imaginary parts of the dispersion relation can

be expressed as

$$D_{\mathrm{r}}(k,\omega) = \frac{c^2 k^2}{\omega^2} - \frac{\omega_{\mathrm{p}}^2}{\omega\left(\omega_{\mathrm{c}} - \omega\right)} \tag{3.8}$$

$$D_{\mathrm{i}}(k,\omega) = \pi \frac{\omega_{\mathrm{p}}^2}{\omega^2 n_{\mathrm{c}}} \left[\frac{\omega}{k} \int_0^\infty 2\pi v_\perp F dv_\perp - \int_0^\infty \left(v_{||}\frac{\partial F}{\partial v_\perp} - v_\perp \frac{\partial F}{\partial v_{||}}\right) \pi v_\perp^2 dv_\perp\right]_{v_{||}=v_{\mathrm{res}}} \tag{3.9}$$

where the first term in the integral in Equation (A.16) has been integrated once by parts, and the subscript h has been discarded with the understanding that the quantity F describes the distribution function for the energetic electrons. From Equations (3.8), (3.9), and (3.7), and noting

$$\left(v_{||}\frac{\partial F}{\partial v_\perp} - v_\perp \frac{\partial F}{\partial v_{||}}\right) = \frac{\partial F}{\partial \alpha}.$$

A convenient representation of the growth rate can be derived [*Kennel and Petschek*, 1966]

$$\gamma = \pi \frac{\omega_{\mathrm{c}}}{n_{\mathrm{c}}} \left(1 - \frac{\omega}{\omega_{\mathrm{c}}}\right)^2 v_{\mathrm{res}} \left[A - \frac{\omega}{\omega_{\mathrm{c}} - \omega}\right] \eta \tag{3.10}$$

where

$$\eta = \int_0^\infty F \, 2\pi v_\perp dv_\perp \bigg|_{v_{||}=v_{\mathrm{res}}}$$

$$A = \frac{\int_0^\infty \dfrac{\partial F}{\partial \alpha} \tan\alpha \, v_\perp dv_\perp}{\eta} \bigg|_{v_{||}=v_{\mathrm{res}}}$$

with α and $\partial F/\partial\alpha$ evaluated only at the resonant velocty, $v_{||} = v_{\mathrm{res}}$. The term η relates to the magnitude of the resonant flux, and the term A is a measure of the anisotropy of the energetic electron distribution function. Noting that η is always positive, the bracketed term in Equation (3.10) controls the sign of the growth rate.

That is, if the condition

$$\omega < \frac{A}{1 + A} \omega_\text{c}$$

is satisfied, then the plasma is unstable to the whistler mode wave of frequency ω. Equivalently stated, the plasma is unstable to whistler mode waves if it is sufficiently anisotropic.

An examination of Equation (3.10) reveals that two sources of free energy exist to drive growth in the whistler mode. Specifically, with the wave frequency, gyrofrequency, and cold electron density given, the only parameter that affect the magnitude of the growth rate are η, and A, which are respectively related to the flux of resonant particles and the anisotropy of the distribution function. The effect of η on wave growth is easy to visualize; a higher electon flux directly relates to an increase in energy available for the interaction. The effect of the anisotropy, however, is more subtle. Knowing that the isotropic Maxwellian distribution represents the equilibrium state of a plasma [*Bittencourt*, 2004, Sec. 7.1.3], an anisotropic distribution function must necessarily contain free energy due to its configuration. A whistler mode wave taps this free energy by reducing the anisotropy of the distribution fuction, on average, through the cyclotron resonant interaction.

3.2 Single Particle Dynamics

The above discussion considers the interaction of the entire electron population with the whistler mode wave. An equivalent treatment is to consider the interaction of each individual electron with the wave and sum the results to obtain the total effect. This method is attractive in that the effect of the wave on the individual electron is easily described by the Lorentz force. The coordinate system in which the interaction is derived is shown in Figure 3.4.

The only spatial coordinate considered is z, which is the distance along the magnetic field line and the direction of ducted wave propagation. The electron velocity vector, $\mathbf{v}$ is decomposed into three components $(v_{||}, v_{\perp}, \zeta)$, corresponding to the velocity of the electron parallel to the static magnetic field direction, the velocity of
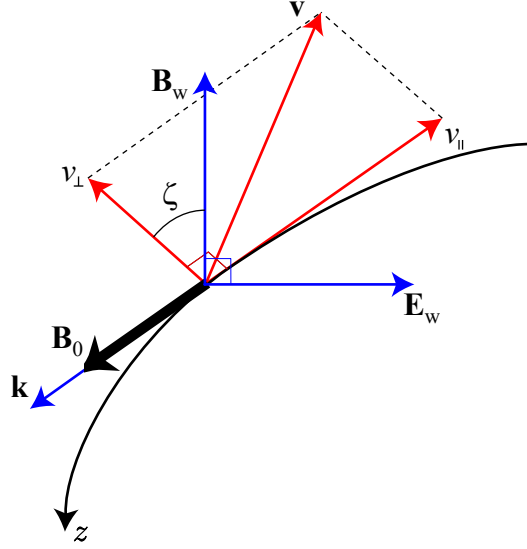
Figure 3.4: The coordinate system used to model the interaction. The whistler mode wave is propagating along the magnetic field line in the $+z$ direction, which corresponds to the magnetic field direction, while the electron is streaming along the field line in the $-z$ direction. The electron velocity vector is decomposed into components perpendicular and parallel to the magnetic field, with a gyrophase angle described between the perpendicular velocity vector and the wave magnetic field vector.

the electron perpendicular to the static magnetic direction, and the gyrophase angle between the perpendicular velocity vector and the wave magnetic field vector, respectively. With a direct application of the Lorentz force, the particle scattering equations are easily derived as (see Appendix B)

$$\frac{dz}{dt} = -v_{\parallel} \tag{3.11}$$

$$\frac{dv_{\parallel}}{dt} = -\frac{q}{m_e} B_{\mathrm{w}} v_{\perp} \sin \zeta \tag{3.12}$$

$$\frac{dv_{\perp}}{dt} = \frac{q}{m_e} B_{\mathrm{w}} \left(\frac{\omega}{k} + v_{\parallel} \right) \sin \zeta \tag{3.13}$$

$$\frac{d\zeta}{dt} = \left[\omega_{\mathrm{c}} - \omega - k v_{\parallel} \right] + \frac{q}{m_e} B_{\mathrm{w}} \frac{\left(\dfrac{\omega}{k} + v_{\parallel} \right)}{v_{\perp}} \cos \zeta \tag{3.14}$$

One aspect to note from Equations (3.11) through (3.14) is that the modification of

the electron velocity is sinusoidally dependent on the the gyrophase angle ζ. However, the gyrophase angle ζ is rapidly varying unless the term in square brackets in Equation (3.14) is small. In other words, the relation

$$\omega_{\text{c}} - kv_{||} \simeq \omega$$

must be satisfied; the particle must be nearly in cyclotron resonance with the wave.

The ratio of Equation (3.13) and Equation (3.12) yields the relation [Das, 1968]

$$\frac{dv_\perp}{dv_{||}} = -\frac{\dfrac{\omega}{k} + v_{||}}{v_\perp}$$

which describes a set of contours in velocity space along which the electrons are moved as a result of the interaction with the whistler mode wave. These contours are sketched in Figure 3.5, for a particular value of the ratio ω/k. For comparison, contours of constant energy (i.e., constant velocity) are plotted as well. This plot provides some intuition concerning the energetics of the whistler mode interaction. Namely, if the electron parallel velocity is increased as a result of the interaction, the electron loses energy to the wave. However, decreasing the electron parallel velocity requires the wave to do work on the electron [$Brice$, 1964].

In the case of a brief linear interaction between the wave and the near-resonant electrons, the initial gyrophase is uniformly distributed (see Equation (A.10)), resulting in an effectively uniform scattering of particles up and down in parallel velocity. However, if the distribution is initially anisotropic, with $A > 0$, then more electrons are scattered up in parallel velocity (down in energy) than the opposite, and the wave is amplified by the interaction. In effect, this observation is a restatement that an anisotropic distribution function is necessary for unstable wave amplification.

The above derivation implicitly assumes a homogeneous magnetic field. The presence of gradients in the magnetic field introduces an additional force on the electrons
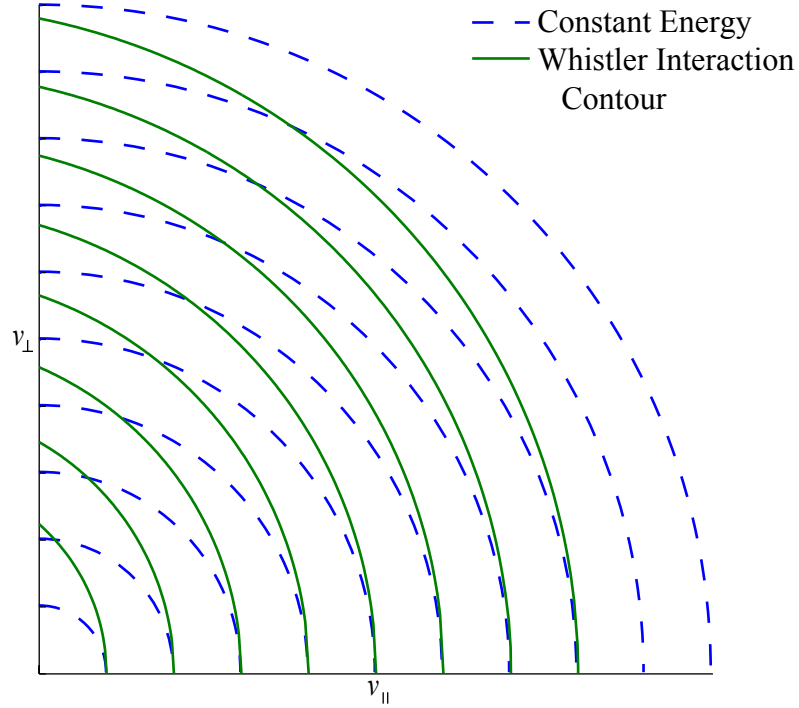
Figure 3.5: Velocity space contours for energetic electrons interacting with whistler mode waves. For comparison, constant energy contours are plotted. An interaction that causes a net decrease in electron parallel velocity results in a net energization of the electrons. Conversely, an interaction that increases the parallel velocity of the electrons results in a loss of energy from the electrons.

that must be included in the equations of motion. Including the effect of the inhomogeneous magnetic field produces the complete equations of motion (see Appendix B)

$$\frac{dz}{dt} = -v_{||} \tag{3.15}$$

$$\frac{dv_{||}}{dt} = -\frac{q}{m_e} B_{\mathrm{w}} v_\perp \sin \zeta - \frac{v_\perp^2}{2\omega_{\mathrm{c}}} \frac{\partial \omega_{\mathrm{c}}}{\partial z} \tag{3.16}$$

$$\frac{dv_\perp}{dt} = \frac{q}{m_e} B_{\mathrm{w}} \left(\frac{\omega}{k} + v_{||} \right) \sin \zeta + \frac{v_{||} v_\perp}{2\omega_{\mathrm{c}}} \frac{\partial \omega_{\mathrm{c}}}{\partial z} \tag{3.17}$$

$$\frac{d\zeta}{dt} = \omega_c - \omega - k v_{||} + \frac{q}{m_e} B_{\mathrm{w}} \frac{\left(\frac{\omega}{k} + v_{||} \right)}{v_\perp} \cos \zeta \tag{3.18}$$

With the wave absent ($B_{\mathrm{w}} = 0$), these equations describe the adiabatic mirroring motion of the energetic particles along the field line, with $v_{||}$ increasing as particles move toward regions of increasing magnetic field strength (i.e., increasing ω_{c}).

3.3   Whistler Mode Waves in the Presence of a Resonant Current

The net effect of the cyclotron resonant interaction of the whistler mode wave and the energetic electrons is the creation of currents in the plasma. These currents are in addition to the natural current that forms in the cold plasma due to the presence of the whistler mode wave, and are composed of energetic electrons whose distribution function has been modified by the interaction. To distinguish them from the cold plasma currents, the currents generated through the cyclotron resonance interaction are hereafter called resonant currents.

3.3.1   Resonant Current Formation

Figure 3.6 conceptually shows, via the interaction with the whistler mode wave, the creation of a resonant current in the plasma. The force on the electron parallel to the magnetic field comes from the $\mathbf{v} \times \mathbf{B}_{\mathrm{w}}$ term in the Lorentz force, which causes the parallel velocity of near resonant electrons to oscillate around the resonant velocity, such that their perpendicular velocity vectors oscillate around the direction antiparallel to the wave magnetic field [*Brice*, 1963]. This behavior, in turn, creates a preferred gyrophase for resonant electrons, and the resulting phase bunching effect creates a helical current formed by the energetic electrons, as shown in Figure 3.6(b) [*Helliwell and Crystal*, 1973]. Note that this bunching effect occurs only for near-resonant electrons, and the bulk of the plasma remains unperturbed (stated mathematically, $\mathbf{k} \cdot \mathbf{E}_w = 0$ for whistler mode waves, so that no longitudinal perturbations in the bulk electron density are present).

The picture presented in Figure 3.6 is illustrative in conceptually introducing the
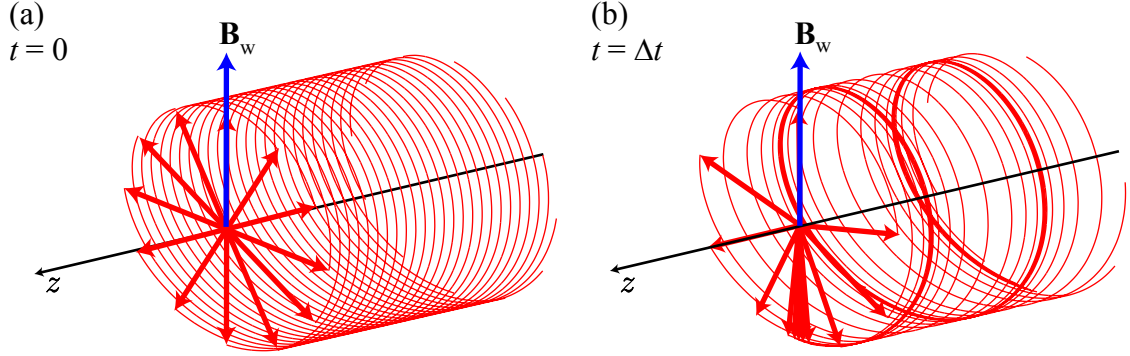
Figure 3.6: Illustration of current formation. (a) Twelve electrons with a single parallel velocity are shown, with their perpendicular velocity vectors uniformly distributed in gyrophase. (b) After some time of interaction Δt, the electrons are no longer uniform in gyrophase; the bunching of electrons acts to create a helical current in the plasma which can amplify or damp the wave. Adapted from *Helliwell and Crystal* [1973]

phase bunched current formation, however it should not be taken as a universal picture. In particular, Figure 3.6 only presents the case for a single velocity of exactly resonant electrons. In practice, the energetic electron distribution contains a range of electron velocities about resonance, such that the less dense regions created by the bunching effect in Figure 3.6(b) would be filled in with electrons that started at different initial parallel velocities. Thus, the detailed nature of the resonant current formation is determined by velocity space gradients in the distribution function [*Dysthe*, 1971].

An alternative, more complete view of resonant current formation can be generated using Equation (A.12), derived in Appendix A, which expresses the linear modification of the distribution function due to the interaction with the whistler mode wave:

$$F_{\mathrm{w}}(v_{||}, v_{\perp}, \zeta) = \frac{qn_{\mathrm{c}}}{m_{\mathrm{e}}\omega_{\mathrm{c}}} \frac{E_{\mathrm{w}}}{2} \left[\frac{-\sin\zeta}{\left(\dfrac{kv_{||} + \omega}{\omega_{\mathrm{c}}}\right) - 1} \right] \left\{ \frac{\partial F_0}{\partial v_{\perp}} + \frac{k}{\omega} \frac{\partial F_0}{\partial \alpha} \right\}$$

In particular, the ζ-variation of the perturbed component of the distribution function, in brackets, represents the phase-bunching effect across a range of velocities near resonance. Furthermore, the term in curly braces determines the sign of the current

generation; distributions with enough anisotropy are bunched in the range of gyrophase that drives growth, while isotropic distributions are bunched in the opposite range of gyrophase.

Figure 3.7 shows the deformation of the distribution function predicted by linear plasma theory, i.e., the value of the perturbed distribution function F_w (see Appendix A) projected onto the $v_{||} - \zeta$ plane, centered at the resonant velocity v_res. In Figure 3.7, the blue color represents a depression (with respect to the initial unperturbed balue) in the value of the electron distribution function, while the red color represents an enhancement. This enhancement or depression as a function of gyrophase results in the formation of a resonant current in the direction of the electron depression. The two cases shown represent the opposite conditions of damping and growth, where Figure 3.7(a) represents the distribution function in the case of linear damping, and Figure 3.7(b) represents the distribution function in the case of linear growth. Since F_w represents the value of the electron distribution, the resonant current is formed pointing in the direction of depressions in the electron density and away from enhancements in the electron density. In the damping case in Figure 3.7(a), the resonant current is formed about the phase angle $3\pi/2$ (i.e., current is directly parallel to the wave electric field), while in the growth case in Figure 3.7(b) the resonant current is formed about the phase angle $\pi/2$ (current is antiparallel to the wave electric field). It is not surprising to note that in the case of damping, $\mathbf{J}_\mathrm{r} \cdot \mathbf{E}_w > 0$, while in the case of growth, $\mathbf{J}_\mathrm{r} \cdot \mathbf{E}_w < 0$.

3.3.2   Modification of the Wave by the Resonant Current

The presence of a whistler mode wave and an energetic electron plasma creates resonant currents that act to amplify or damp the wave. The first-order effect of these resonant currents on the waves can be derived from Maxwell's equations and the expression for the cold current carried by the whistler mode wave, as shown in Appendix C. The derivation neglects the effect of the displacement current because of the high index of refraction of the whistler mode, assumes that the inhomogeneity of the medium is slowly varying on wavelength scales, and assumes that the wave
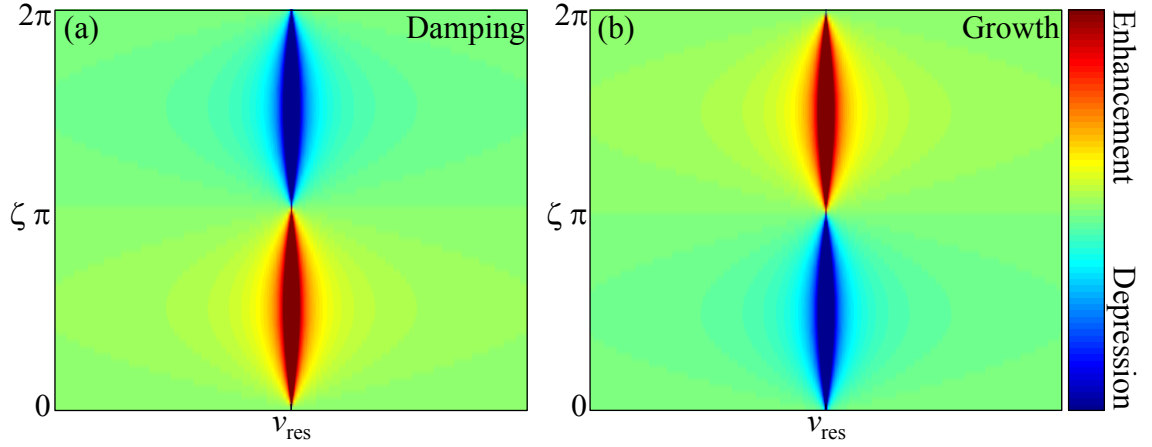
Figure 3.7: The perturbed component of the distribution function, projected onto the $v_\| - \zeta$ plane. (a) Isotropic case, where damping is predicted. (b) Anisotropic case, where growth is predicted.

is essentially monochromatic. Under these conditions, Equation (C.9) and Equation (C.12) describe the evolution of the wave amplitude and phase caused by the resonant current,

$$\frac{\partial B_\mathrm{w}}{\partial t} + v_\mathrm{g}\frac{\partial B_\mathrm{w}}{\partial z} = -\frac{\mu_0}{2}v_\mathrm{g}J_\mathrm{E} \tag{C.9}$$

$$\frac{\partial \phi}{\partial t} + v_\mathrm{g}\frac{\partial \phi}{\partial z} = -\frac{\mu_0}{2}v_\mathrm{g}\frac{J_\mathrm{B}}{B_\mathrm{w}} \tag{C.12}$$

where B_w is the amplitude of the interacting wave and ϕ is the additional phase accumulated by the wave during the interaction.

An examination of these wave equations reveals the picture shown in Figure 3.8, where the resonant current J_r is decomposed into J_E and J_B, components in the direction of the wave electric field and magnetic field, respectively. As a more general condition than the linear growth described above (see Section 3.3.1), the wave grows when the resonant current vector has any component antiparallel to the wave electric field, that is, $0 < \psi < \pi$, where ψ is the angle between the resonant current vector and the wave magnetic field. Wave damping occurs when the resonant current vector is in the opposite range. In addition, the phase of the wave advances when $\pi/2 < \psi < 3\pi/2$.
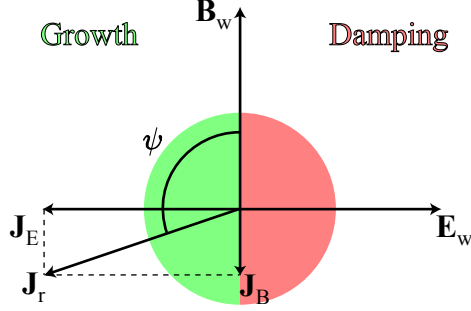
Figure 3.8: Resonant current configuration. The resonant current is divided into two components, J_E parallel to the wave electric field and J_B parallel to the wave magnetic field. Growth occurs when the angle between the resonant current vector and the wave magnetic field vector is between 0 and π.

In the case of linear growth above, the quantity ψ can take on the values of $\pi/2$ or $3\pi/2$, depending on the gradients in the distribution function. These values of ψ correspond to pure growth or damping, respectively, with no phase modification. The more general definition of growth from Equation (C.9) and Equation (C.12) allows for the additional effect of phase advance or retardation to occur during the wave-particle interaction, which is a purely nonlinear component of the interaction.

Finally, the growth rate calculated from resonant currents generated from the interaction (as opposed to velocity-space gradients in the electron distribution function, as in Equation (3.10)) can be determined from Equation (C.9), using [*Omura and Matsumoto*, 1982]

$$\gamma_{\mathrm{nl}} = \frac{1}{B_{\mathrm{w}}} \frac{dB_{\mathrm{w}}}{dt} = -\frac{\mu_0}{2} v_{\mathrm{g}} \frac{J_E}{B_{\mathrm{w}}} \qquad (3.19)$$

3.4 Particle Trapping

The chief effect that is believed to contribute to the nonlinear aspects of the VLF triggered emission instability (e.g., temporal growth, triggering of emissions, generation of sidebands, etc.) is particle trapping, where electrons become trapped in the wave potential well and remain in resonance with the wave over long distances. Consider
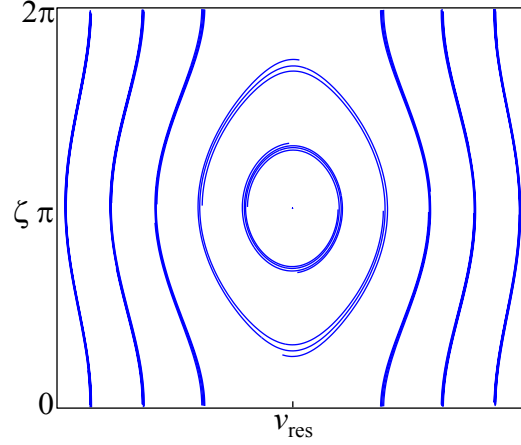
Figure 3.9: Particle trap formed by the interaction between the coherent wave and resonant electrons. Two populations of electrons are formed: the gyrophase ζ of particles near resonance is limited to a fixed range of values, representing a set of particles trapped by the wave potential; the gyrophase of untrapped particles rotates freely in ζ.

the equations of motion, Equations (3.12), (3.13), and (3.14), which define the interaction between individual electrons and the whistler mode wave in a homogeneous medium. Applying these equations and tracing electron trajectories in the $v_{||}$–ζ plane yields the picture in Figure 3.9, where there are clearly two separate populations of electrons. The set of electrons whose parallel velocity is near the resonant velocity are trapped by the wave potential, and their phase is constrained to a fixed range of values. In contrast, particles well away from resonance are untrapped; the phase of these electrons can take on any value during the interaction.

This picture, however, does not take into account the effect of the inhomogeneity of the geomagnetic field, which tends to ensure that the interaction time between a resonant particle and the wave is relatively brief. That is, considering the resonance condition

$$\omega_c - kv_{||} \simeq \omega,$$

ω_c, k, and $v_{||}$ all change as a function of distance along the field line due to the changing magnetic field. This tendency to disrupt resonance is illustrated in Figure 3.10, where the resonant velocity as a function of position along the field line $v_{res}(z)$ is
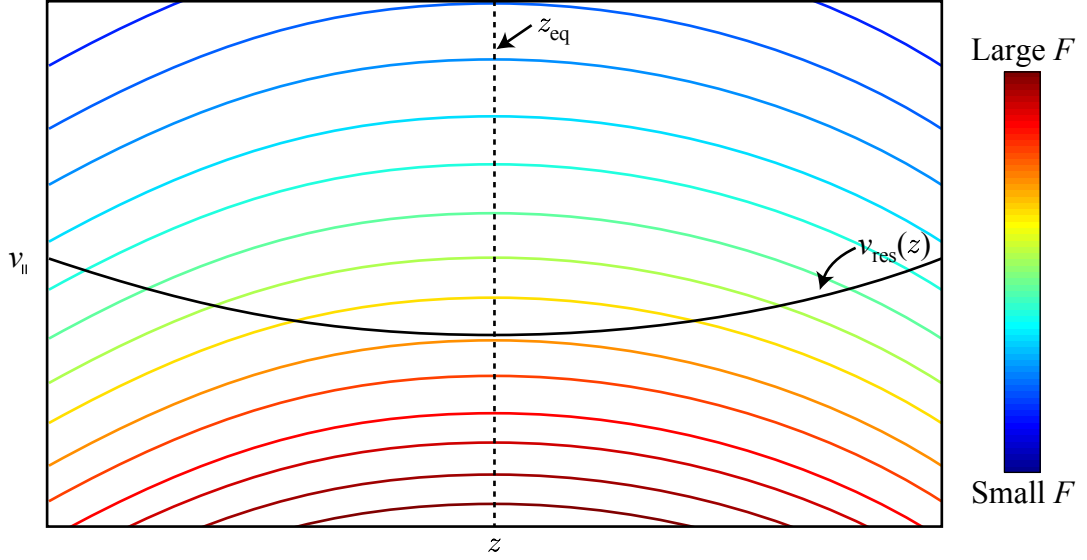
Figure 3.10: The variation in resonance velocity as a function of position along the dipole field line z, centered about the magnetic equator, for a constant frequency wave. Shown also are adiabatic trajectories of electrons, that is, the variation in $v_{||}$ as the electron moves along the magnetic field line due to the changing strength of the magnetic field. The color of the trajectories is representative of the value of the distribution function, which, by Liouville's theorem, is constant along each trajectory.

plotted along with the natural variation in particle velocity due to the effect of the changing magnetic field (see Equation (B.9)). In every case including near the magnetic equator, the distance over which the electron velocity is close to the resonant velocity of the wave is finite. Thus, to extend the length of the interaction with a single electron, the wave must act on the electron to keep it in resonance.

This extension of the interaction length can occur if the magnitudes of the wave force terms are sufficiently large relative to the magnitudes of the inhomogeneity terms in the equations of motion. In that case, a small portion of the energetic electron population is trapped by the wave and is forced to remain in resonance. The difference between the two populations is shown in Figure 3.11, where trajectories are traced for a single electron in two cases: first, for a fictitious zero-amplitude wave, and second, for a particle traveling through a constant, 10 pT wave field. In the first case, shown by the blue trajectories in Figure 3.11, the zero-amplitude wave has no
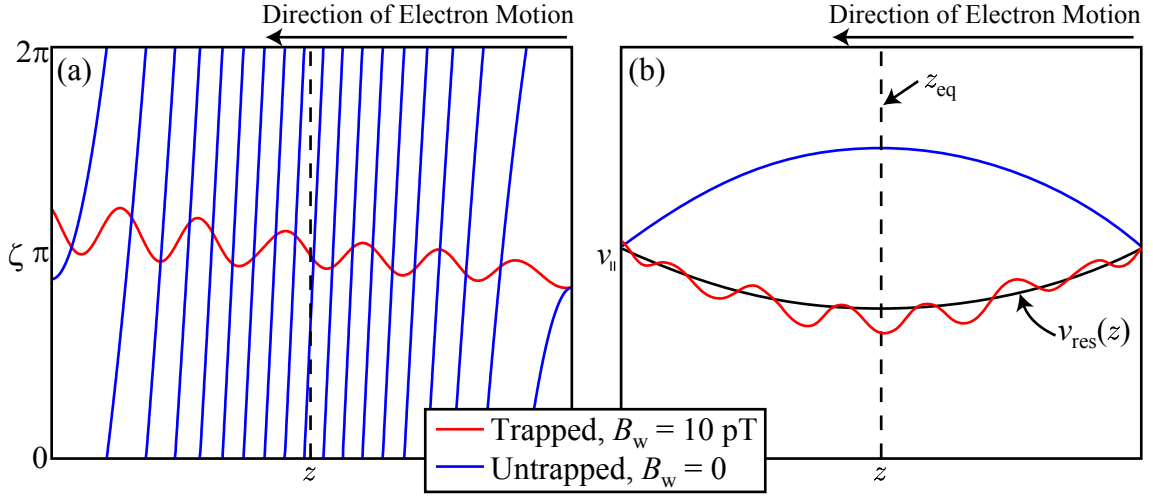
Figure 3.11: A comparison between trapped and untrapped trajectories through phase space as a function of distance along the field line z. The trapped particle trajectory is shown in red, while the untrapped particle trajectory is shown in blue. Electron motion is from right to left in the figure. (a) Particle trajectories in gyrophase ζ. (b) Particle trajectories in velocity parallel to the magnetic field $v_{||}$.

effect on electron motion. The electron begins in resonance with the wave, and the angle between the electron perpendicular velocity vector and the wave magnetic field vector is slowly varying, as shown on the right side of Figure 3.11(a). However, as the electron moves down the field line toward the magnetic equator into regions of lower magnetic field strength, the variation in the magnetic field causes the electron parallel velocity to increase, while the resonant velocity of the wave equivalently decreases, as shown in Figure 3.11(b). This separation in velocity space causes a rapid rotation in relative gyrophase, allowing for little net interaction between the particle and the wave. This rapid rotation continues until the electron is brought back into resonance as the electron moves away from the equator by the symmetric variation of the magnetic field (on the left side of Figure 3.11(a) and Figure 3.11(b)).

Increasing the wave amplitude to 10 pT creates a different picture, as illustrated by the red curve in Figure 3.11. In this case, the gyrophase of the electron in Figure 3.11(a) is initially slowly varying, as in the untrapped case. However, as the

electron continues down the field line toward the equator, the gyrophase remains slowly varying. That is, the action of the wave on the electron keeps it in resonance with the wave despite the effect of the varying ambient magnetic field strength. In Figure 3.11(b), the action of the wave keeps the electron parallel velocity close to the resonant velocity of the wave. Thus, the electron remains in resonance with the wave along the length of the field line shown, allowing for much more significant modification of the properties of the electron than in the untrapped case.

Because trapped electrons are held to the resonant velocity of the wave, the velocities of the trapped electrons can be significantly modified relative to the modification of the velocities of untrapped particles. *Nunn* [1974] demonstrates that the particles with the greatest energy change contribute the most to current formation; indeed, the phase angle of the resonant current ψ is close to the stable phase about which the trapped particles oscillate. To see this result, first consider how the position of the phase space "island" created by the particle trap evolves due to the effect of the magnetic field inhomogeneity. Figure 3.12 shows the phase space trajectories for electrons in the $v_{||}-\zeta$ plane at five separate points along the magnetic field line, symmetrically and evenly spaced about the magnetic equator, where the electron trajectories are in blue and the gyrophase position $\zeta=\pi$ is denoted by a dashed line. With the exception of the left-most plot (where the wave is not strong enough to maintain the trap in the presence of the inhomogeneity), the electron trajectories all display a separation between the trapped population and untrapped population. The trajectories of the trapped electrons form a phase space island at the resonant velocity around some stable phase ζ_0.

Several points are immediately obvious from Figure 3.12. Despite the symmetry in the placement of the positions used to calculate phase diagrams, the trajectories are not symmetric. This asymmetry occurs because the senses of the inhomogeneity are opposite on either side of the equator. That is, electrons streaming down the field line are first in a region where the ambient magnetic field strength is decreasing, then transition into a region where the magnetic field strength is increasing. This change in sign of the magnetic field gradient creates the asymmetry in the diagrams even though the magnitudes of the gradients at symmetric positions are the same.
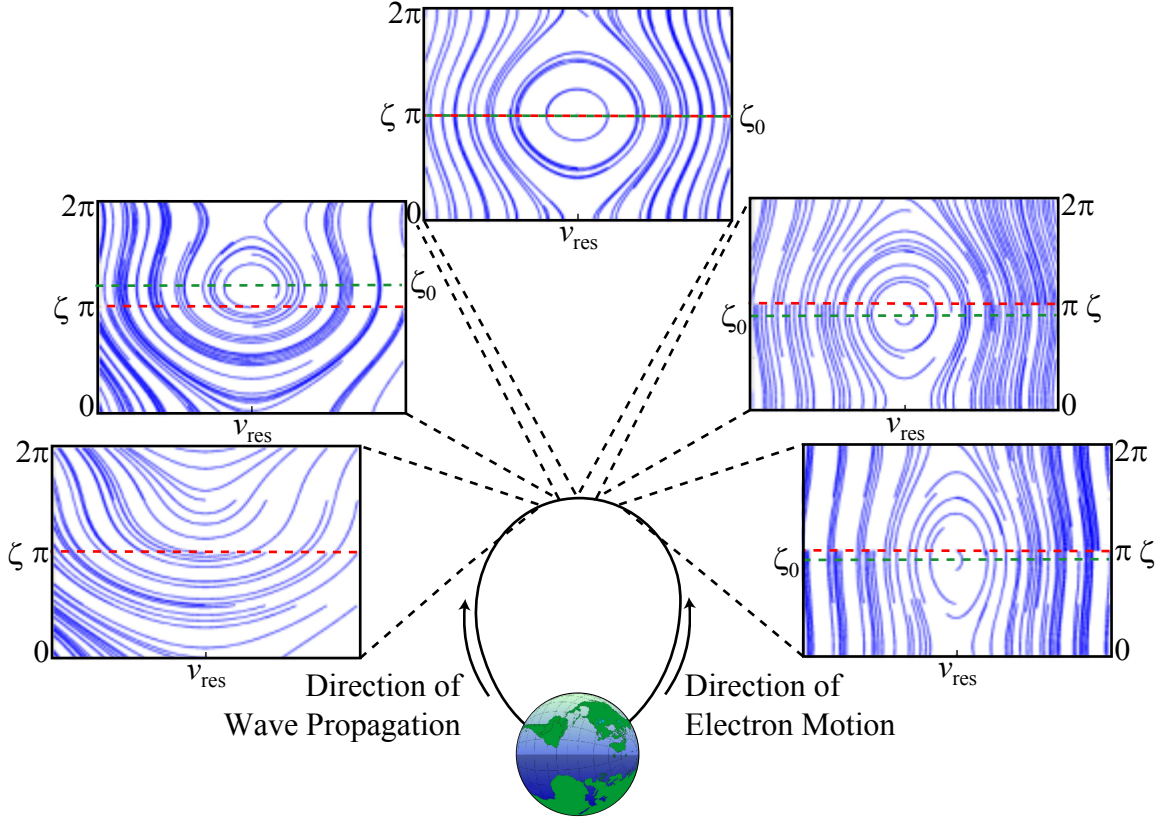
Figure 3.12: Phase diagrams showing trajectories of electrons in the $v_{||} - \zeta$ plane at five locations along the geomagnetic field line. The locations are evenly and symmetrically spaced about the geomagnetic equator. In each diagram, the phase $\zeta = \pi$ is marked by a dashed red line, and the stable phase $\zeta = \zeta_0$ for the phase-space island is marked by a dashed green line, when present.

Particularly, before the electrons cross the equator, $\zeta_0 < \pi$, while after the electrons cross the equator, $\zeta_0 > \pi$.

Another salient point from Figure 3.12 is that even in the presence of the inhomogeneity, particles that are initially on trapped trajectories remain trapped, while particles that are on untrapped trajectories remain untrapped. In other words, there are no trajectories that cross from the phase space islands into the phase space "sea." This condition is a result of the formation of a separatrix in phase space, which prevents the mixing of the trapped and untrapped populations.

Comparison of Figure 3.11(b) and Figure 3.10 reveals that a phase space electron

hole is formed by the existence of the separatrix. Specifically, Figure 3.11(b) shows the trend for trapped electrons to be held at the resonant velocity of the wave as they stream toward the equator. Liouville's theorem (see Appendix D) requires that the phase space density along electron trajectories is constant. Thus, the colored trajectories in Figure 3.10 schematically represent the ambient value of the electron distribution function as a function of z and $v_{||}$, showing that trapped electrons are carried from a relatively tenuous region, where they are originally trapped, into a region of higher ambient phase space density as they move toward the equator. This transport creates a relative depression of the energetic electron phase space density in the range of phase space contained within the trapping separatrix, creating a phase space electron hole [*Dowden et al.*, 1978; *Omura and Summers*, 2006]. Because this phase space hole represents a depression in the electron density relative to the ambient electron population, the net integration of the distribution function (see Equation (A.13)) creates a current vector that points toward the phase space hole, as shown schematically in Figure 3.13.

3.5 Energy Exchange

The contours of the whistler mode interaction from Section 3.2, along with the dynamics of trapped particle motion, define the energetics of the VLF triggered emission instability. Notably, an interaction that causes an increase in electron parallel velocity also causes a decrease in electron perpendicular velocity and a corresponding decrease in total electron energy. A decrease in electron parallel velocity has the opposite effect. Thus the available energy in this interaction comes primarily from the perpendicular energy of the interacting electrons. This energy must be divided between the increase in electron parallel energy, additional wave energy, and heating of the cold electron population [*Matsumoto and Kimura*, 1971].

In particular, the resonant current formation from the electron hole creation in the range of phase that contributes to growth ($0 < \zeta_0, \psi < \pi$) occurs in the region where the electrons are moving toward the equator. However, the resonant velocity decreases as the electrons approach the equator, causing the parallel velocity of the
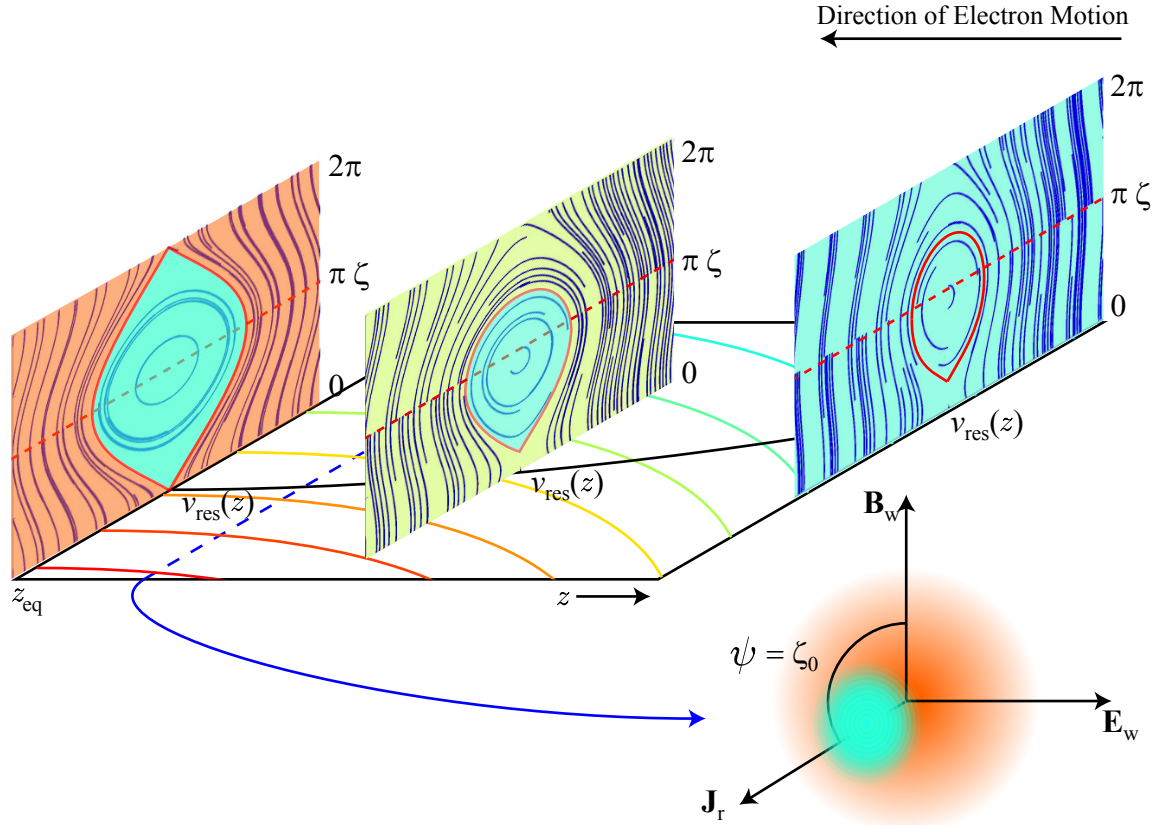
Figure 3.13: Schematic illustration of phase space electron hole current formation. Depicted in isometric view are three phase diagrams of $v_{||}$–ζ planes at three different locations along the field line. Shown for comparison is the resonant velocity and contours of constant energetic electron density in the $z-v_{||}$ plane, where the colors of the contours represent the unperturbed energetic electron density at each $(z, v_{||})$ position. The equivalent colors, representing the phase space density of the untrapped and trapped electrons, are shown in the $v_{||}$–ζ phase diagrams. The dashed blue line represents a single position along the field line where the electron distribution is projected into the $\zeta-v_{\perp}$ plane, showing the region of reduced density corresponding to the phase space electron hole. Because of the reduced energetic electron density in the range of gyrophase associated with the phase space electron hole, the integrated resonant current vector points toward the stable phase ζ_0 of the electron trap.

trapped electrons to equivalently decrease, which leads to an energization of the trapped particle population. In contrast, the untrapped electrons are scattered, on average, up in parallel velocity as they move toward the equator, which corresponds to a net decrease in electron energy [*Bell and Inan*, 1981; *Omura and Matsumoto*, 1982]. Thus, the energy for the unstable nonlinear interaction comes from the linear interaction between the untrapped electrons and the wave. This energy contributes to wave growth as well as to the energization of the trapped electrons.

3.6   Theories of Saturation

In light of the above discussion of the dynamics of the wave-particle interaction between ducted whistler mode waves and energetic electrons that drives the VLF triggered emission instability, the proposed mechanisms for the process of saturation can be divided into two categories: anisotropy destruction and energy removal.

- Anisotropy destruction: As can be seen from Equation (3.10), the sources of free energy for the whistler mode instability are the flux of resonant electrons and the anisotropy of the resonant energetic electron distribution. The energetics of wave growth require that the net motion of electrons in velocity space is up in parallel velocity and down in perpendicular velocity, corresponding to a reduction in the anisotropy of the distribution function. This reduction naturally leads to a reduction of the free energy available for growth, with the reduction in anisotropy resulting in an increase of entropy in the system. Since linear theory predicts no long-term deformation of the static distribution function, the effects of the whistler mode wave on the electron distribution function can only be revealed by a quasilinear or nonlinear treatment of the problem.

  A number of mechanisms for this anisotropy destruction have been proposed. In particular, the development of a trapping separatrix and the resultant stirring of the electrons within the separatrix can produce rapid and highly localized destruction of anisotropy. This destruction can occur in one electron pass through the interaction region, for high amplitude waves, or for multiple electron bounces

during the whistler interaction, for lower amplitude waves [*Dysthe*, 1971].

Another proposed mechanism for the destruction of anisotropy is diffusion along the whistler interaction contours. In this mechanism, an individual electron passes through the interaction region several times, experiencing multiple discrete interactions with the whistler wave. Over the process of these many interactions, the electron experiences a random walk along the whistler interaction contours. This random-walk process also results in a net reduction in anisotropy of the distribution and corresponding termination in growth [*Cornilleau-Wehrlin and Gendrin*, 1979; *Omura and Summers*, 2004].

- Energy removal: The theory developed for this type of saturation has relied on the presence of a whistler duct. If the whistler duct is "leaky" in some nonlinear manner such that the energy from higher amplitude waves escapes from the duct faster than for lower amplitude waves, then there exists some maximum amplitude at which the rate of energy conversion to the whistler mode waves from the energetic electrons equals the rate of energy loss from wave leakage out of the duct. This leakage would, in effect, cause the growth of the wave to stall. Other nonlinear effects, such as generation of additional wave modes through wave scattering by density irregularities, also represent energy sinks to the system. Thus, amplitude-dependent modes of energy loss could also cause saturation of the instability. [*Nunn*, 1990]

Chapter 4

Saturation Characteristics

Data from the Siple wave injection experiment (see Chapter 2) provide 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.

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, and this variety is reflected in the complexity of the transmitted formats. To isolate the saturation phenomenon, only relatively simple transmissions are considered, which are typically embedded within more complex formats. Specifically, constant-frequency transmissions are examined, when the transmission is only a single frequency in the form of a continuous wave. Of 1986 transmissions, virtually all constant frequency transmissions contain a power ramp, i.e., the transmission power is increased slowly (typically at 10 dB/s) from some initial level to a full-power level. The purpose of this power ramp is to study the instability threshold (see Section 2.4.1), 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].

Additionally, because the signature of saturation can easily be masked by interference between signals received that have traveled on various paths with different delays, events are selected which show minimal evidence of excitation of multiple

magnetospheric ducts. Furthermore, only data from periods of low background noise, with signal-to-noise ratios greater than 20 dB, are considered.

This chapter presents observed characteristics of the VLF triggered emission instability at saturation, using data from the Siple wave injection experiment. The discussion first covers the amplitude effects at saturation, follows with the description of effects identified in the frequency domain, and concludes with analysis of the phase of signals showing saturation.

4.1 Amplitude Effects At Saturation

Oscillations in saturated whistler mode signals were first observed by *Bell and Helliwell* [1971]. Since saturation and subsequent re-growth are 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.

4.1.1 Long-period Oscillations

Figure 4.1 shows an event from 8 December 1986. Figure 4.1(a) 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 4.1(b) 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 growth. The period of the oscillation is $\sim$2 s, and this pattern of growth to saturation followed by suppression is termed the "long-period saturation oscillation."

Suppression of the input wave during the triggering of a VLF emission has been
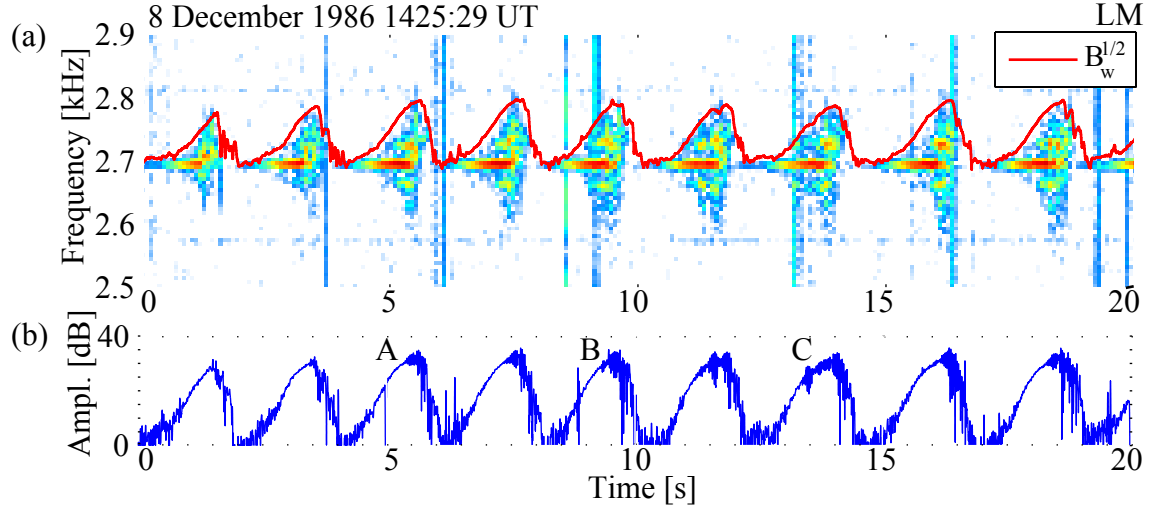
Figure 4.1: Siple transmission from 8 December 1986 showing saturation characteristics. The Siple transmission is a constant frequency key-down signal with a transmission frequency of 2.7 kHz. (a) Time-frequency spectrogram showing 20 seconds from the event. The development of wideband wave energy is clearly visible. Overlaid is a plot of the square-root of the normalized narrowband amplitude at the transmitted frequency, with a 170 ms delay. 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 2.7 kHz.

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 (below the initially received level, (see Figure 2.4(c))) until the emission rises significantly in frequency. Once the emission has risen in frequency by $\gtrsim 70$ Hz, the wave energy reappears at the transmitted frequency, growing at the same rate as it grew initially (see Section 2.4.1). The event in Figure 4.1 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, suppression appears to be an inherent part of the saturation process.
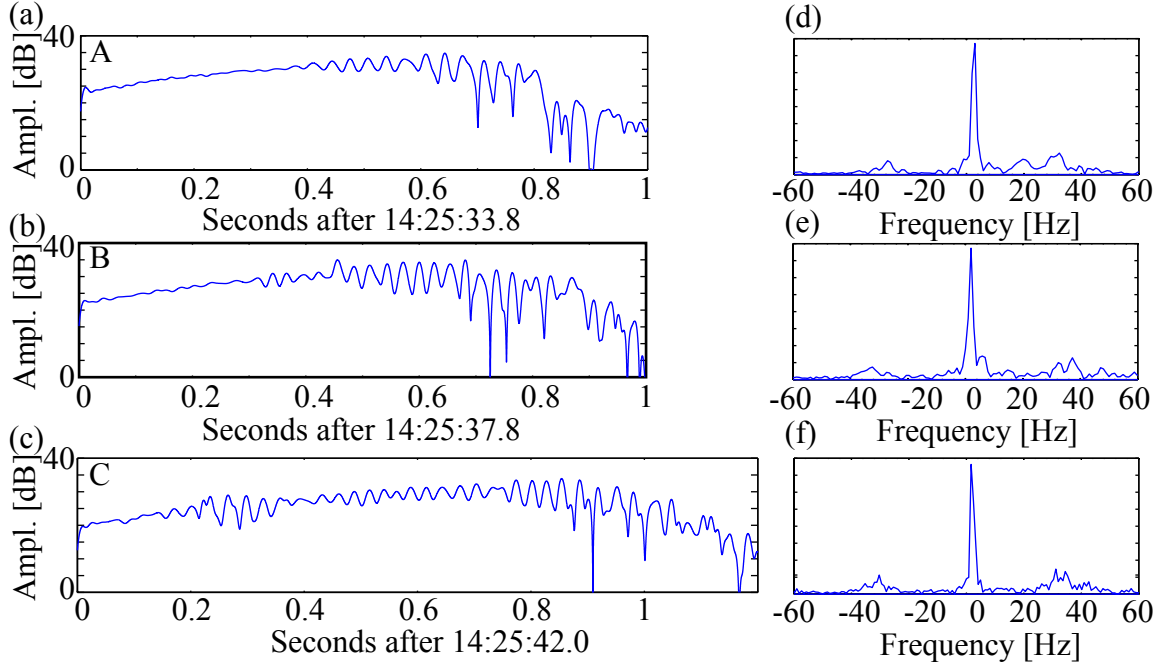
Figure 4.2: Siple transmission from 8 December 1986 showing saturation characteristics. The Siple transmission is a constant frequency key-down signal with a transmission frequency of 2700 Hz. (a-c) Expanded amplitude plots of maxima labeled "A", "B", and "C", respectively, in Figure 4.1(b). The plots show short period oscillations develop near saturation. (d-f) Fourier transforms of the data shown in (c), (d), and (e), respectively. The frequency content of the short period oscillations in all three cases is concentrated near 30 Hz.

Raghuram et al. [1977b] found that whistler mode two-hop echoes from Siple transmissions can cause a disruption or suppression of the growth process (see Section 2.4.4). Typical two-hop magnetospheric echo times for Siple signals were ∼4 s. Because of the shorter ∼1.5 s onset of the long-period oscillation shown here, it cannot be attributed to such echo-induced suppression.

4.1.2 Short-period Oscillations

In addition to the long-period oscillation, a short-period oscillation is also observed at saturation. The three panels, Figure 4.2(a-c), show expanded amplitude plots at

saturation, corresponding to the periods marked "A", "B", and "C" in the 20 s amplitude plot of Figure 4.1(b). 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 ∼20-40 Hz, in the spectra shown in Figure 4.2(d-f).

Figure 4.3 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 use of standard narrowband filtering to measure clear oscillations. Thus, the tilt in Figure 4.3(a) is in the opposite sense of the tilts in Figure 4.3(b) and Figure 4.3(c) because the transmitted frequencies are respectively below and above the frequency that corresponds to the minimum integrated path delay along the field line. Both Siple transmissions in Figure 4.3(b) and Figure 4.3(c) 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 4.3(a) remains relatively constant in frequency, and little variation in the oscillation period is observed.

4.2 Frequency Effects at Saturation

4.2.1 Sidebands

Sidebands in the triggered emissions process (see Section 2.4.2) 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]. The *Sa and Helliwell* [1988] 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.
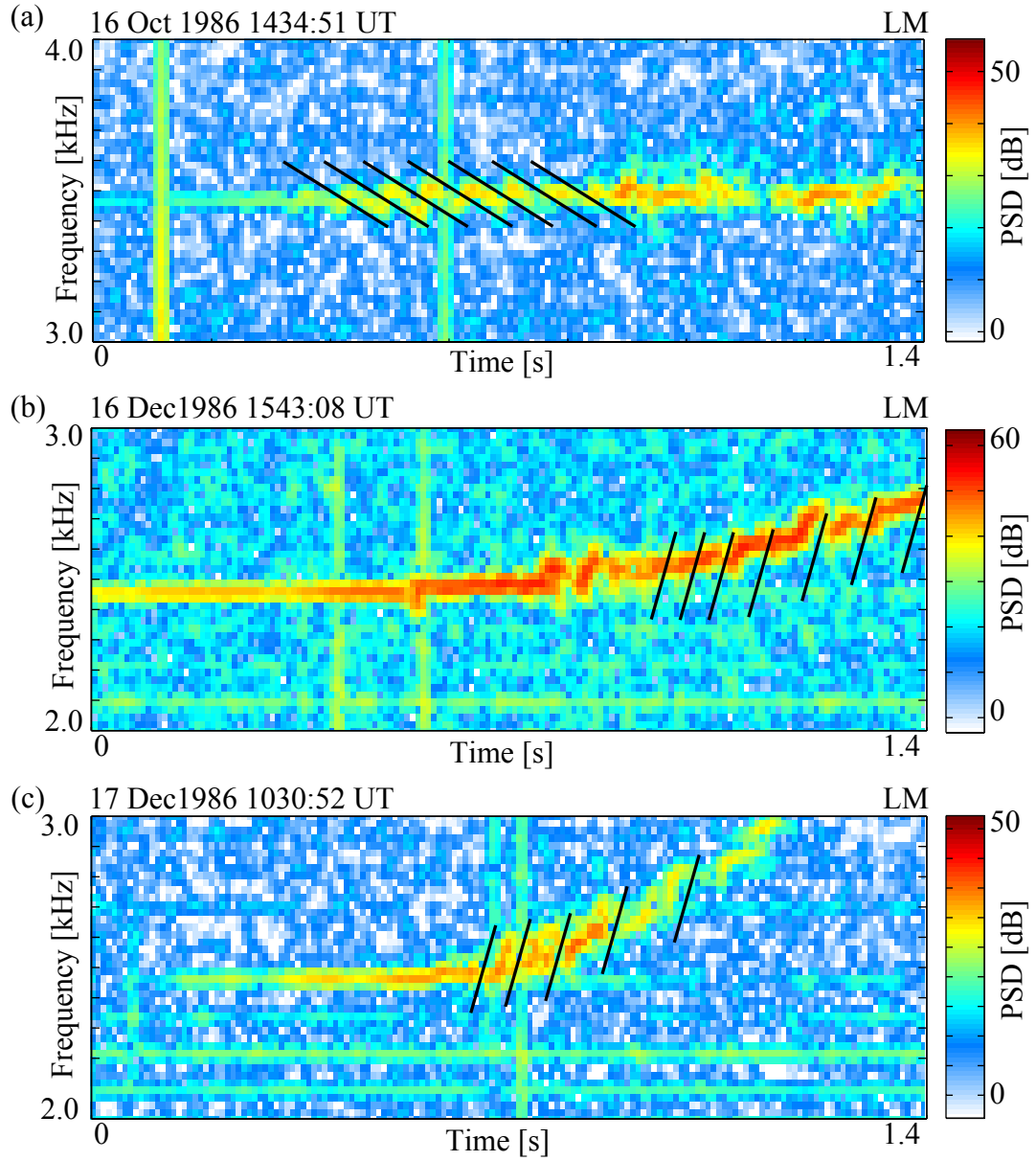
Figure 4.3: Time-frequency spectrograms of three events showing short-period amplitude oscillations associated with saturation. Minima in the oscillations are highlighted by lines. The angle of the lines is attributed to whistler mode dispersion.

A simpler explanation for the existence of sidebands is 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 4.2(d),(e), and (f) are the Fourier transforms of the time series data shown in Figures 4.2(a),(b), and (c), respectively. All three panels show sidebands ∼30 Hz above and below the frequency of the transmitted signal. These sidebands are frequency domain manifestations of the amplitude modulation associated with the short period oscillations that are visible in the amplitude plots.

Sidebands are a common occurrence at saturation. In addition to the event shown in Figure 2.5, Figure 4.4 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 then ensues, leading to the development of short-period oscillations, albeit less clearly than in the 8 December case. The frequency plots in Figure 4.4(d) display 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.

4.2.2 Frequency Spreading

Other frequency domain effects are also detectable. In addition to sidebands, the spectrogram in Figure 4.1(a) 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, $B_\mathrm{w}^{1/2}$, measured in a 20 Hz bandwidth around 2700 Hz, with a 170 ms delay, representing a temporal lag between the wave reaching a certain amplitude before the production of the equivalent wideband waves. 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 $B_\mathrm{w}^{1/2}$.
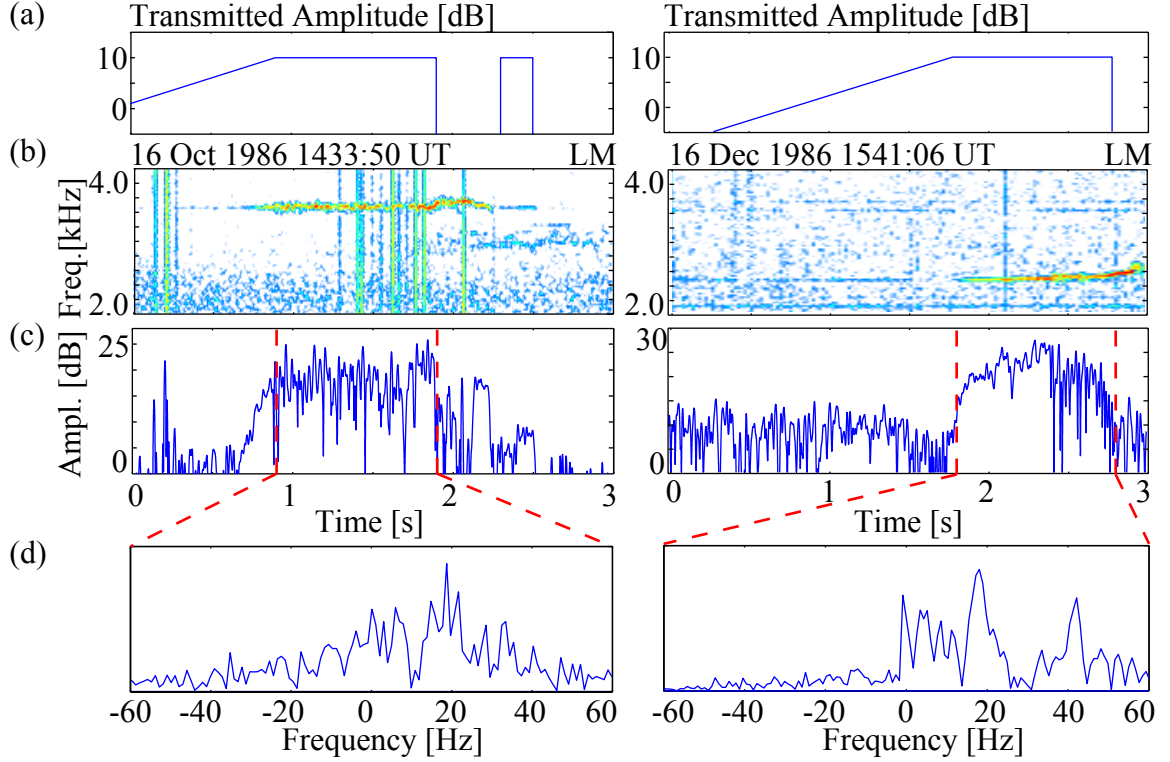
Figure 4.4: Two additional events showing sidebands at saturation, on 16 October 1986 and 16 December 1986. In both cases, the spectral content of the sidebands has a frequency spacing of approximately 20 Hz. (a) Variation of transmitted amplitude as a function of time, shifted to match the received amplitude at Lake Mistissini. (b) Time-frequency spectrogram of received signals. (c) Narrowband amplitude plots of the received signals in 160 Hz bandwidth about the transmitted frequencies. (d) Fourier transform of the narrowband amplitude data between the red lines in (c).

This measurement of the frequency spread of the wave allows an estimate to be made for wave amplitudes during the growth and saturation process. Since the frequency spread is proportional to $B_{\mathrm{w}}^{1/2}$, and consequently to the trapping bandwidth (see Equation (4.1)), which is the frequency of oscillation of electrons trapped in the wave potential well, the spread is likely caused by the scattering of electrons trapped in the potential well of the wave. This observation of frequency spreading thus allows a bound to be placed on the trapping bandwidth on the lower end (larger frequency spreads are potentially lost in the noise), and therefore provides an estimate of in-duct

wave amplitudes at saturation. With the trapping bandwidth given as [*Omura et al.*, 1991]

$$f_{\mathrm{tr}} \equiv \frac{1}{\pi} \left(k v_\perp \frac{q}{m_{\mathrm{e}}} B_{\mathrm{w}} \right)^{1/2} \tag{4.1}$$

the cyclotron resonance condition

$$f_{\mathrm{c}} - f = \frac{k v_{||}}{2\pi} \tag{4.2}$$

and the pitch angle of interacting particles α defined in Equation (3.4), the formula for the wave amplitude as a function of the trapping bandwidth is

$$B_{\mathrm{w}} = \frac{\pi m_{\mathrm{e}}}{2q} \frac{f_{\mathrm{tr}}^2}{f_{\mathrm{c}} - f} \cot \alpha \tag{4.3}$$

where ducted propagation has been assumed, with wave number k parallel to the ambient magnetic field; q is the fundamental charge; m_{e} is the rest mass of the electron; $v_{||}$ and $v_\perp$ are the velocities of the interacting particles respectively parallel and perpendicular to the geomagnetic field; B_{w} is the wave amplitude; $f_{\mathrm{c}} = \omega_{\mathrm{c}}/2\pi$ is the electron gyrofrequency; f is the transmitted wave frequency; and $f_{\mathrm{tr}} = \omega/2\pi$ is the trapping bandwidth.

The observable frequency spread measured in Figure 4.1(a) is $\sim$220 Hz around the transmitted frequency. From group delay measurements of Siple transmissions during the same hour, we find that the observed signals propagated near $L=4.7$ and obtain the result shown in Figure 4.5. 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 to 60 degrees at that bandwidth requires a minimum in-duct wave amplitude at saturation of $\sim$50 to 100 pT 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 at least 5 pT at saturation.
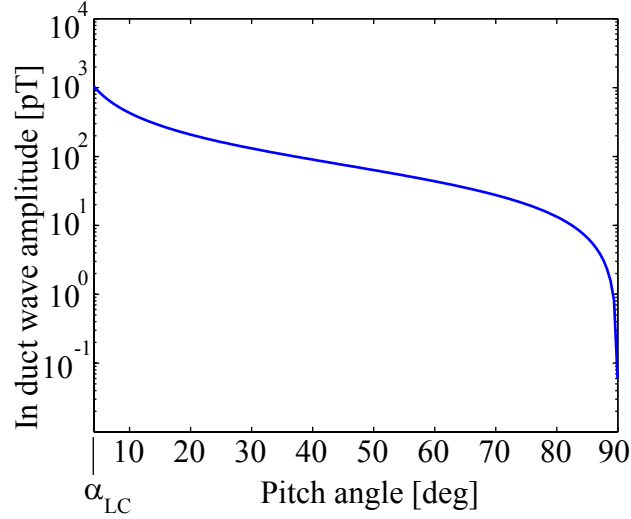
Figure 4.5: Estimate of the minimum wave amplitude required to produce the frequency spread observed in Figure 4.1(a) as a function of the pitch angle of trapped particles.

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, epoch averaging can be used 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 4.6. Figure 4.6 is, effectively, an average spectrogram of all of the long-period oscillations, plotted as a three-dimensional surface. Figure 4.6(a) shows the surface from an isometric perspective. The color in the figure is equivalent to the power spectral density, with the red representing the most intense spectral component and blue representing the weakest spectral component. Looking along the time axis, Figure 4.6(b) 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 fall-off 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. Moreover, Figure 4.6(c), which views the three-dimensional surface from Figure 4.6(a) from a perspective pointing along the positive frequency axis, shows a clear division between the existence of sidebands at saturation and the lack of sidebands during growth. That is, the ridge associated with the first sideband appears at 0.9 s into the record, which is approximately the same time as the onset of saturation. Thus, the spectral presence of sidebands is identified as the frequency-domain manifestation of amplitude oscillations during saturation, rather than being caused by the more exotic processes proposed by *Brinca* [1972] or *Sa* [1990].

4.2.3   Incoherent Wave Energy at Saturation

Figure 4.6(d) shows the frequency content of the wave at saturation. This measurement can be used to create an analytical model of the saturation frequency content. In the figure, the 'x's and 'o's represent the wave power spectral density for frequencies below and above the transmitted wave frequency f_0, respectively. The value of the abscissa is the absolute value of the wave frequency f relative to the transmitted wave frequency, $|f - f_0|$. Plotted against these points are two least-squares linear fits. The first corresponds to a fit of the points where $|f - f_0| \leq 10$ Hz. The second corresponds to a fit of points with $|f - f_0| \geq 20$ Hz. Thus, the wideband wave spectrum can be conceptually divided into two parts. The first, corresponding to frequencies near f_0, is a small spread in the wave energy associated with the generation of resonant currents by particles deeply trapped in the wave potential as the coherent wave is amplified (see Chapter 3). The second part of the wave spectrum, which roughly corresponds to wave frequencies outside of the first sideband, is attributed to the generation of currents associated with the erratic detrapping of particles near saturation (see Chapter 5).

Based on the character of the data in Figure 4.6(d), the power spectral density of
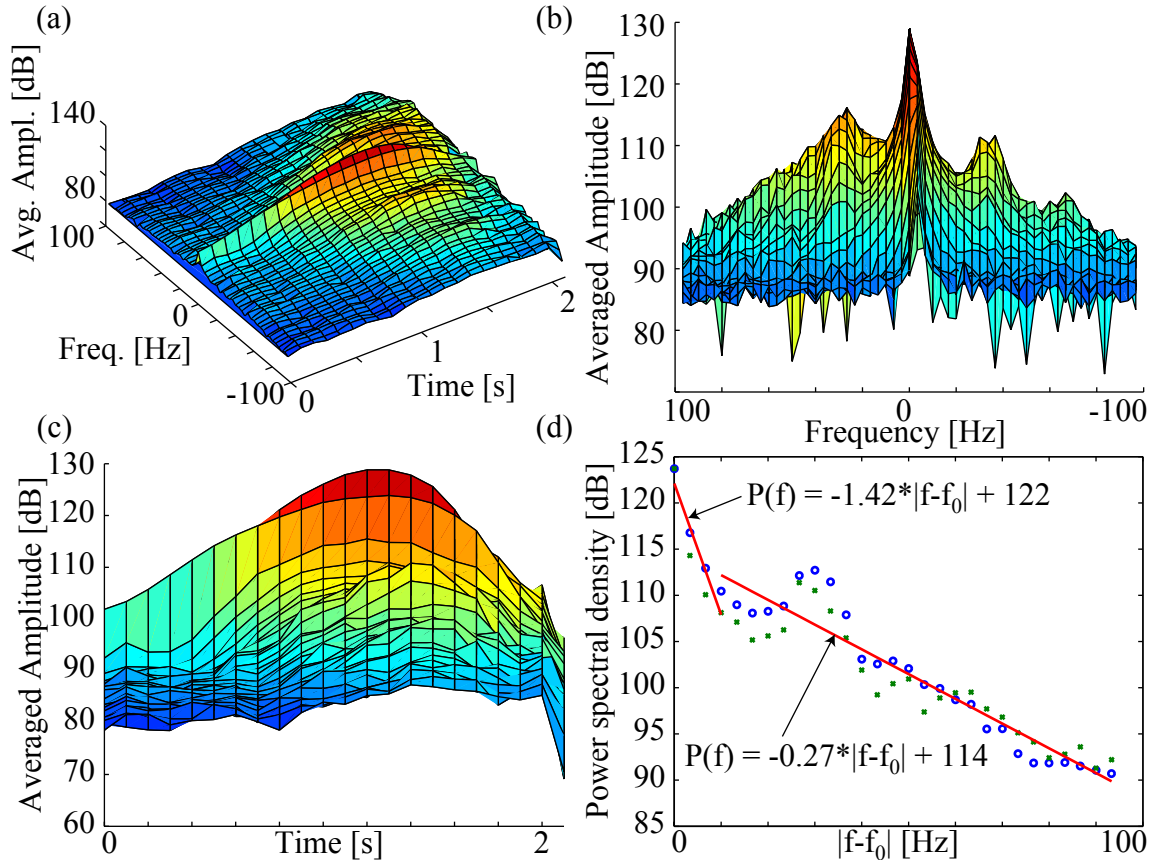
Figure 4.6: Frequency-amplitude measurements for analysis of the event in Figure 4.1. Each long-period oscillation is considered an individual epoch, and the frequency measurements are averaged over the epochs, producing a single long-period oscillation. The color is equivalent to the power spectral density, with the red representing the most intense spectral component and blue representing the weakest spectral component. 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 fall-off with increasing wave amplitude. (a) Average spectral behavior, shown in an isometric perspective. (b) Same as (a), looking up the time axis. (c) Same as (a), looking up the frequency axis. (d) Frequency spreading at saturation, with two piece-wise linear fits to the data. The first fit represents the frequency content of the coherent (i.e. with bandwidth ¡ 10 Hz), and the second fit represents the exponential fall-off of the incoherent wave energy and sidebands.

the wave, $P(f)$, is modeled as a piecewise-exponential envelope around f_0, such that,

$$
P(f) = \begin{cases} P_0 A_1 e^{-\left(s_1 \frac{|f - f_0|}{\sqrt{B_{\mathrm{w}}}}\right)} & |f - f_0| \leq f_1 \\ P_0 A_2 e^{-\left(s_2 \frac{|f - f_0|}{\sqrt{B_{\mathrm{w}}}}\right)} & |f - f_0| > f_1 \end{cases} \tag{4.4}
$$

where P_0 is the power spectral density of the wave at f_0; A_1 and A_2 are constants representing the intercepts of the linear fits for the two parts of the spectrum; s_1 and s_2 are constants of proportionality relating the square root of the wave amplitude to the spreading bandwidth; and f_1 is the frequency at which the two exponentials intersect.

To aid in the construction of a model for this process (see Chapter 6) and to quantify the contribution of the various spectral components, the total wave power is calculated for three cases: a 1 Hz bandwidth wave, a wave including the spectral components near f_0 (i.e., with the spectral form defined by the first exponential in Equation (4.4)), and a wave including both spectral components (i.e., for $|f - f_0| \leq f_1$ and $|f - f_0| > f_1$). For the narrowband wave, an exponential envelope with a 1 Hz half-power bandwidth is assumed, the total power is

$$
\int_0^\infty P(f) df = \int_0^\infty P_0 e^{-(0.69|f - f_0|)} df = 2.9 P_0, \tag{4.5}
$$

For the near-f_0 component of the spectrum,

$$
\begin{aligned}
\int_0^\infty P(f) df &= \int_0^\infty P_0 A_1 e^{-\left(s_1 \frac{|f - f_0|}{\sqrt{B_{\mathrm{w}}}}\right)} df \\
&= \frac{\sqrt{B_{\mathrm{w}}}}{s_1} A_1 P_0 \left[2 - e^{-\left(s_1 \frac{f_0}{\sqrt{B_{\mathrm{w}}}}\right)} \right] \\
&\simeq 2 \frac{\sqrt{B_{\mathrm{w}}}}{s_1} A_1 P_0
\end{aligned} \tag{4.6}
$$

where the fact that $s_1 f_0/\sqrt{B_{\mathrm{w}}} \gg 1$ has been used to neglect the contribution of the exponential term. With the data from Figure Figure 4.6(d), the contribution of this component to the total power of the wave at saturation can be calculated. With the -1.42 dB/Hz estimated value of the exponential fall-off in the wave spectrum, with a 122 dB intercept and 123.7 dB saturation amplitude, the ratio $s_1/\sqrt{B_{\mathrm{w}}}$ is $0.163\,\mathrm{Hz}^{-1}$, and $A_1 = 0.68$, giving a total power of $8.3P_0$, which is ~ 3 times the power of the equivalent narrowband wave.

To determine the effect of the second portion of the wideband spectrum, the total power is again calculated, assuming again that f_0 is sufficiently large that exponentials of $-(s_1 f_0/\sqrt{B_{\mathrm{w}}})$ and $-(s_2 f_0/\sqrt{B_{\mathrm{w}}})$ may be neglected,

$$
\begin{aligned}
\int_0^\infty P(f)df &\simeq 2\left[\int_{f_0}^{f_1} P_0 A_1 e^{-\left(s_1 \frac{|f-f_0|}{\sqrt{B_{\mathrm{w}}}}\right)} df + \int_{f_1}^\infty P_0 A_2 e^{-\left(s_2 \frac{|f-f_0|}{\sqrt{B_{\mathrm{w}}}}\right)} df \right] \\
&= 2P_0\left[\frac{\sqrt{B_{\mathrm{w}}}}{s_1} A_1 \left(1 - e^{-\left(s_1 \frac{f_1-f_0}{\sqrt{B_{\mathrm{w}}}}\right)}\right) + \frac{\sqrt{B_{\mathrm{w}}}}{s_2} A_2 e^{-\left(s_2 \frac{f_1-f_0}{\sqrt{B_{\mathrm{w}}}}\right)} \right] \\
&= 2\frac{\sqrt{B_{\mathrm{w}}}}{s_1} A_1 P_0 - 2\frac{\sqrt{B_{\mathrm{w}}}}{s_1} A_1 P_0 e^{-\left(s_1 \frac{f_1-f_0}{\sqrt{B_{\mathrm{w}}}}\right)} + 2\frac{\sqrt{B_{\mathrm{w}}}}{s_2} A_2 P_0 e^{-\left(s_2 \frac{f_1-f_0}{\sqrt{B_{\mathrm{w}}}}\right)}
\end{aligned}
$$
(4.7)

From Figure 4.6(d), the value of the exponential fall-off is -0.27 dB/Hz, the intercept is 114 dB, and $f_1 - f_0 = 6.3$ Hz. Correspondingly, $s_2/\sqrt{B_{\mathrm{w}}} = 0.31\,\mathrm{Hz}^{-1}$, and $A_2 = 0.107$. Inserting these values into Equation (4.7), recognizing that the first term is simply the power associated with the near-f_0 spectral component, yields a contribution of $\sim 2.9P_0$ for the second component alone.

From the above calculations, one can conclude that the wideband frequency components created during the growth process contribute significantly to the total power of the wave. Compared to a narrowband wave, the amplified wave contains ~ 3 times the power in the near-f_0 portion of the spectrum, and an equivalent amount of power in the second portion of the spectrum. Comparing scattering by narrowband and

wideband waves, *Inan* [1987] found that narrowband waves were $\sim$3.5 times more effective. With this consideration, the scattering by the near-f_0 remains significant and should be included in any model of the wave-particle interaction, while the contribution of the second portion of the spectrum may be included as a correction.

4.3   Phase Effects at Saturation

Presuming that no free running plasma emissions are triggered by a Siple transmission, the phase behavior at saturation is related to the amplitude characteristics of the saturated signals, in particular following the same temporal pattern of the short-period oscillations [*Paschal and Helliwell*, 1984]. Figure 4.7 shows a case in which a 10 s constant frequency, constant power transmission is received at Lake Mistissini. Narrowband amplitude and phase plots are displayed for a 1.7 s segment of the data in a 160 Hz bandwidth around the transmitted frequency of 3.48 kHz. In this segment, the signal reaches saturation and irregular short-period amplitude oscillations develop. The phase data in Figure 4.7(c) show that no significant phase advance occurs through the growth and saturation phase of the instability, with one exception. The phase is relatively constant except during deep nulls in the received intensity, where the phase becomes erratic. This behavior is expected in that the amplitude of the signal during the nulls is on the order of the intensity of the noise, allowing the noise to control the rate of phase advance. However, near 2 s in the record, the phase does not return to same relative phase, as would be expected for the disruption of a coherent wave and as observed by *Paschal and Helliwell* [1984]. This result indicates some level of phase accumulation during the instability that is either disrupted or enhanced during these amplitude nulls.

To clarify this point, assume that the received signal is advancing with some phase rate Δf relative to the reference phase at frequency f_0. Thus, the phase of the received signal advances Δf cycles in one second. At time t, the phase of the signal is ϕ. Some disruption in the signal occurs such that the signal amplitude drops sharply for some time period Δt, then returns to its original value. If, during that interval, the process creating that signal remains coherent, the phase of the signal advances

by $\Delta f \Delta t$ cycles, such that the value of the phase of the signal is $\phi + \Delta f \Delta t$ when the amplitude of the signal is reestablished. In the case shown in Figure 4.7(c), $\Delta f \simeq 0$, so one would expect that the signal phase would remain unchanged as a result of the disruption, as it is during the disruptions at 1.75 s and 2.4 s into the record. However, the disruption at 1.95 s into the record causes a discrete phase shift, indicating some disruption in the coherence of the process generating the signal.

In the case of the 8 December event, the phase behavior is much clearer (see Figure 4.8). During the onset of short-period oscillations, the low-amplitude portions of the oscillations show a tendency for increasing phase. As the amplitude of the signal increases as part of the oscillation, the phase returns to the regular phase advance established during the growth portion of the instability. As the saturation portion of the instability progresses, the depth of the short-period minima increases. Correspondingly, the phase of the received signal diverges farther from the regular phase advance behavior. Eventually, the change in phase associated with the minima reaches nearly a full cycle, no longer returning to the regular advance character. At that point (near 0.7 s in Figure 4.8(b) and Figure 4.8(d)), the phase-coherent growth process becomes fully disrupted and the signal begins to be suppressed back to near the initial level.
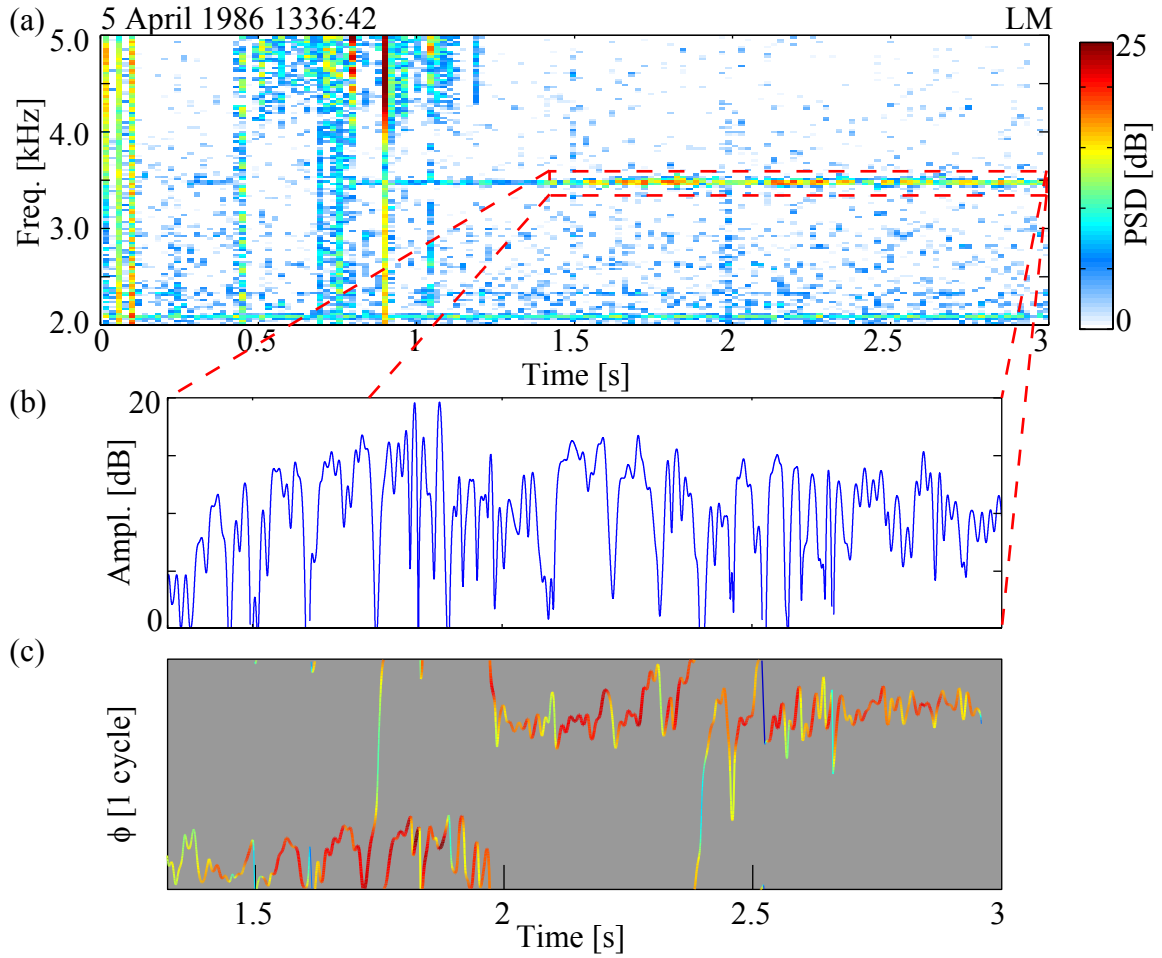
Figure 4.7: Phase behavior at saturation. The received signal is part of a 10 s constant frequency, constant power, 3.48 kHz pulse received at Lake Mistissini on 5 April 1986. The phase characteristics show typical phase response at saturation, where amplitude minima during the short-period oscillations correspond to abrupt changes in the phase of the received signal. (a) Time-frequency spectrogram of three seconds of data containing the initial portion of the pulse. (b) Narrowband amplitude plot of the pulse at saturation in a 160 Hz bandwidth around the transmitted frequency. (c) Phase plot of the same narrowband data, where the color of the trace corresponds to the amplitude of the signal.
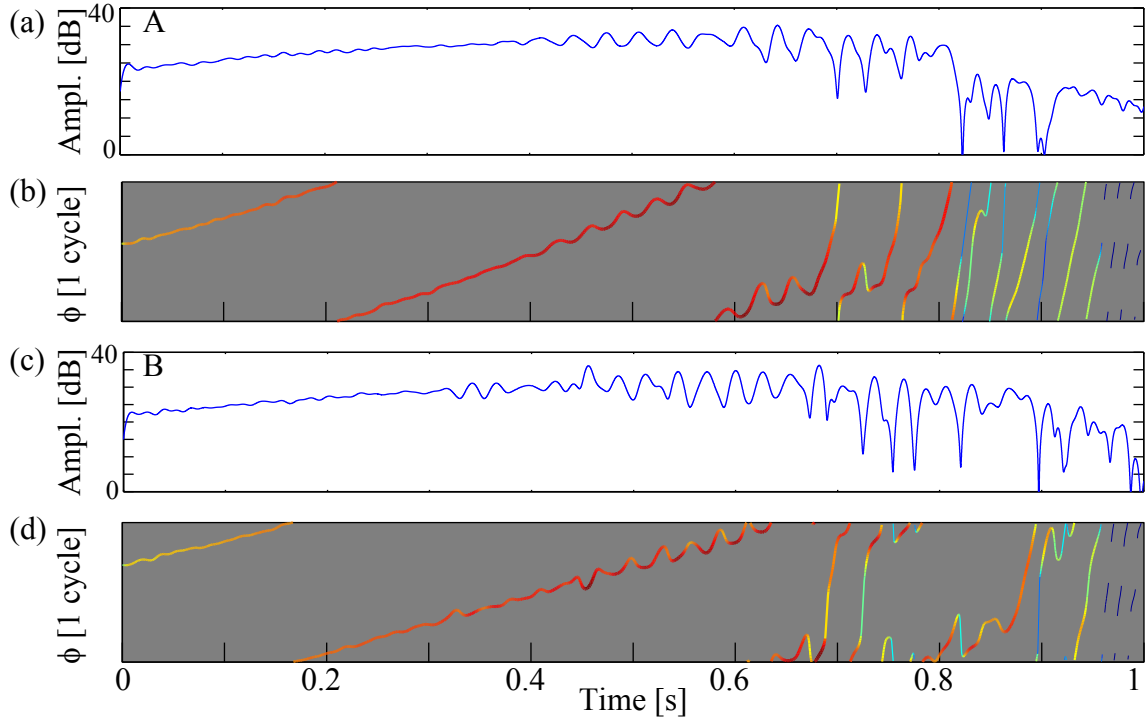
Figure 4.8: Phase behavior at saturation during the 8 December 1986 event. (a) The narrowband amplitude plot of segment "A" from Figure 4.1 (b) Phase plot for the same segment of narrowband data, where the color of the plot corresponds to the amplitude of the signal. (c) The narrowband amplitude plot of segment "B" from Figure 4.1 (d) Phase plot for the same segment of narrowband data, where the color of the plot corresponds to the amplitude of the signal.

Chapter 5

The Model of VLF Triggered Emissions

This chapter presents a new model for the simulation of the VLF triggered emission instability. Other models that have been applied to this problem are either computationally expensive, so that iterative studies of the problem are impractical, or rely on non-physical restrictions to produce outputs consistent with experimental results, so that the results of the simulation have questionable validity. A model is presented here that is computationally inexpensive (i.e., all modeling runs were performed in $\sim$2 hours on a desktop PC), but without non-physical restrictions other than those which are necessarily required by the use of a discretized code.

In particular, self-consistent computer simulations using the Particle-in-Cell formulation [*Omura and Summers*, 2004; *Katoh and Omura*, 2006; *Lampe et al.*, 2006] reproduce many of the characteristics of VLF triggered emissions, including rising frequency emissions and saturation. However, these simulations do not elucidate the underlying components of the instability that control the nonlinear growth process or saturation. For example,*Omura and Summers* [2004] show that the process of quasilinear diffusion is apparent in the end state of the energetic electron distribution following interactions with an amplified whistler mode wave. That is, 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 various aspects of the triggered emission process, such as saturation.

In comparison, computer simulations using the Vlasov Hybrid model [*Nunn*, 1990, 1993; *Nunn et al.*, 1997; *Smith and Nunn*, 1998; *Nunn et al.*, 2003, 2005] 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. 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 insight into the physical processes involved can be developed. 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 stability.

This chapter proceeds by developing an algorithm for the new model introduced in this dissertation, relying on the theoretical discussions in Chapter 3. Once the model is developed, results from the model are presented, showing that it reproduces results that are consistent with the observations of VLF triggered emissions from the Siple wave injection experiment described in Chapter 2. In addition, the characteristics of saturation of the instability, shown in Chapter 4, are also reproduced by the model, the results of which are used to provide a description of the physical mechanism of the nonlinear growth and saturation characteristics of the instability. Furthermore, studies are presented that demonstrate the sensitivity of the modeling results to the various input parameters of the model. Finally, the chapter concludes by discussing the advantages and limitations of the model and the conditions under which the

assumptions used to develop the model may break down.

5.1   Model Development

As mentioned above, the particle-in-cell method can very accurately model the physics of the interaction, but has so far been either computationally expensive or limited to considering very simple hot electron distributions. The Vlasov Hybrid model, on the other hand, has many attractive characteristics such as better computational efficiency as well as the tracking of the distribution function, which is more revealing of the evolution of the state of the energetic electrons. Because of these attractive features, a similar model is developed here, tracking the distribution function as it evolves in time and space, but without the use of any artificial parameters or assumptions concerning saturation.

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, ducted propagation can be assumed. That is, magnetospheric field-aligned cold plasma density perturbations act as guiding structures to keep the whistler mode wave propagation direction closely aligned to the magnetic field, which is necessary for signal reception at the conjugate ground station [*Helliwell*, 1965, Sec. 3.6]. This assumption allows variations transverse to the magnetic field to be neglected, so that the model can be constructed in a single spatial dimension.

The wave magnetic field is specified as

$$\widetilde{\mathbf{B}}_{\mathrm{w}} = B_{\mathrm{w}} \exp i(\phi + \Phi) \quad ; \quad \Phi = -\omega t + \int_{z_0}^{z} k(\omega, z')dz'$$

where B_{w} is the wave amplitude, i is the imaginary constant, z_0 is a position where $\Phi = 0$ at time $t = 0$, Φ is the rapid spatial and temporal variation of the phase of the wave field, and ϕ is any remaining phase variation after the rapidly varying contribution Φ is removed. The quantity ω is the angular wave frequency and k is the wave number, which varies spatially due to the magnetic inhomogeneity and variation of the cold plasma density along the field line. Here, z is the distance along

the geomagnetic field line, and the wave propagates northward in the $+z$-direction. 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 that is created by the resonant interaction between the wave and the energetic electrons, the following equations can be derived for the wave fields from Maxwell's equations and the linearized equation of motion for the cold electrons (See Appendix C),

$$\frac{\partial B_w}{\partial t} + v_g \frac{\partial B_w}{\partial z} = -\frac{\mu_0}{2} v_g J_E \tag{5.1}$$

$$\frac{\partial \phi}{\partial t} + v_g \frac{\partial \phi}{\partial z} = -\frac{\mu_0}{2} v_g \frac{J_B}{B_w} \tag{5.2}$$

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 Ψ 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 Φ and the cold plasma dispersion relation.

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,

$$J_B = -q \int_{v_{\parallel_1}}^{v_{\parallel_2}} \int_0^\infty \int_0^{2\pi} v_\perp^2 F(v_\parallel, v_\perp, \zeta) \cos \zeta \, dv_\parallel dv_\perp d\zeta \tag{5.3}$$

$$J_E = q \int_{v_{\parallel_1}}^{v_{\parallel_2}} \int_0^\infty \int_0^{2\pi} v_\perp^2 F(v_\parallel, v_\perp, \zeta) \sin \zeta \, dv_\parallel dv_\perp d\zeta \tag{5.4}$$

where $v_\parallel$ is the electron velocity parallel to the geomagnetic field, $v_\perp$ is the electron velocity perpendicular to the geomagnetic field, 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 ζ is the angle between the perpendicular velocity vector and the wave magnetic field vector.

The key component of any simulation of the VLF triggered emission instability lies in the manner in which the resonant currents are calculated, and here the approach

differs from the technique used in the Vlasov Hybrid code [*Nunn*, 1990, 1993]. To perform the calculation, the evolution of the distribution function as a function of time and space under the influence of the whistler mode wave must be determined. If the distribution function at time t is known, Liouville's theorem (see Appendix D) determines the distribution function at time $t + \Delta t$. That is,

$$F(t + \Delta t, v_{||}, v_\perp, \zeta, z) = F(t, v_{||} - \delta v_{||}, v_\perp - \delta v_\perp, \zeta - \delta\zeta, z - \delta z) \qquad (5.5)$$

where the parameters $\delta z, \delta v_{||}, \delta v_\perp$, and $\delta\zeta$ are the total changes in $z, v_{||}, v_\perp$, and ζ, respectively that were seen between t and $t + \Delta t$ by the particle that is at the phase space position $(v_{||}, v_\perp, \zeta, z)$ at time $t + \Delta t$. Using the particle equations of motion [Equations (3.15) through (3.18)], to first order

$$\delta z = -v_{||}\Delta t \qquad (5.6)$$

$$\delta v_{||} = -\left[\frac{q}{m_e}v_\perp B_w \sin\zeta + \frac{v_\perp^2}{2\omega_c}\frac{\partial\omega_c}{\partial z}\right]\Delta t \qquad (5.7)$$

$$\delta v_\perp = \left[\frac{q}{m_e}(v_p + v_{||})B_w \sin\zeta + \frac{v_{||}v_\perp}{2\omega_c}\frac{\partial\omega_c}{\partial z}\right]\Delta t \qquad (5.8)$$

$$\delta\zeta = \left[\omega_c - k\left(v_p + v_{||}\right) + \frac{q}{m_e}\left(\frac{v_p + v_{||}}{v_\perp}\right)B_w \cos\zeta\right]\Delta t \qquad (5.9)$$

where q is the fundamental charge, m_e is the rest mass of the electron, $v_p = \omega/k$ is the wave phase velocity, and ω_c is the angular electron gyrofrequency.

One complicating factor in the use of a numerical model is that the value of the distribution function can only be defined at discrete points in phase space. The phase space point on the right hand side of Equation (5.5) may take on a value that is not necessarily on the defined grid, as shown in Figure 5.1. As a result, the value for this point must be interpolated from the surrounding points in phase space. The interpolation scheme used in the model is a first order volume-weighted method.

Finally, note that the wave parameters (v_p, k) in Equations (5.6) through (5.9) 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, k is calculated
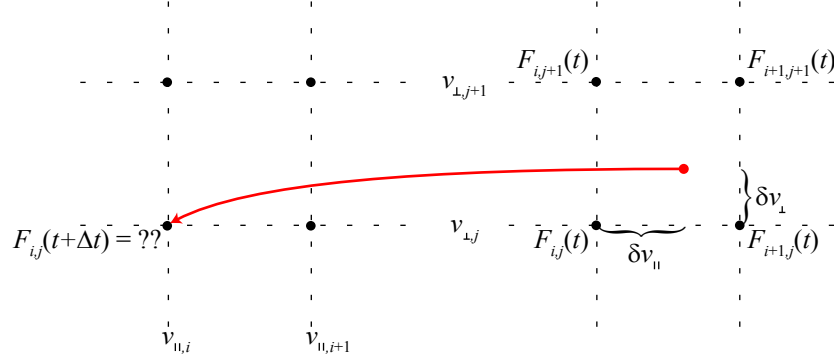
Figure 5.1: A schematic representation of the need for interpolation. The grid on the right represents the fixed points on the phase space grid where the distribution function is defined at time t, shown in two dimensions. The grid on the left represents the points on the same grid at time $t + \Delta t$, which are unknown. The particle which arrives at the grid point (i, j) at time $t + \Delta t$ may start at a point in phase space at time t that is not necessarily on the defined grid, as shown in red.

using the approximation,

$$k = \frac{\partial}{\partial z} (\phi + \Phi) \tag{5.10}$$

to determine the actual wave number as the instability evolves, v_{p} is calculated using the cold plasma dispersion relation (see Equation (3.2)),

$$v_{\mathrm{p}} = \frac{k\omega_{\mathrm{c}}}{k^2 + \dfrac{\omega_{\mathrm{p}}^2}{c^2}} \tag{5.11}$$

This methodology represents an algorithm for modeling of the VLF wave growth and triggering process. An initial distribution function $F_{i,j,k,l}^0$ is defined, where the superscript 0 indicates that the value of distribution function is represented at time $t = 0$, and the subscripts i, j, k, l indicate that the value of the distribution function is represented at the position in phase space $(v_{\|_1} + i\Delta v_\|, j\Delta v_\perp, k\Delta\zeta, z_{\min} + l\Delta z)$, where $z_{\min}$ is the location of the wave input to the the modeled space, and $\Delta v_\|, \Delta v_\perp, \Delta\zeta$, and Δz are the resolution parameters of the system. The initial wave field values $B_{\mathrm{w}l}^0$ and ϕ_l^0 are initialized to zero, k_l^0 is initialized to the wave number corresponding to the initial wave frequency, and the rapid phase variation Φ is initialized as $\Phi_l^0 =$

$-k_l^0(z_{\min} + l\Delta z)$. From here, the process becomes iterative:

1. For each (i, j, k, l), determine the parameters $\delta z, \delta v_{||}, \delta v_\perp$, and $\delta \zeta$ using Equation (5.6) through Equation (5.9). Determine $F_{i,j,k,l}^{n+1}$ by interpolating F^n to find the value corresponding to $F^n(v_{||1} + i\Delta v_{||} - \delta v_{||}, j\Delta v_\perp - \delta v_\perp, k\Delta \zeta - \delta \zeta, z_{\min} + l\Delta z - \delta z)$.

2. Calculate the resonant current using the discretized form of Equation (5.3) and Equation (5.4),

$$
J_{\mathrm{B}l} = -q \sum_i \sum_j \sum_k v_\perp^2 F_{i,j,k,l}^{n+1} \cos \zeta \Delta v_{||} \Delta v_\perp \Delta \zeta
$$

$$
J_{\mathrm{E}l} = q \sum_i \sum_j \sum_k v_\perp^2 F_{i,j,k,l}^{n+1} \sin \zeta \Delta v_{||} \Delta v_\perp \Delta \zeta
$$

3. Update the field parameters B_{w}^n and ϕ^n according to the wave update Equation (5.1) and Equation (5.2), respectively. Update the rapid variation of the wave phase according to

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

4. Calculate k_l^{n+1} and $v_{\mathrm{p}l}^{n+1}$ using Equation (5.10) and Equation (5.11), respectively.

5. Iterate the above steps until the simulation is complete.

Note that the growth term on the right hand side of Equation (5.2) 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 ϕ at an equivalently high rate, producing frequency components that violate Nyquist considerations and affect system stability. Thus, at each time step, the value of this term is limited so that ϕ cannot advance more than π 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. The necessity for such anti-alias filtering is a natural consequence of the use of a discrete time step.

5.2   Modeling Results

Before the model is used to examine the dynamics of the growth and saturation processes of the VLF triggered emission instability, it is first run several times to attempt to reproduce results observed from the Siple wave injection experiment (see Chapter 2). The observation of the expected behavior from the model serves as a method to qualify the modeling results. After the qualification of the model, results for a single run that reproduces all of the aspects of saturation identified in Chapter 4 are presented. With the results from this run, the detailed aspects of the growth and saturation process in the instability can be examined.

5.2.1   Qualification Runs

Before using the model to analyze the saturation process, several qualification runs are performed to demonstrate that the model produces the observed behavior. First, the model is run with the goal of reproducing an experimental event, which is shown in Figure 5.2. The 200 ms constant-frequency pulse at 2.22 kHz that was transmitted at 1529:02.6 UT on 10 December 1986 is modeled, shown in the lower left of the spectrogram in Figure 5.2(a). The narrowband amplitude in a 80-Hz bandwidth around the transmitted frequency is shown as the solid blue line in Figure 5.2(b). 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 ramps allow the determination of the path of signal propagation ($L \simeq 4.69$) and equatorial cold electron density ($\sim$94.1 cm^{-3}) using dispersion analysis [*Sazhin et al.*, 1992]. From the dispersion analysis, the whistler mode delay for the 2.2 kHz pulse is 3.77 s, which corresponds to an arrival time of 0.17 s into the narrowband amplitude plot. The 200 ms pulse duration is marked by the vertical lines. The spectrogram and the narrowband plot both show wave energy arriving prior to this point as well as after the termination of the pulse. The early arrival is attributed to amplification on two weak ($\sim$20 dB less intense) paths with significantly shorter delays, highlighted by arrows in Figure 5.2(a). 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.

With the path properties, wave frequency, and pulse length determined, the only adjustable parameters left to choose are the input wave amplitude and the ambient (i.e., initial) 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, as discussed in Section 2.4.7. These observations measured the wave magnetic field amplitude to be ∼0.01 to 0.05 pT, for signals that have not propagated across the magnetic equator, where the majority of linear and nonlinear amplification is expected to occur [*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 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 4.1, 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 in which Siple transmissions were received on spaceborne receivers corresponded to transmissions that included the resonant frequency of the tuned transmitter. Therefore, the input wave amplitudes are expected to be reduced relative to those measured on spacecraft, and thus for modeling purposes, input wave amplitudes ≤ 0.01 pT are considered herein.

For the functional form of the initial hot electron distribution, the distribution function from *Bell et al.* [2000] is selected, which is derived from energetic electron flux measurements aboard the POLAR spacecraft for energies less than 20 keV. The total hot electron flux is adjusted to give the desired growth characteristics. Consistent with *Nunn et al.* [2005], the simulation results are not particularly sensitive to the choice of the 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 can strongly interact with the resonant particles within ∼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 well. Given
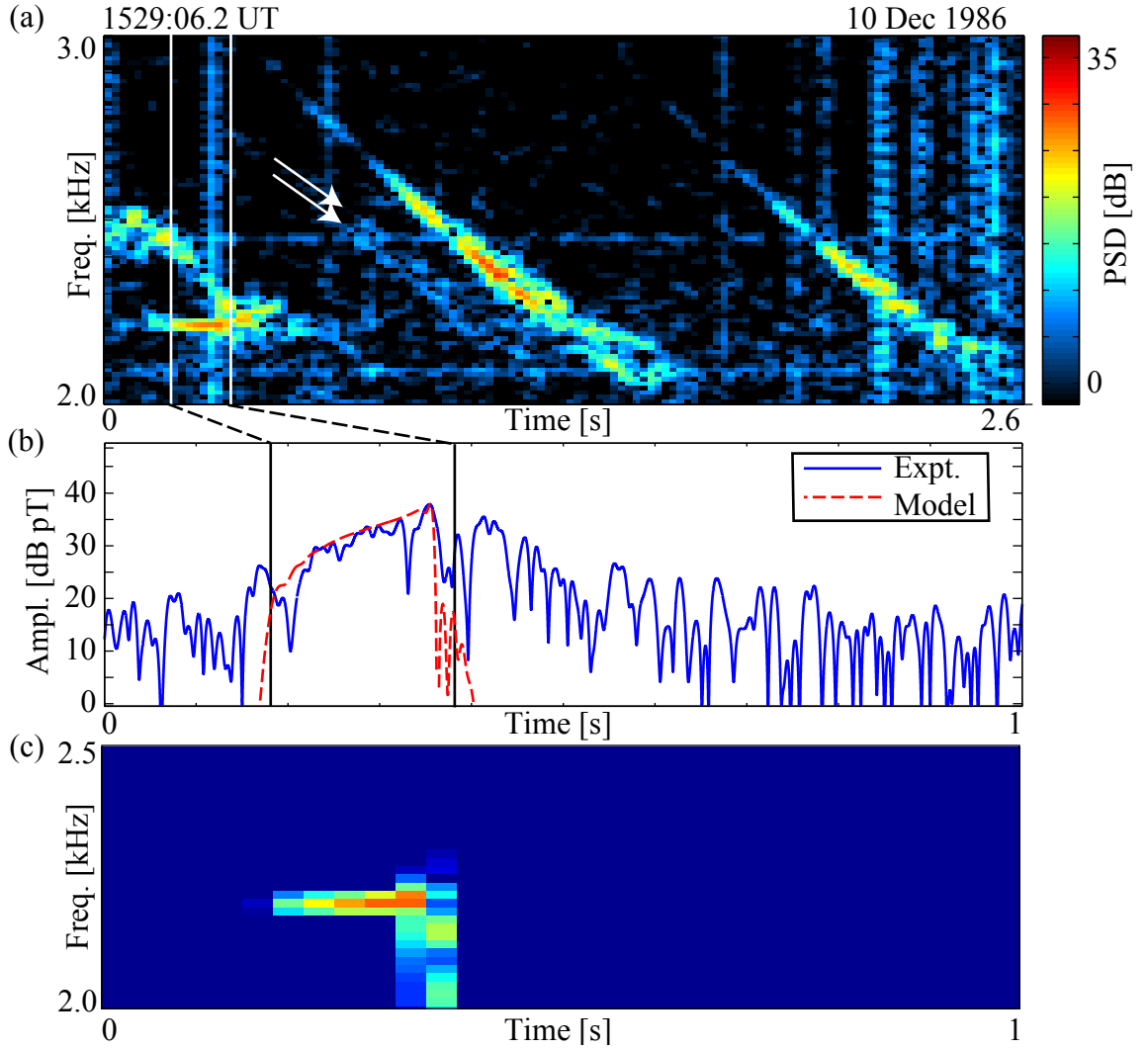
Figure 5.2: Triggered emission event used to qualify the saturation model. (a) Time-frequency spectrogram of the event. The modeled pulse is at 2220 Hz, and appears $\sim$0.2 s into the record. The 200 ms duration of the pulse is marked in the spectrogram using vertical lines. Dispersion analysis of the two following ramps shows a propagation path of $L = 4.69$ and equatorial cold electron density of 94.1 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.

input amplitudes $\sim$0.01 pT, the necessary amplification requires a strongly unstable distribution, corresponding to linear growth rates at the equator $\gtrsim$ 150 dB/s.

The various parameters controlling the resolution of the simulation are $\Delta t =$ 0.5 ms, $\Delta z = 21147$ m, $\Delta v_{||} = 50000$ m/s, $\Delta v_{\perp} = 1 \times 10^7$ m/s, $\Delta \zeta = \pi/12$. The parameters $\Delta t, \Delta v_{||}, \Delta v_{\perp}$, and $\Delta \zeta$ are selected by systematically halving their values until the simulation result does not change between iterations. The spatial step Δz is selected to satisfy the Courant condition for the most limiting of $v_{||}, v_{\mathrm{g}}$, and v_{p}, given Δt [*Courant et al.*, 1967]. The parameter $v_{||_1}$ is varied at each z-position so that the range of $v_{||}$ values is a 9×10^6 m/s range centered on the resonant velocity at each z, corresponding to the input wave frequency. This range of $v_{||}$ values is also found by performing the simulation with successively larger ranges until the simulation result did not change between iterations. The perpendicular velocity is modeled over the range 1×10^7 m/s to 15×10^7 m/s, corresponding to an energy range of 284 eV to 63.9 keV, which is expected to cover the entire $v_{\perp}$ range of interacting particles (i.e., a pitch angle range of 15.5° to 76.5° at the equator). The relative phase ζ of the interacting distribution is modeled between 0 and 2π. The spatial extent of the model is -2000 km to 2000 km along the field line, centered around the magnetic equator. Modeling larger regions does not significantly affect the simulation results.

The simulation is 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 red line in Figure 5.2(b). After transiting the modeled region, the wave arrives at the output of the region with an amplitude of 10 pT, with linear growth being responsible for the amplification from the initial 0.006 pT level to the level first received. Subsequently, the wave grows exponentially (in a non-linear fashion) to a maximum amplitude of 90 pT at an observed growth rate of 61 dB/s. The growth rate and total growth ($\sim$19 dB) fall well within the range of parameters reported for the Siple experiment (see Chapter 2). Like the experimental result, shown as a solid blue 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.

To further qualify the model, the input wave amplitude is reduced to 0.003 pT and the simulation is repeated, with the expectation that the amplified pulse will not reach saturation prior to termination. As shown in Figure 5.3(a), such is indeed the case. Here, the initial amplitude of the pulse (as it arrives at the model output) as well as the observed growth rate are reduced relative to the 0.006 pT pulse, and the amplitude of the signal continues to increase until the termination of the pulse (i.e. no saturation is observed). The accompanying spectrogram in Figure 5.3(b) shows that the pulse termination triggers a rapidly rising emission. The pulse duration is reduced to 100 ms, and the nature of the emission is changed to a faller, as shown in Figure 5.3(c). This effect of longer pulses triggering risers and shorter pulses triggering fallers was observed by *Helliwell and Katsufrakis* [1974], and the fact that the model reproduces this effect serves to increase confidence in the model.

5.3 Saturation Results

Now turning attention to the study of saturation, the input wave amplitude is set to 0.006 pT and the pulse length is increased to 800 ms. The results are shown in Figure 5.4 and Figure 5.5. Figures 5.4(a) and (b) show the spectrogram and amplitude of the output wave, respectively, as a function of time. Figures 5.5(a), (b) and (c) 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 in Figure 5.5 is spaced 0.5 ms after the previous column. The minimum wave amplitude required for particle trapping as a function of position is indicated with a color bar on the right hand side of Figure 5.5(a). The equator is marked by a white horizontal line. All of the plots in Figure 5.4 and Figure 5.5 share the same horizontal axis, with the simulation time running from 0 s to 1 s. From Figures 5.4(a) and (b), one can see that the long period oscillations, short period oscillations, and wave frequency spread that were observed in Chapter 4 are all reproduced by the model. There are three consecutive periods of growth followed by wave frequency 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
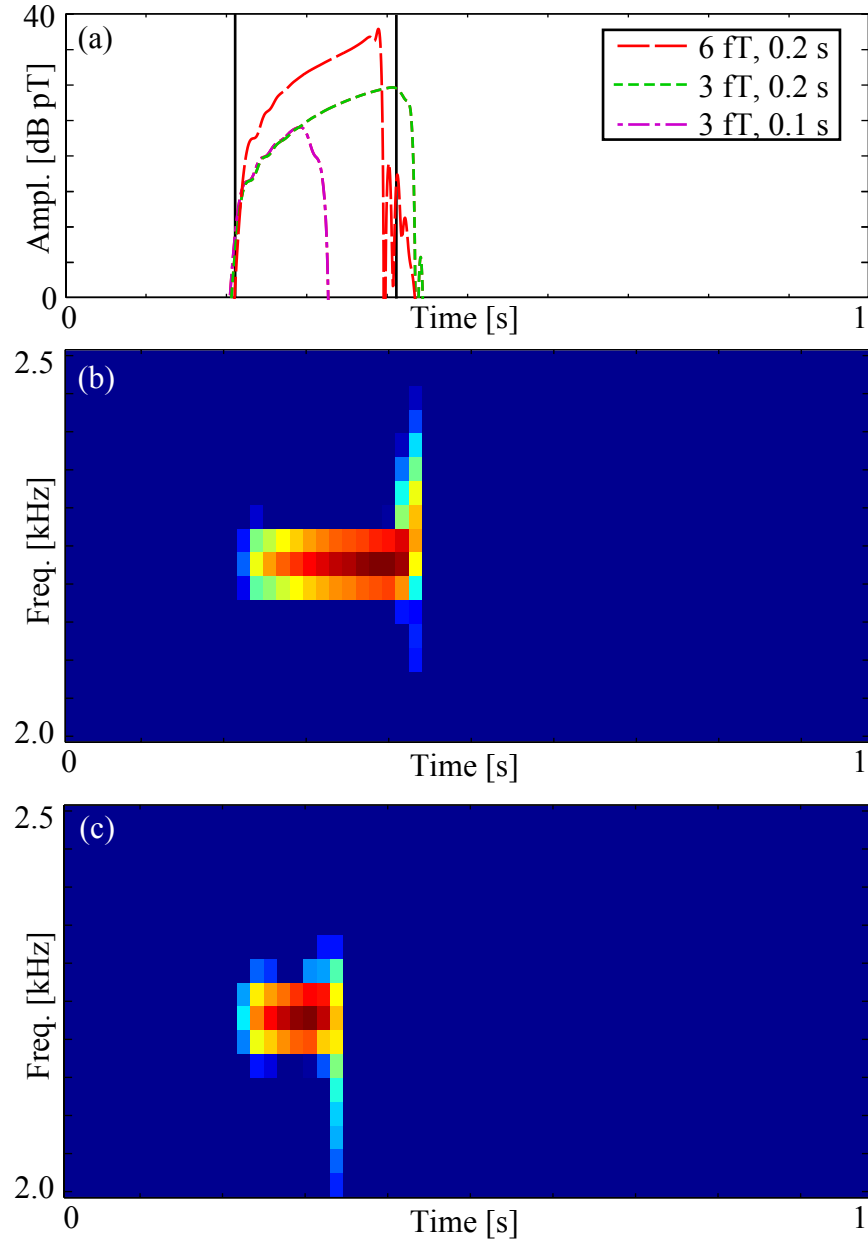
Figure 5.3: Qualification runs of the model. (a) Amplitude traces for the three qualification runs. The first, 6 fT, 0.2 s pulse is also shown in Figure 5.2(b), shows a very weak VLF emission triggered at termination by the extension of wave energy beyond the expected length of the pulse (marked by black lines). The other two, each with an input wave amplitude of 3 fT, also show termination triggering. (b) A 0.2 s pulse triggers a rising emission. (c) A 0.1 s pulse triggers a falling emission.

continue through three oscillations as the wave amplitude drops to the suppressed level.

The dynamics of the saturation process are captured in Figures 5.5(a), (b), and (c). As the initial part of the pulse propagates toward the equator (along the dashed line segment $\overline{\text{AB}}$ in Figure 5.5(b)), 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 due 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 due to the increasing effect of the inhomogeneity. The amplification of the wave continues in a manner consistent with that predicted by linear theory (see Section 3.1.4), with the wave amplitude increasing as a function of position, such that the amplitude at the initial 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 $B_{\text{w}} \simeq B_{\text{tr}}$ is increased. South of Point D, the wave amplitude is not sufficient to trap the resonant electrons, while the wave amplitude north of Point D is great enough to cause particle trapping. Thus, resonant currents are enhanced as southward-streaming particles transition from a region where the wave field dominates and trapping can occur (north of Point D) to a region in which the magnetic inhomogeneity dominates and trapping is not possible (south of Point D).

Once the wave energy has filled the simulation space, there are two regions of enhanced growth. The northern growth rate enhancement causes the wave amplitude just south of Point D to be increased above the local trapping level, which enhances the growth rate even farther south, and so on, such that the enhanced growth rate is seen to move southward as time progresses, along the line segment $\overline{\text{DE}}$. 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
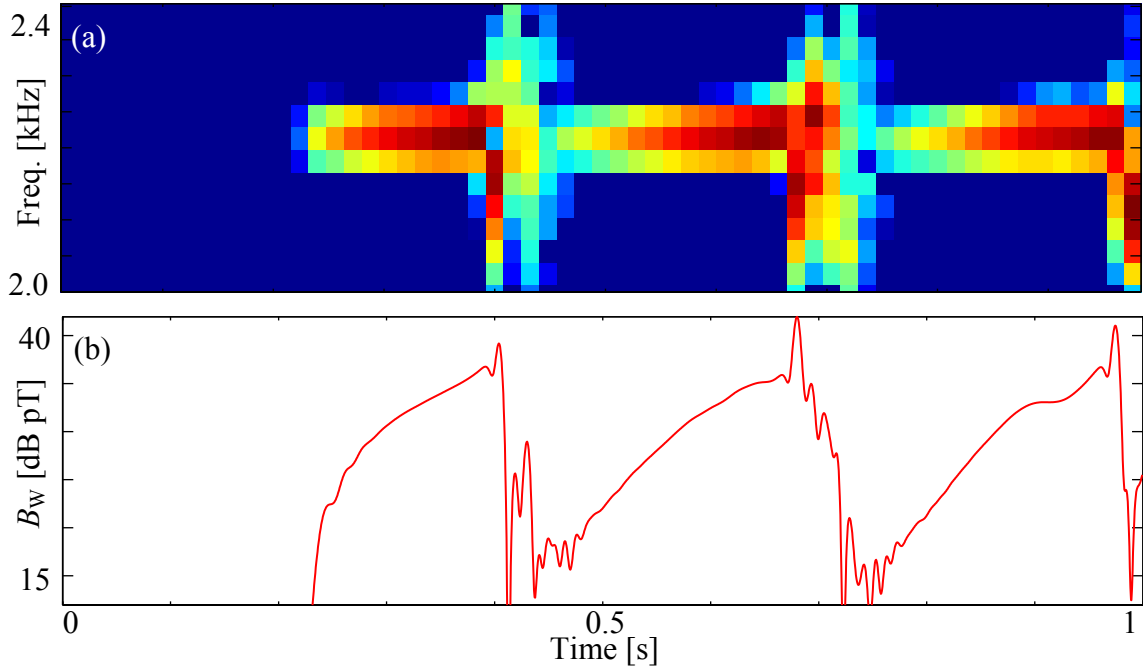
Figure 5.4: Output of the model showing saturation effects. The model is driven with a 6 fT pulse with pulse duration of 0.8 s. (a) Time frequency spectrogram of the model output, showing frequency spread at saturation and three long period oscillations. (b) Amplitude trace of the model output, also showing short period oscillations developing at saturation.

along the line segment $\overline{\text{CE}}$. When the two regions of enhanced growth rate merge, the growth south of the equator is strongly enhanced.

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). Figure 5.5(c) shows a plot of the wave number as a function of position and time during the simulation. The initial value for the wave number at the equator is 1.10×10^{-3} m^{-1}, and the initial k value decreases smoothly to 1.08×10^{-3} m^{-1}, resulting in initial values of k in the plot being represented by a pale blue or white color. In the plot, k values that differ significantly from the initial k value are represented by dark blue (for values less than the initial k value) or dark red (for values greater than
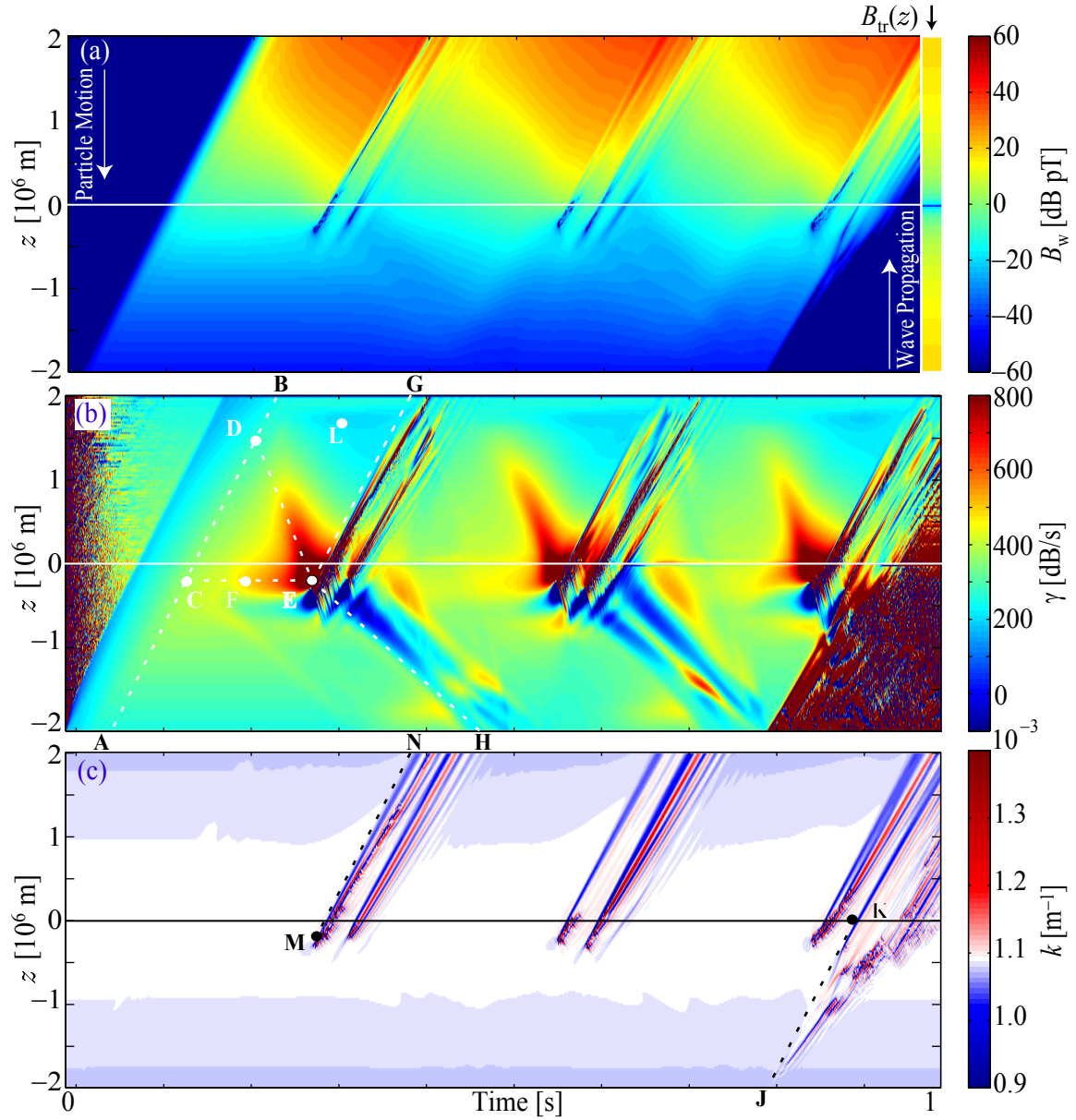
Figure 5.5: Modeled parameters during the saturation run of Figure 5.4. (a) Wave amplitude B_w as a function of time and position within the modeled region. The magnetic equator is marked by a white line, and the equivalent color scale for the minimum trapping amplitude is shown at the right. (b) Measured growth rate γ as a function of time and position within the modeled region. (c) Measured wavenumber k as a function of time and position within the modeled region. The initial value of the wave number varies smoothly as a function of position in the range $1.08 \sim 1.10 \times 10^{-3}$ m^{-1}

the initial k value). From the plot, it is clear that these significantly different wave numbers, corresponding to frequencies that differ from the transmitted frequency, are developed coincident with saturation (Point M in Figure 5.5(c), which corresponds to Point E in Figure 5.5(b)). At this point, the high-growth structure and off-frequency wave components propagate northward to the receiver, along line segments $\overline{\text{EG}}$ and $\overline{\text{MN}}$, while the location of negative growth appears to stream southward along line segment $\overline{\text{EH}}$. This process then repeats for the other two long period oscillations.

It is clear by comparing Figures 5.5(b) and (c) 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 M. 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 5.5(c) along line segment $\overline{\text{JK}}$. Thus, it appears that generation of additional wave frequencies is also associated with gradients in wave amplitude.

Returning to the point that the growing wave amplitude results in a transition from a trapped condition to an untrapped condition for a group of particles as they stream southward through the wave field, the effect of the transition from a trapped condition to an untrapped condition can be seen in Figure 5.6, which shows test particle trajectories in gyrophase ζ through the wave field, as developed at 0.25 s and 0.28 s into the simulation, respectively before and at saturation, as a function of position z. Figure 5.6(a) shows the growth rate (solid green line, calculated using Equation (3.19)) and the wave number (dashed blue 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.

Figure 5.6(b) shows test particle trajectories through the wave field at $t = 0.25$ s. The initial $v_\parallel$ and phase of the particles are chosen so that the particles are initially trapped in the high amplitude portion of the wave. Three different initial perpendicular velocities are shown, which sample the peak of the distribution function as a
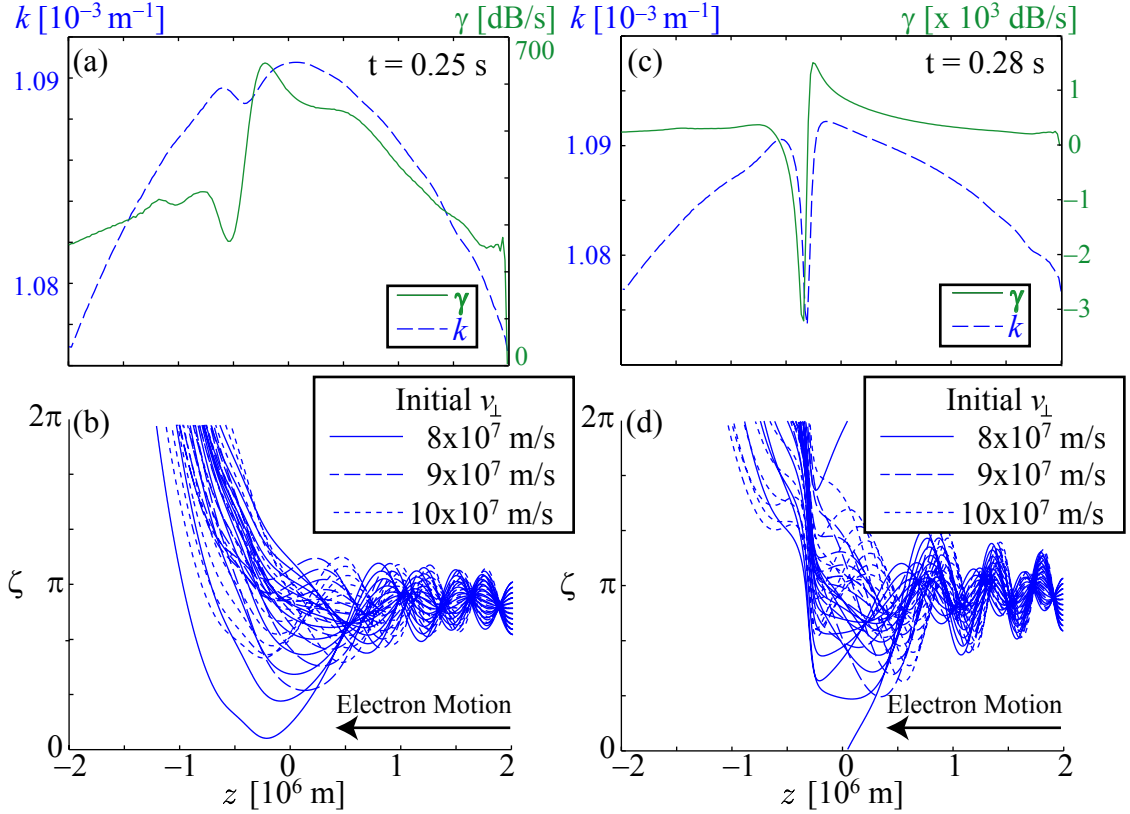
Figure 5.6: Evolution of the phase of trapped particle trajectories at two separate times during the simulation shown in Figure 5.4 and Figure 5.5. (a) and (c) Wave number k (solid line) and growth rate γ (dashed line) as a function of position at time $t=0.25$ s and 0.28 s, respectively. Both quantities show the effect of the de-trapping of particles and the evolution of the phase space electron hole. (b) and (d) Particle trajectories for initially trapped electrons and three different values of initial perpendicular velocity. Particles are de-trapped 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.

function of $v_\perp$ at the resonant velocity. As these particle 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 characteristics.

As the particles in Figure 5.6(b) 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 de-trapped. 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 due to the effect of the inhomogeneity, see Figure 3.10). This increase causes the relative phase angle ζ of the particle gyrophase relative to the wave phase to increase for every de-trapped particle. In turn, the phase space position of the electrons that were formerly trapped swings to some location $\zeta > \pi$ before the de-trapping electrons begin to lose their bunched character. At phase space locations with $\zeta > \pi$, the resonant current J_E switches sign, causing wave growth rates to decrease.

Initially, this transition of the resonant currents from growth to damping is slow, and the formerly trapped electrons mix with the ambient plasma before damping can occur. However, as the wave intensity and gradients increase, as shown in Figures 5.6(c) and (d), 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 formerly trapped electrons rotate through ζ-space. Specifically, as the wave is amplified the number of trapped electrons increases, which increases the magnitude of the resonant current. This increase leads to larger deviations of ϕ due to the effect of J_B, which drives the value of k down (note that in the frame of the wave, ϕ is seen to increase, as expected from Equation (5.2)). This decreasing k causes the resonant velocity of the wave to increase, causing in turn the formerly trapped electrons to rotate more rapidly in the wave reference frame. If the resonant currents are strong enough, then k is strongly reduced, and the strong de-trapping current transitions quickly from strong growth to damping.

When the damping becomes strong enough, the wave is absorbed by the plasma,

resulting in $B_\mathrm{w} \to 0$. This reduction in wave intensity causes the term on the right hand side of Equation (5.2) to become large, causing rapid variations in ϕ and a resulting spread in k, which is observed at the output as the generation of wideband waves at saturation. Because of the rapid variations in ϕ, the phase angle ψ (see Figure 3.8) of the resonant current structure relative to the wave field varies equivalently. Thus, the effect of J_E also varies rapidly, which is observed as rapid changes from damping to growth in the region of the de-trapping currents, as seen in Figure 5.5(b). The absorption of the wave also disrupts the amplification process and the entire spatial growth structure begins to propagate northward. As the region of rapid ϕ 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.

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 5.5(b), 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 (see Section 3.5). Specifically, parallel velocity of the trapped particles is decreased as they stream toward the equator, as the trapped particles are forced to remain in resonance with the wave (see Figure 3.10). Equivalently, the trapped particles are accelerated in perpendicular velocity, which corresponds to an increase in total energy for the trapped particles. The source of energy for this acceleration of the trapped population and the source of energy for the growing wave come from the energy of the resonant, untrapped particles. As the wave grows, the size of the trapped particle population increases, and more energy is diverted to accelerating the trapped particles. This increase in trapped particle population results in less energy available for wave growth and a corresponding decrease in the observed growth rate.

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 $\overline{\mathrm{EH}}$. The velocity of the negative growth region is approximately

10^7 m/s, which is significantly less than the equatorial resonance velocity of 3.6×10^7 m/s, and therefore cannot be caused by a single group of depleted particles. This negative growth region is in fact due to the destruction of anisotropy caused by the maximum amplitude portion of the wave packet. As the wave packet propagates northward from the point where saturation occurs (Point E), it leaves behind itself a depleted population (i.e., a group of electrons with reduced free energy available for wave growth in a small range of parallel velocities around the local resonant velocity) that streams southward on adiabatic trajectories (see Figure 3.10). 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 $\overline{\text{EH}}$.

The presence of this negative growth region affects the wave amplitude during the second and third long period oscillations. Waves injected between Point B and Point G in Figure 5.5(a) experience less total amplification as they approach the equator, which results in less growth enhancement in the region just south of the equator for the following long period oscillations. Furthermore, because of this reduction in total amplification, the wave amplitude at the output of the interaction region at 0.45 s in Figure 5.4(b) is reduced below the initially received level at 0.23 s.

Inan et al. [2004a] and *Platino et al.* [2006] report simultaneous observations of chorus emissions aboard multiple spacecraft. These observations are remarkable in that they show a doppler shift of apparently identical emissions, indicating a moving generation region. The estimated parameters of this motion showed that the interaction region was moving at a velocity of $0.3 - 1.2 \times 10^7$ m/s over a spatial region of 400 to 6000 km. In comparison, the modeling results here observe that the maximum growth region associated with VLF triggered emissions ($\overline{\text{DE}}$ in Figure 5.5(b)) moves at $\sim 0.5 \times 10^7$ m/s over a spatial region of $\sim$1500 km. This indicates that the mechanisms for the VLF triggered emission instability and chorus may be the same or very similar, confirming previous modeling work [*Omura and Summers*, 2006]. Futhermore, the results also indicate that the model presented here, with minor modification, may be useful in modeling the growth phase of chorus emissions.

From the saturation modeling results, some generalizations about the mechanism

of the instability can be made. The nonlinear growth of the wave is not driven by the trapping of electrons *per se*. As noted above, 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 amplitude grows spatially from the condition of no trapping $(B_{\mathrm{w}} < B_{\mathrm{tr}})$ to one in which trapping occurs $(B_{\mathrm{w}} > B_{\mathrm{tr}})$. Before reaching this region, counter-streaming trapped electrons experience a deep wave potential well and a separatrix that prevents them from mixing with the ambient electrons, forming the phase space electron hole. In the region of wave amplitude gradients, the separatrix disappears, and the electrons constituting the phase space electron hole (i.e., the "hole electrons") begin to mix with the ambient energetic electrons that have not been trapped by the wave. For some distance that depends on the magnitude of the wave amplitude gradients, the hole electrons remain in resonance with the wave, enhancing the linear currents produced by ambient energetic electrons. After that distance, the hole electrons mix with the ambient energetic electrons, and the formerly trapped population becomes indistiguishable from ambient population. If the wave amplitude gradients in the growth region are modest, the transition of the hole electrons from trapped to untrapped (and mixed) is smooth, resulting in an equivalently modest enhancement of growth. As the nonlinear growth continues and the wave gradients become sharper, the loss of the trap becomes much more abrupt, and the de-trapping hole electrons stay more tightly bunched for a longer distance, enhancing resonant currents. However, if the hole electrons stay bunched long enough, such that they start to lose resonance with the wave (due to the effect of the magnetic inhomogeneity) before they mix with the ambient population, the resonant currents start to affect the wave phase in a manner that drives the loss of resonance faster, eventually causing saturation, as discussed above.

5.4 Sensitivity Studies

The results of sensitivity studies using the model are shown in Figure 5.7. The studies are performed by setting the initial, base configuration of the model to the

initial values used to create the results in Figure 5.2. Specifically, the values of pulse length, $l_{\text{pulse}} = 400$ ms; equatorial cold electron density, $n_{\text{c,eq}} = 94.1$ el/cm^3; input wave amplitude, $B_{\text{in}} = 6$ fT; path of propagation, $L = 4.7$; linear growth rate, $\gamma_{\text{lin}} = 325$ dB/s; and transmitted frequency, $f = 2.22$ kHz are used. Each of these parameters are varied individually in small increments about their base value to produce the results shown. For the base configuration case, the pulse is observed to arrive at the output of the interaction region at 0.22 s and terminate at 0.62 s, with a growth rate of 60 dB/s and a saturation level of 38 dB pT (or 79 pT).

From the results, the following conclusions about sensitivity of each parameter can be made:

1. Pulse length l_{pulse}: The variation in modeled output as a function of pulse length is shown in Figure 5.7(a). Unsurprisingly, the growth and saturation characteristics are insensitive to the length of the input pulse. In this case, saturation is not evident until the pulse length is extended beyond 200 ms. The maximum amplitude of the signal increases with the length of the pulse, which is consistent with experimental observations [*Helliwell and Katsufrakis*, 1974].

2. Cold equatorial electron density $n_{\text{c,eq}}$: The variation in modeled output as a function of equatorial cold electron density is shown in Figure 5.7(b). The linear growth rate calculated from the hot electron distribution function is held constant, so the variation in cold electron density primarily affects the speed of propagation through the modeled region. The lowest cold electron density corresponds to the highest speed, which can be observed through the initial arrival time of the pulses at the output of the modeled region. The lowest cold electron density also corresponds to the lowest observed growth rate, while the saturation level is largely unaffected by the variation in cold electron density.

3. Input wave amplitude B_{in}: The variation in modeled output as a function of input wave amplitude is shown in Figure 5.7(c). Here, the effect of the threshold for the instability is clearly visible [*Helliwell et al.*, 1980]. (Note that the leading edges of the pulse show some rising-amplitude effect. This is due to the slow turn-on of the pulse in the model to avoid any ringing response of the simulation
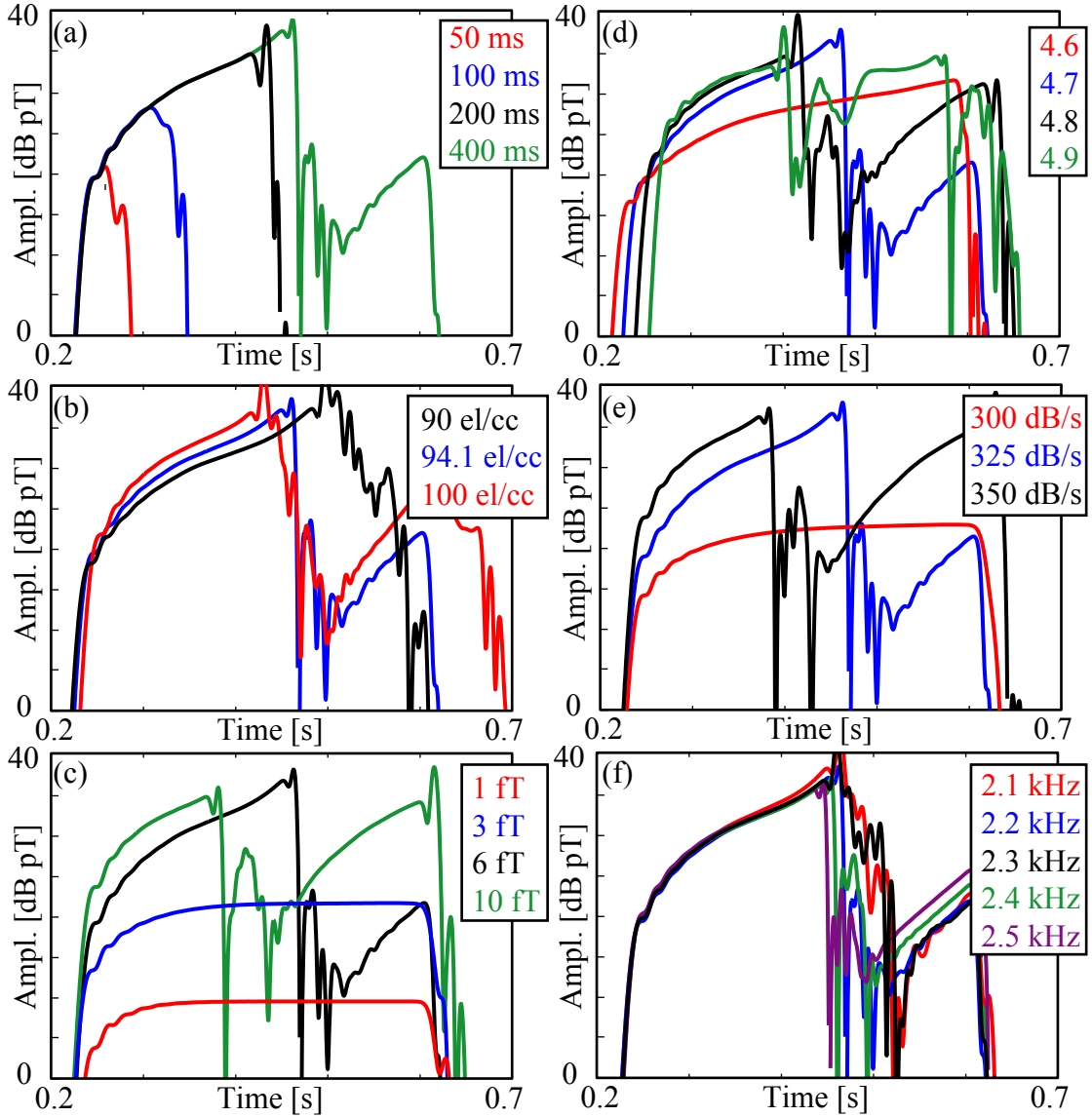
Figure 5.7: Sensitivity studies. The base point for the studies is close to the initial conditions used to produce the results in Figure 5.2. That is, $l_{\text{pulse}} = 400$ ms, $n_{\text{c,eq}} = 94.1$ el/cm^3, $B_{\text{in}} = 6$ fT, $L = 4.7$, $\gamma_{\text{lin}} = 325$ dB/s, and $f = 2.22$ kHz. Each parameter is varied individually and the resulting amplitude plots are shown. (a) Variation of the pulse length, l_{pulse}. (b) Variation of the cold equatorial electron density, $n_{\text{c,eq}}$. (c) Variation of the input wave amplitude, B_{in}. (d) Variation of the propagation path, L. (e) Variation of the linear growth rate, γ_{lin}. (f) Variation of the transmitted frequency, f.

to high frequency signal components.) Additionally, the rising initially received amplitude of the pulse corresponds directly to the increase in input amplitude. Finally, the growth rate and saturation level do not appear to have a dependence on the amplitude of the pulse.

4. Propagation path, L: The variation in modeled output as a function of the path of propagation is shown in Figure 5.7(d). The lower L-values correspond to faster wave propagation velocities and earlier pulse arrival time (recall that cold electron density is held constant). At lower L-values, the initial arrival level of the pulse decreases because the higher wave propagation velocity corresponds to a lower per-meter linear growth rate. As a result of this reduction in linear amplification, the wave propagating at $L = 4.6$ is amplified to a value near the equator that is closer to the threshold level. The proximity to the threshold level results in a lower nonlinear growth rate for the wave propagating at $L = 4.6$, while the other waves have approximately the same growth rate. The saturation level appears to decrease as the L-values increase.

5. Linear growth rate, γ_{lin}: The variation in modeled output as a function of the calculated linear growth rate of the initial energetic electron distribution is shown in Figure 5.7(e). As the linear growth rate increases, the initial arrival amplitude increases. Lower values of the linear growth rate are not sufficient to amplify the wave to the threshold level, and as a result, the lowest growth rate does not show nonlinear growth. For the higher linear growth rates, the saturation level and nonlinear growth rates are unaffected.

6. Transmission frequency, f: The variation in modeled output as a function of the transmitted frequency is shown in Figure 5.7(f). The modeling results do not appear to be significantly changed by the variation of the transmitted frequency. In particular, the modeled form of the distribution function from *Bell et al.* [2000] of

$$F_{\text{eq}} \propto \frac{\sin^{25}\alpha}{v^4} \tag{5.12}$$

gives a constant amount of anisotropy as a function of resonant velocity [*Gurnett and Bhattacharjee*, 2005, Sec. 9.3.2.4], and thus the initial distribution function is not particularly selective as a function of frequency. In contrast, the experimental results are often highly selective in frequency [*Helliwell and Katsufrakis*, 1978; *Dowden et al.*, 1978; *Golkowski et al.*, 2008]. This contrast with the experimental results implies that the initial distribution function may have additional gradients than those represented in Equation (5.12).

The following general conclusions can be made from the sensitivity studies:

- The saturation level is approximately constant over range of parameters considered. The only parameter that appears to affect the saturation level is the magnitude of the inhomogeneity. The higher L-values correspond to a weaker inhomogeneity in the magnetic field. In light of the above discussion of the growth mechanism, the bunch of particles corresponding to the de-trapping electron hole must travel farther down the field line at higher L-values before they are significantly out of resonance with the wave, which causes the generation of off-frequency wave components that eventually drives saturation. Thus, the particles must be less strongly bunched for higher L-values to maintain their coherence in a weaker inhomogeneity for the longer distance, resulting in lower saturation levels.

- The maximum level of the signal, without saturation, is determined by a combination of the growth rate and the amount of time that the pulse is allowed to grow. This result, combined with the tendency of falling frequency emissions to be triggered by shorter length pulses [*Helliwell and Katsufrakis*, 1974], implies that the rate of change of the frequency of the resulting termination triggered emission is controlled by the amplitude at termination. Higher amplitude pulses trigger risers, while lower amplitude pulses trigger fallers. An extension of this effect is that the amplitude of the pulse at termination determines the amount of additional wave energy that is generated; as the termination level increases, the length of the pulse is extended.

- The nonlinear growth rate appears to be affected by two parameters: the proximity to the threshold level and the cold electron density. Holding the signal amplitude close to the threshold level results in a reduction of the nonlinear growth rate relative to the growth rate of a signal that is significantly above the threshold level. This effect is very sensitive; a 5 dB increase in the amplitude of the input signal results in a full transition between an unamplified signal and a signal that grows at the full growth rate. This result is consistent with the observed rapid onset of amplification after a signal is increased steadily above the threshold level [*Helliwell et al.*, 1980].

5.5   Discussion of the Model

This model for the VLF triggered emission phenomenon is based on resonant particle currents that are calculated directly from a modeled energetic electron distribution function. The model can simulate the saturation process in triggered emissions, showing all of the characteristics experimentally observed at saturation, and allowing an understanding of the saturation process. The model is most similar to the Vlasov Hybrid Simulation [*Nunn*, 1990, 1993], and a discussion of the differences between the two models reveals the advantages and limitations of the model presented here. In particular, the chief advantage of the model specified here is that it does not require the use of an additional (artificial) saturation factor or repeated filtering of the wave field, and that the results may thus be presented without caveats about these artificial factors.

The primary difference between the model presented here and the Vlasov Hybrid Simulation is the way in which the evolution of the distribution function is tracked. The Vlasov Hybrid code randomly seeds the modeled phase space with "simulation particles" and tracks the evolution of these individual particles, always referring back to the original distribution (thus, the Vlasov Hybrid model also contains an appeal to Liouville's Theorem). As a result, the simulation particles act as a random sampling of the distribution function in phase space. In contrast, the model here samples the distribution function at predetermined, regular phase-space intervals, referring back

to the distribution from the previous time step. This method allows interpolation over a regular grid, as compared to interpolation over a the set of irregularly spaced simulation particles in the Vlasov Hybrid code. This use of a regular grid in interpolation greatly reduces the computational noise.

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 phase-space dimensionality of the simulation) with greater than 99.5% probability. However, with one million grid points, the density law does not ensure coverage for some 5000 points, requiring additional interpolation over the grid to resolve these values [*Nunn et al.*, 2005]. Furthermore, there are undoubtedly a number of points that rely on 1-2 particles to generate a value during the interpolation. In comparison, the scheme presented here always has 2^n values with which interpolation is performed.

The advantage that the Vlasov Hybrid has over the approach used here is that it can address a pathologically disturbed distribution. That is, 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 the case of the model presented here, 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.

Additionally, the code used here 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, there is no re-centering step in which the range of modeled parallel velocities and range of allowed k values are adjusted according to the distribution of k 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 Equation (5.1) and Equation (5.2). Specifically, the deriviation requires the assuption that $\partial\phi/\partial t \ll \omega$, which can be violated if the frequency of the emission difers significantly from the initial wave frequency. Additionally, at

saturation the spatial phase structure (i.e., $\phi(z)$) of the wave becomes very fine, and the choice of Δz may not be adequate to model the full frequency evolution of the wave, stunting the growth of off-frequency components.

One limitation that both codes share is the assumption of the monochromatic nature of the input wave. At any point in the wave space, there is only one wave number defined. Thus, the models 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 a frequency domain treatment may be appropriate.

Chapter 6

Diffusion Modeling of VLF Triggered Emissions

The presence of wideband waves at saturation (see Chapter 4) implies that at least a portion of the wave-particle interaction of the VLF triggered emission instability involves waves that are not purely monochromatic. The interaction between a single-frequency, whistler mode wave and counter-streaming electrons in cyclotron resonance with the wave is deterministic in nature, as shown in Chapter 3. However, the interaction of counter-streaming electrons with wideband wave components is stochastic, and is typically modeled as a random-walk process, which results in a diffusion of electrons in phase space along contours of whistler-mode interaction (see Figure 3.5) [*Kennel and Petschek*, 1966].

Thus, the existence of wideband wave energy at saturation indicates that at least a portion of the wave-particle interaction can be modeled using quasilinear pitch angle diffusion. This treatment is attractive for a number of reasons. First, pitch angle diffusion operates on the whole distribution function, and the interaction with individual particles does not need to be tracked, as the application of Liouville's theorem in Chapter 5 requires. This neglect of individual particle motion allows for a significant savings in computational overhead. The diffusion treatment also enables the specification of the inputs to the model, as well as the expression of results from the model, in terms of properties of the total distribution function, such as anisotropy.

These global specifications of the properties of the distribution are intuitively easier to model, calculate, and understand. Additionally, pitch angle diffusion is insensitive to the instantaneous value of ζ, the angle between the electron velocity vector and the wave magnetic field vector; diffusion coefficients are calculated based on the resonance condition (see Equation (6.3)). Thus, the modeling of the wave phase and distribution gyrophase is not necessary, resulting in a reduction of the dimensionality of the model. These simplifications produce a model that is computationally lean, allowing for quick repetition to explore the dynamics of the process.

It should be noted, however, that the use of pitch angle diffusion to model the wave-particle interaction is strictly not justified. The half-power bandwidth of the wave at saturation in Figure 4.6 is $\sim 5-10$ Hz. For that event, the equatorial resonant velocity is $\sim 1.5 \times 10^7$ m/s and the interaction region is ~ 4000 km in length. Although a given particle experiences many small scattering events within the modeled region due to the tails of the wideband frequency spread, the particle experiences $\sim 1-4$ significant scattering events due to the highest amplitude portion of the wave spectrum. Consequently, the actual physical process only vaguely resembles the random walk necessary for a rigorous diffusion treatment. Thus, the results from this modeling effort are at best illustrative, but do serve as a touchstone for understanding how the results of a non-self-consistent diffusion model differs from the results of fully self consistent models, such as the one described in Chapter 5. This understanding is useful since diffusion models are frequently used to calculate radiation belt lifetimes, while coherent interaction models are required for modeling the VLF triggered emission instability.

The discussion in the remainder of this chapter begins by developing the diffusion model used to treat the scattering of particles by the wideband waves developed at saturation. In addition, a qualitative heuristic model is needed to generate the nonlinear growth aspect of the VLF triggered emission instability, so one is proposed based on the calculated linear growth rate of the modified distribution function. The model is then applied to simulate the 8 December 1986 event in Figure 4.1, and conclusions are drawn about the validity and usefulness of such a diffusion-based approach.

6.1 Model Description

The proposed diffusion model for the wave-particle interaction at saturation in the VLF triggered emission instability is simple, encapsulating only the processes that are important to the application of a diffusion treatment. The particle interactions with the wideband wave energy are modeled using pitch angle diffusion, and nonlinear effects, such as wave growth and frequency spread, are modeled empirically. By excluding the complicating factors of individual particle scattering (and consequently, particle trapping) and resonant current generation, two benefits are gained. First, the size of the computation is reduced, enabling quick examination and rapid iteration of the process. Second, by only modeling a small portion of the overall process, the relative contribution of diffusion processes can be determined.

6.1.1 Pitch Angle Diffusion

As in the full treatment in Chapter 5, the interaction region is modeled spatially in the z-dimension. At every point z, the waves are defined only in terms of the power spectral density, $P_z(\omega)$ as a function of angular wave frequency, ω. The electron distribution function, $F_z(v_{||}, v_\perp)$, is defined in terms of the electron velocity parallel and perpendicular to the ambient magnetic field, $v_{||}$ and $v_\perp$, respectively.

In the following discussion, the notation $F_z\big|_{i,j}$ is used to denote the value of the electron distribution function at the ith parallel and the jth perpendicular velocity values, respectively, at position z. That is, $F_z\big|_{i,j} = F_z(v_{||}[i], v_\perp[j])$. Similarly, for the diffusion coefficient, $D_z\big|_{i,j} = D_z(v_{||}[i], v_\perp[j])$.

With this treatment, the particles are scattered in velocity space by the wideband wave energy using the quasilinear pitch angle diffusion equation [*Walt*, 1994, ch. 7] at each time step,

$$\frac{\partial F}{\partial t} = \frac{1}{v_\perp} \left[v_\perp \frac{\partial}{\partial v_{||}} - (v_{||} + v_\mathrm{p}) \frac{\partial}{\partial v_\perp} \right] \left\{ D v_\perp \left[v_\perp \frac{\partial F}{\partial v_{||}} - (v_{||} + v_\mathrm{p}) \frac{\partial F}{\partial v_\perp} \right] \right\} \qquad (6.1)$$

which is discretized using a second order alternating forward-difference backward-difference technique to yield,

$$
\begin{aligned}
F_z^s\big|_{i,j} = {} & F_z^n\big|_{i,j} \\
& + \frac{\Delta t}{q} \Bigg\{ \frac{1}{2\Delta v_\parallel^2} v_\perp^2[j] \Big[D_z\big|_{i+1,j} \Big(F_z^n\big|_{i+1,j} - F_z^n\big|_{i,j} \Big) \\
& \hspace{3.5cm} + D_z\big|_{i,j} \Big(F_z^n\big|_{i+1,j} - 2F_z^n\big|_{i,j} + F_z^n\big|_{i-1,j} \Big) \\
& \hspace{3.5cm} - D_z\big|_{i-1,j} \Big(F_z^n\big|_{i,j} - F_z^n\big|_{i-1,j} \Big) \Big] \\
& - \frac{1}{2\Delta v_\parallel \Delta v_\perp} \Big[D_z\big|_{i+1,j} \big(v_\parallel[i+1] + v_\mathrm{p} \big) v_\perp[j] \Big(F_z^n\big|_{i+1,j} - F_z^n\big|_{i+1,j-1} \Big) \\
& \hspace{3.5cm} + D_z\big|_{i,j} \big(v_\parallel[i] + v_\mathrm{p} \big) v_\perp[j] \Big(F_z^n\big|_{i,j+1} - 2F_z^n\big|_{i,j} + F_z^n\big|_{i,j-1} \Big) \\
& \hspace{3.5cm} - D_z\big|_{i-1,j} \big(v_\parallel[i-1] + v_\mathrm{p} \big) v_\perp[j] \Big(F_z^n\big|_{i-1,j+1} - F_z^n\big|_{i-1,j} \Big) \Big] \\
& - \frac{1}{2\Delta v_\parallel^2} \frac{v_\parallel[i] + v_\mathrm{p}}{v_\perp[j]} \Big[D_z\big|_{i,j+1} v_\perp^2[j+1] \Big(F_z^n\big|_{i,j+1} - F_z^n\big|_{i-1,j+1} \Big) \\
& \hspace{3.5cm} + D_z\big|_{i,j} v_\perp^2[j] \Big(F_z^n\big|_{i+1,j} - 2F_z^n\big|_{i,j} + F_z^n\big|_{i-1,j} \Big) \\
& \hspace{3.5cm} - D_z\big|_{i,j-1} v_\perp^2[j-1] \Big(F_z^n\big|_{i+1,j-1} - F_z^n\big|_{i,j-1} \Big) \Big] \\
& + \frac{1}{2\Delta v_\parallel^2} \frac{v_\parallel[i] + v_\mathrm{p}}{v_\perp[j]} \Big[D_z\big|_{i,j+1} \big(v_\parallel[i] + v_\mathrm{p} \big) v_\perp[j+1] \Big(F_z^n\big|_{i,j+1} - F_z^n\big|_{i,j} \Big) \\
& \hspace{3.5cm} + D_z\big|_{i,j} \big(v_\parallel[i] + v_\mathrm{p} \big) v_\perp[j] \Big(F_z^n\big|_{i,j+1} - 2F_z^n\big|_{i,j} + F_z^n\big|_{i,j-1} \Big) \\
& \hspace{3.5cm} - D_z\big|_{i,j-1} \big(v_\parallel[i] + v_\mathrm{p} \big) v_\perp[j-1] \Big(F_z^n\big|_{i,j} - F_z^n\big|_{i,j-1} \Big) \Big] \Bigg\} \quad (6.2)
\end{aligned}
$$

where the superscript n represents the distribution at the current time step, and the superscript s represents the distribution after the scattering has been calculated. The diffusion coefficient increases as the wave grows, and the stability of the difference equation, Equation (6.2), depends on the value of the diffusion coefficient. Therefore, the factor q is introduced to ensure stability of the calculation, and the scattering update is iterated q times each time step, effectively decreasing the size of the time step used in the diffusion calculation. The value of q is automatically increased in accordance with the value of the diffusion coefficient. Thus, at z-locations with larger

wave amplitudes (and corresponding diffusion coefficient values) more iterations of the diffusion update are required than locations with smaller wave amplitudes. The use of q to ensure computational stability is justified as long as the diffusion coefficient, $D_z\big|_{i,j}$, changes slowly on time scales of Δt. $D_z\big|_{i,j}$ is calculated from [*Walt*, 1994, ch. 7]

$$D_z\big|_{i,j} = \frac{1}{2}\left(\frac{q}{m_e}\right)^2 \frac{1}{1 + v_{||}[i]/v_g(z, \omega_{\text{res},i})}\left(1 - \frac{v_p(z, \omega_{\text{res},i})v_{||}[i]}{v_{||}^2[i] + v_\perp^2[j]}\right)^2 P_z^n(\omega_{\text{res},i}) \quad (6.3)$$

where $v_g(z, \omega_{\text{res},i})$ is the group velocity of the wave and $v_p(z, \omega_{\text{res},i})$ is parallel velocity of the wave at z with frequency $\omega_{\text{res},i}$. $\omega_{\text{res},i}$ is the wave frequency that is resonant with particles that have parallel velocity $v_{||}[i]$, respectively. $P_z^n(\omega_{\text{res},i})$ is the power spectral density of the wave at that frequency, and is determined using an exponential frequency spread,

$$P_z^n(\omega_{\text{res},i}) = P_z^n(\omega_0)A_1 \exp\left(-s_1\frac{|\omega_{\text{res},i} - \omega_0|}{\sqrt{B_w}}\right) \quad (6.4)$$

where ω_0 is the frequency of the transmitted wave. The factor s_1 is a constant, the value of which is determined by measuring the frequency spread as seen in the data, such as that in Figure 4.6. Similarly, A_1 is an empirical constant relating the frequency spread to the wave power at ω_0.

The waves and particles are propagated up and down the field line, respectively, using a one-dimensional convection update. The wave power spectral density at the next time step $n + 1$ is calculated as

$$P_z^{n+1}(\omega) = P_z^g(\omega) + v_g(z, \omega)\frac{\Delta t}{\Delta z}P_{z-1}^g(\omega) \quad (6.5)$$

where the superscript g designates the power spectral density following the application of the wave growth update (see Section 6.1.2), and $v_g(z, \omega)$ is the group velocity. Similarly, the update equation for the distribution function is

$$F_z^{n+1}(v_{||}, v_\perp) = F_z^s(v_{||}, v_\perp) + v_{||}\frac{\Delta t}{\Delta z}F_{z+1}^s(v_{||}, v_\perp) \quad (6.6)$$

where the superscript s designates the electron distribution function following the application of the quasilinear scattering. The temporal spacing, Δt, is selected to ensure that the dynamics of the saturation process are well captured; in other words, Δt is chosen so that the diffusion coefficient D_z does not change significantly due to wave growth during a single time step, so that the use of the q-factor is justified. For a given Δt, the mesh spacing Δz is selected to ensure Courant stability for the maximum v_{g} or $v_{\parallel}$ modeled in the simulation.

6.1.2 Wave Growth

Without the modeling of the gyrophase dependence of the distribution function, resonant current formation cannot be calculated. Thus, modeling the effect of wave growth requires the use of the mathematically simplest formulation that captures the effect of nonlinear growth. *Karpman et al.* [1974] found that the nonlinear growth increases in accordance with the sign of the linear growth rate. Therefore, the waves at each time step are taken to be amplified or attenuated according to

$$P_z^g(\omega) = G_z^n P_z^n(\omega) \tag{6.7}$$

where the superscript g represents the wave power spectral density after growth, and G_z^n is the contribution of nonlinear growth during time step n, calculated using the exponential update equation

$$G_z^n = G_z^{n-1} + g_{\mathrm{nl}} \, \gamma_{\mathrm{lin}}(z, \omega_0)\Delta t \tag{6.8}$$

where $\gamma_{\mathrm{lin}}(z, \omega_0)$ is the linear growth rate calculated from the distribution function using Equation (3.10), and g_{nl} is a proportionality constant relating the linear growth rate to the nonlinear growth rate. The linear growth rate $\gamma_{\mathrm{lin}}(z, \omega_0)$ is calculated as an average weighted by the relative proportion of trapped particles in the neighboring velocity bins. The value of g_{nl} is selected to give nonlinear growth rates that are comparable to those observed, and is held constant throughout the simulation. The value of the nonlinear growth contribution G_z^n is reset to unity if the value of the

linear growth rate goes negative at any point before the wave enters the growth region. Since the wave is thus attenuated prior to entering the growth region, it cannot reach the required amplitude for trapping and subsequent nonlinear growth, which was demonstrated in Chapter 5 to be caused by the transition to a trapping state. The weighting of the calculated linear growth rate $\gamma_{\text{lin}}(z, \omega_0)$ as well as the reset of G_z^n are necessary to account for the fact that the effect of particle trapping is not explicitly encapsulated in the model [*Nunn*, 1974].

6.2   Results

Using the above model, an attempt is made to quantitatively simulate the saturation phenomena observed in the 8 December 1986 event. For the functional form of the initial electron distribution, F_0, the distribution from *Bell et al.* [2000] is again used, which is fit to an average of many measurements with the HYDRA instrument on the POLAR spacecraft. The total flux is adjusted for a reasonable, but arbitrary, equatorial linear growth rate of 150 dB/s at the resonant frequency. The linear growth rate selection becomes arbitrary due to the effect of the adjustable parameter g_{nl}; a lower linear growth rate would lead directly to a higher value for g_{nl}, and *vice versa*.

The input wave is a monochromatic wave at 2.7 kHz with input amplitude of 2 pT. The amplitude is significantly higher than the range of input amplitudes considered in Chapter 5. This difference is due to the fact that the diffusion based model does not consider the effect of linear growth, which acts to amplify the wave from the input of the modeled region as it propagates toward the equator. As a result, the input amplitude must be equivalently higher to get the amount of scattering at the equator necessary to cause saturation at amplitudes that are consistent with empirical measurements.

The calculation is performed at $L = 4.7$ using the cold electron density defined by *Carpenter and Anderson* [1992]. The magnetic field is modeled as a centered dipole. The modeled spatial region is from $z =$[-2000 km, 4000 km], which contains the entire trapping region throughout the growth and saturation cycle. The electron

distribution was modeled in the range $v_{||}$ =[500 eV, 1500 keV] and $v_{\perp}$ =[1keV, 50 keV], which contains the resonant energy range and the entire spreading bandwidth of the wave. The temporal spacing was selected to be $\Delta t = 0.001$ s, which is a factor ~ 50 shorter than any of the oscillations observed at or near the transmitted frequency. Additionally, the modeled event shows a maximum instantaneous nonlinear growth rate of approximately 35 dB/s. Since the diffusion coefficient varies as the square of the wave amplitude, the maximum change of the diffusion coefficient magnitude within a single time step is 0.07 dB, or 0.8%, meaning that the use of the q-factor is justified given this timestep size. The value of the phenomenological growth constant g_{nl} is 0.0205.

The value for the spreading constant, s_1 is determined from the superposed epoch data from Figure 4.6. For each temporal point in the epoch, the slope of the frequency spread is determined from a least squares fit. Figure 6.1 shows the resulting slope values as a function of $B_{\mathrm{w}}^{-1/2}$. Since the low amplitude frequency spread measurements are based on fewer data points, they display larger variance than the high amplitude frequency spread measurements, as seen in Figure 6.1. To account for this heteroscedascity, a weighted fit is used to estimate s_1. Before the empirical value of s_1 can be fully determined, it should be noted that the data in Figure 4.6 is on an unnormalized decibel scale, and must first be calibrated to in-duct wave amplitudes. Thus, the above method can estimate the empirical value of s_1 as a function of the average maximum wave amplitude, as shown by the solid blue line in Figure 6.2. Finally, the model is run with multiple values of s_1, yielding the modeled average maximum wave amplitude as a function of s_1, also shown in Figure 6.2 as the dashed green line. The intersection of the two curves represents the "best" value of s_1, i.e. the value at which the frequency spread and saturation level most closely match those seen in the data. The value of s_1 used was 4.31×10^{-7}, giving a maximum average amplitude of 73.2 pT. Combining this inferred value of the in-duct wave amplitude with the results from Figure 4.5 implies (within the caveats of the diffusion formulation) that the trapping motion of particles with pitch angles $\alpha \gtrsim 45$ deg is responsible for the generation of the wideband wave energy observed in Figure 4.1(a). Note also that the saturation amplitude is consistent with the values determined in Chapter 5.
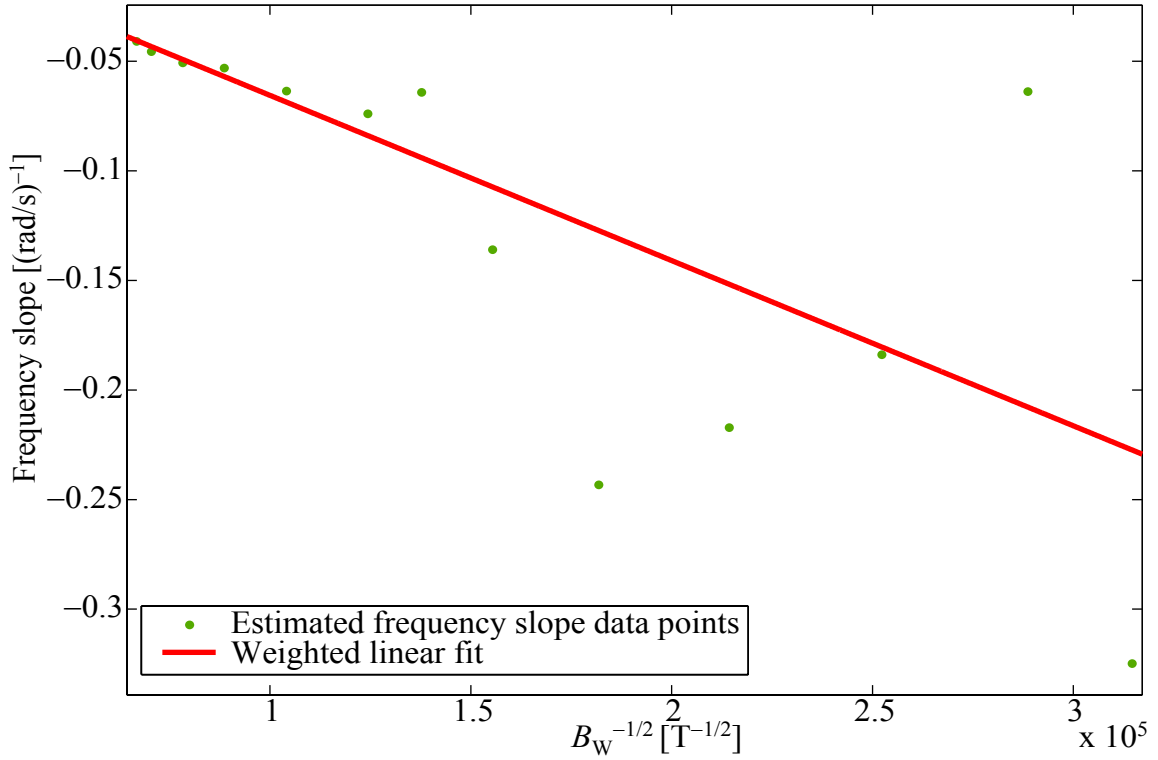
Figure 6.1: Measurement of the rate of amplitude fall-off with frequency as a function of $B_w^{-1/2}$. Individual points in the plot are minimum least squares estimates of the amplitude fall-off rate for each point in the superposed epoch of Figure 4.6. The fitted line represents the weighted least squares fit of the data, where the weights are taken from the standard errors of each of the amplitude fall-off rate estimates. The wideband spreading constant, s_1, is determined from the slope of the fitted line and the in-duct wave amplitude.

Figure 6.3(a) shows the results of the simulation in the form of the amplitude output of the model. The amplitude is integrated over a 131 Hz band around the center frequency of 2700.121 so that its values can be compared to the filtered amplitude measurement from Figure 4.1. The absolute maximum amplitude corresponds to 94.2 pT (which is a ∼2 dB difference from the average of the maximum amplitudes of 73.2 pT). The amplitude plot also shows the characteristic shape of the long period oscillations. The short period oscillations are notably absent.

Figure 6.3(a) shows a direct comparison between the model output and the data from the 8 December 1986 event, demonstrating good agreement, although short
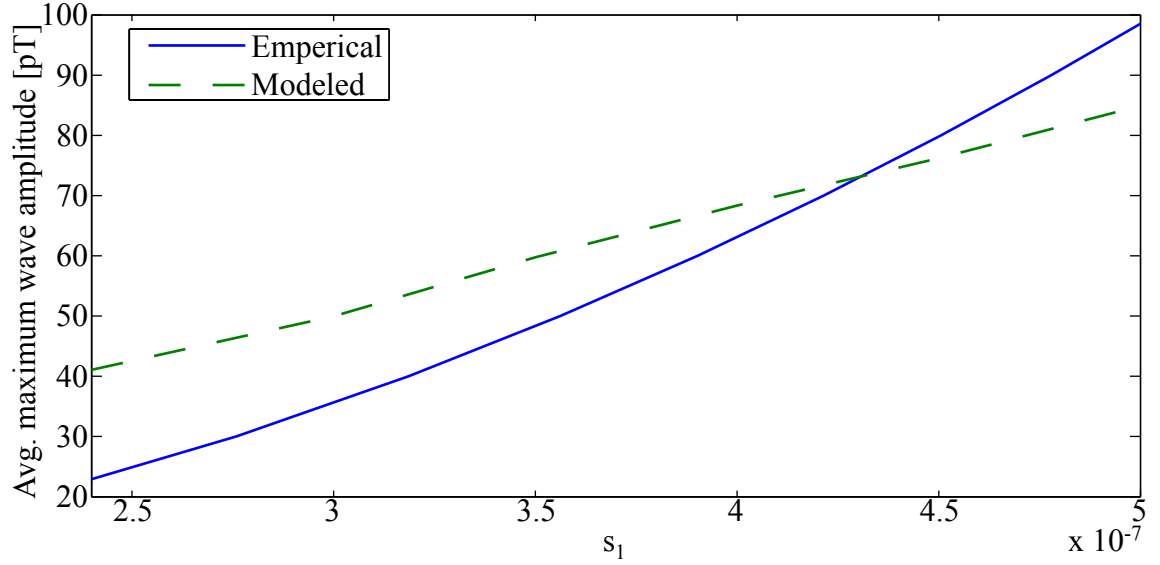
Figure 6.2: Behavior of the wideband spreading constant, s_1, versus average maximum amplitude of long period oscillations for the 8 December 1986 event. The empirical trace shows the measured value of s_1, calibrated to the assumed in-duct average maximum amplitude. The modeled trace shows the average maximum amplitude as determine by the model for selected values of s_1. The two traces cross at $s_1 = 4.31 \times 10^{-7}$, which corresponds to an average maximum amplitude of 73 pT.

period oscillations are not reproduced. It should be noted that the actual period of the long-period oscillations as observed on 8 December varies between 1.96 s and 2.39 s. To achieve this same period variation from the model, the energetic electron flux input is adjusted as a function of time, as shown in Figure Figure 6.3(b). This variation of the energetic electron flux causes a corresponding change in the linear growth rate and nonlinear growth rate, since the two are directly related in the model. The changing growth rate then controls the long-period oscillation period by changing the length of time for the wave to grow to saturation.

6.3   Conclusions

The observations in Chapter 4 detail the characteristics of the triggered emission process at saturation, including the generation of long and short-period oscillations
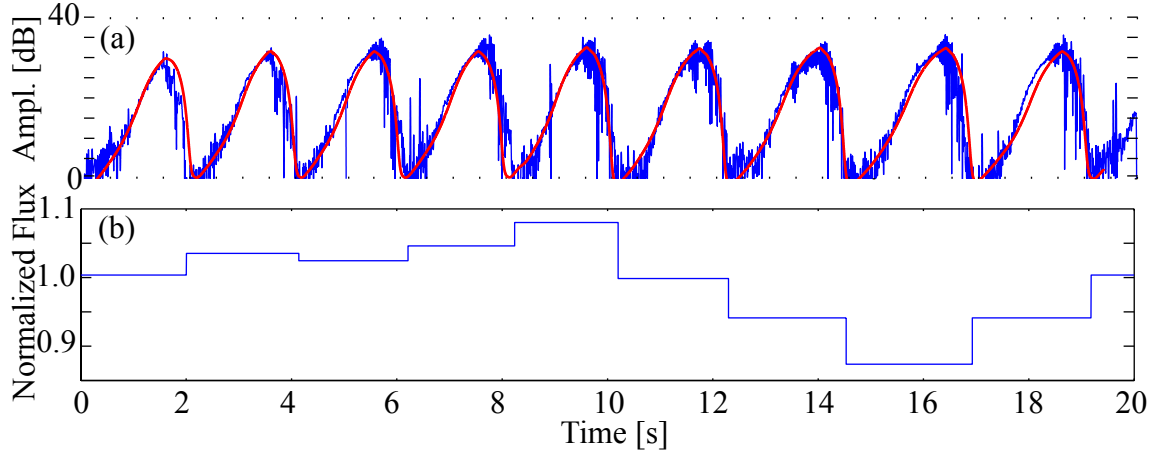
Figure 6.3: Output of the saturation model. (a) Amplitude output (red) showing characteristic shape of long-period oscillations, with good agreement with the event from Figure 4.1 (blue). Note that short period oscillations and sidebands are not reproduced. (b) Energetic electron flux variation used to reproduce the variation in the long-period oscillations seen in the data.

associated with saturation. In addition, it is observed that the generation wideband (∼100 Hz) incoherent radiation is associated with the trapping of particles in the potential well of the coherent triggering wave. From that observation, the above model is developed around the basic premise that these incoherent waves may be responsible for the destruction of the anisotropy in the energetic electron distribution and thus bring about the saturation of the growing waves.

Despite the simplicity of the diffusion-based model, it is able to reproduce a remarkable number of qualitative and quantitative characteristics of the data. By modeling the generation of broadband waves and their subsequent interaction with the electron distribution, the model can replicate other aspects of the saturation process, including the shape of the long-period saturation oscillations, the timing of the oscillations, and wave amplitudes at saturation.

The simple nature of the model does not support conclusions regarding the nature of the input energetic electron distribution. Because of the selection of g_{nl} to give the required growth rates, the only requirement on the electron distribution is that it is anisotropic in a range of parallel velocities near that resonant with the input

wave. Macroscopic changes in the magnitude of anisotropy or total energetic electron flux are masked by the choice of g_{nl}. For example, if one were to increase F_0 for the simulation of the 8 December event by 10%, the only detectable difference in the results would be an equivalent 10% reduction of g_{nl}. However, g_{nl} is held constant throughout the simulation. Once g_{nl} is specified, the only remaining free parameter is F_0. Therefore, a conclusion can be drawn that the variation in the period of the long-period oscillations must be accompanied by some change in the input flux. In the case of this simulation, the variation is attributed to an overall variation in the energetic electron flux, but mathematically the period variation could also be brought about by a change in the electron anisotropy. However, as the primary cause of the electron anisotropy is convection from the magnetotail (the region of the magnetosphere which extends away from the sun), it is considered less likely to vary on the observed timescales.

By necessity, there is one phase-specific effect that must be modeled heuristically, that of the nonlinear growth due to particle trapping. Without growth, there can be no saturation. However, the simplest possible model is chosen for that effect based on the theoretical considerations of *Karpman et al.* [1974]: nonlinear growth proportional to the free energy available for growth in the calculational form of the linear growth rate. In effect, that leaves only one adjustable parameter, the constant of proportionality g_{nl}, available to make the model replicate the data.

One limitation of the diffusion model is that it does not replicate the short-period amplitude oscillations observed in Chapter 4. As noted in Chapter 5, the cause of the short-period amplitude oscillations is the brief disruption and renewal of the resonant current that is enhanced by the de-trapping electron hole. Since the nonlinear growth accumulation parameter, G_z, is fully reset as soon as the linear growth rate goes negative, the quick renewal of the resonant current enhancement cannot be reproduced using this simple nonlinear growth formulation developed above.

There are two major limitations of the diffusion model that relegate it to the status of being no more than a curiosity. First, the nearly-monochromatic nature of the waves, even at saturation, makes a diffusion treatment of the problem invalid.

Second, the more general model in Chapter 5 shows that there is no causal relationship between wideband waves and saturation. In fact, the relationship is reversed; wideband waves are most significant when the wave is suppressed near saturation. With an understanding of these deficiencies of the diffusion based model, it is hard to ascribe any predictive power to the model. However, with some manipulation (i.e., the specification of the heuristic growth formulation [Equation (6.7)] and wave frequency spread [Equation (6.4)]), the model can return values of the long period oscillation and saturation amplitude that are in the range of values expected from the experimental observations.

Chapter 7

Summary and Suggestions for Future Work

This dissertation is a complete work, representing a significant contribution toward the understanding of the VLF triggered emission instability. Specifically, by examining the aspects of the instability associated with satruation, this work provides a unique and novel explanation for the underlying physical mechanism behind the nonlinear growth and saturation effects in the instability. As demonstrated herein, the physical mechanism for the growth and saturation of the instabiltiy lies in the spatial transition from a state where the wave interacts weakly with the plasma to one where the wave can organize the plasma into two populations. Effectively, one population is energized and then allowed to mix with the other, creating a nonlinear, unstable interaction.

The summary of the historical, observational, and theoretical aspects provided by this document is unified by this new understanding of the mechanism of the instability. That is, the complex observational results from wave injection experiments are condensed under this unifying framework, allowing for communication of the characteristics of the phenomenon without a great deal of nuance. As such, the discussion presented is competely approachable for a reader unfamiliar with the topic. Directly, the new, detailed observations of the saturation phenomenon presented in Chapter 4

appear to fit into the previous experimental framework in a manner so straightforward that the obesations seem to follow directly from the discussion in Chapter 2 and Chapter 3.

This new, unifying understanding of the instability has its genesis in the results of the model developed in Chapter 5. The model itself is grown out of an understanding of the Vlasov Hybrid Simulation [*Nunn*, 1990, 1993], with specific steps taken to overcome the limitations of this older formulation. As a result, the algorithm and implimentation of the new model are simple enough to belie its superior capability in modeling the instability.

Despite the contributions of this dissertation, there is still much to be done with regard to the VLF triggered emission instability. Aspects of the instability remain unaddressed include the triggering of the free-running emissions, relating the instability to natural activity, and simulating multi-wave phenomena.

7.1 Triggering Free-Running Emissions

The detailed mechanism behind the triggering of the free-running plasma emission during the evolution of the instabiltiy remains unexplained. Based on the considerations presented in this dissertation, it is likely that the emission is spawned by a self-sustaining generation region composed of very steep wave amplitude gradients. In particular, the direction of the frequency sweep of the transmission is determined by the sense of the inhomogeneity at the point of detrapping. This specification is similar to the argument proposed by *Helliwell* [1967] and *Roux and Pellat* [1978], but instead of a single bunch of electrons radiating as they stream down the field line (and correspondingly lose their coherence due to any initial spread in velocities), the generation region is relatively stationary (compared to the amount of motion cited by *Helliwell* and *Roux and Pellat*), with a continuous stream of de-trapping electrons rotating out of resonance with the wave due to the effect of the inhomegeneity. The nature of the inhomogeneity (i.e. increasing or decreasing magnetic field strength) determines the sense of relative gyrophase rotation of the detrapping stream, and hence the rising or falling frequency nature of the emission.

This theoretical discussion, however, must be confirmed, and is currently beyond the capability of the model defined in Chapter 5. Specifically, the emissions generated by the model do not endure for long, even on pulse termination. More explicitly, the model does continue to produce wave energy near the equator, but the wave energy rapidly tapers to a level 40-50 dB below the saturation level, which is inconsistent with experimental results. Two aspects of the model may be responsible for this result and warrant further study. First, the spatial phase structure at saturation or pulse termination can become very fine. Thus, the spatial step size may be too coarse to accurately sample the steep wave amplitude gradients within the generation region. Second, the derivation of the wave update equation in Appendix C neglects many terms in the use of the monochromacity assumption. Some of these terms, particularly the higher order derivative terms and the spatial derivative terms of the phase of the resonant current may be responsible for the set up and maintenance of the self sustaining generation region. Thus, the modeling and subsequent explanation of the free-running triggered emissions may require running the code with a much finer time step, or re-deriving the update equation to include the second order terms that may become important at the point of triggering, when steep wave amplitude gradiants are established.

7.2   Relation to Natural Activity

One of the observational aspects of the VLF triggered emission instability that has not been quantified is the relation of the instability to the presence of natural noise. There has been some work showing that the presence of a hiss band can be particularly advantageous for activating the instability, and the effect of the hiss band on the energetic electron distribution helps to explain the fine frequency-selectivity of the phenomenon [*Trakhtengerts et al.*, 2001; *Pasmanik et al.*, 2002]. These results are also consistent with the anecdotal operational use of the top of a hiss band to encourage Siple reception (see secrefoperations). However, this effect has never been studied or quantified. Such an experimental study would help to limit the conditions under which the energetic electrons activate the instability.

7.3   Multi-wave Effects

One of the key assumptions in the use of the wave update equation is the assumption of a single wave in the modeled region, i.e., only one k value exists at every point in the region. However, many of the phenomena associated with the instability, such as entrainment, the coherence bandwidth effect, and the triggering of only rising emissions on long transmissions require the interaction between two or more waves. Additionally, *Serra* [1984] showed, using a test particle analysis, that two equal amplitude waves, closely space in frequency, can act to disrupt the trapping process. This effect has yet to be observed in a self-consistent code. However, to model multi-wave effects, the model presented here must be significantly modified. In particular, the wave update equation must be re-derived and discretized without any assumptions of monochromacity, and a frequency domain treatment may be more appropriate, with the caveat that the wave number is not constant across the modeled region, and must be handled appropriately.

Appendix A

Whistler Mode Dispersion

This Appendix derives the whistler mode dispersion relation using linear plasma theory. The whistler mode dispersion relation is first derived assuming a cold plasma with no thermal energy, then derived assuming an energetic distribution of electrons.

A.1   The Cold Plasma Dispersion Relation

The derivation assumes that the waves are propagating in a cold, uniform, magnetized plasma, in the direction along the magnetic field line. The frequency range considered is $\Omega_s \ll \omega < \omega_c$, where Ω_s is the cyclotron frequency for any ion species s. This allows the assumption that the ions in the plasma are stationary during the interactions with the whistler mode wave.

The derivation begins by linearizing the parameters of interest. For example, the total magnetic field in the plasma $\widetilde{\mathbf{B}}$ is divided into two components, a static component $\mathbf{B}_0$ and a small varying component $\widetilde{\mathbf{B}}_{\mathrm{w}}$ due to the action of the whistler mode wave. The total electric field and electron velocities are similarly decomposed,

$$
\begin{aligned}
\widetilde{\mathbf{B}} &= \mathbf{B}_0 + \widetilde{\mathbf{B}}_{\mathrm{w}} & \mathbf{B}_0 &>> \widetilde{\mathbf{B}}_{\mathrm{w}} \\
\widetilde{\mathbf{E}} &= \mathbf{E}_0 + \widetilde{\mathbf{E}}_{\mathrm{w}} & \mathbf{E}_0 &\simeq 0 \\
\widetilde{\mathbf{v}} &= \mathbf{v}_0 + \widetilde{\mathbf{v}}_{\mathrm{w}} & \mathbf{v}_0 &\simeq 0
\end{aligned}
\tag{A.1}
$$

where additional assumptions have been noted. Specifically, the wave magnetic field is small compared to the static magnetic field, the plasma is not influenced by any static electric fields, and there is no bulk motion of the electrons.

Starting with the Lorentz force, applying the linearized quantities with the assumptions from Equation (A.1) yields,

$$m_e \frac{\partial \widetilde{\mathbf{v}}}{\partial t} = -q \left(\widetilde{\mathbf{E}} + \widetilde{\mathbf{v}} \times \widetilde{\mathbf{B}} \right)$$
$$\frac{\partial \widetilde{\mathbf{v}}_w}{\partial t} = -\frac{q}{m_e} \widetilde{\mathbf{E}}_w - \frac{q}{m_e} \left(\widetilde{\mathbf{v}}_w \times \mathbf{B}_0 \right) \tag{A.2}$$

where m_e is the rest mass of the electron and q is the fundamental charge.

The z direction is defined to be along the static magnetic field. That is, $\mathbf{B}_0 = B_0 \hat{\mathbf{z}}$, where $\hat{\mathbf{z}}$ is the unit vector in the z direction. Since the direction of wave propagation is only in that direction, $\mathbf{k} = k\hat{\mathbf{z}}$. Similarly, the variation in the wave quantities is only transverse to the static magnetic field, allowing Equation (A.2) to be reduced to two dimensions transverse to $\hat{\mathbf{z}}$. Assuming that time varying quantities can be expressed in the form $e^{-i\omega t}$, and using the definition of the gyrofrequency, $\omega_c = qB_0/m_e$, Equation (A.2) becomes

$$\begin{bmatrix} -i\omega & \omega_c \\ -\omega_c & -i\omega \end{bmatrix} \widetilde{\mathbf{v}}_w = -\frac{q}{m_e} \widetilde{\mathbf{E}}_w \tag{A.3}$$

in a rectangular coordinate system.

The current carried by the cold electrons, $\widetilde{\mathbf{J}}_c$, and the conductivity tensor $\overleftrightarrow{\boldsymbol{\sigma}}$ are defined by

$$\widetilde{\mathbf{J}}_c = -n_c q \widetilde{\mathbf{v}}_w = \overleftrightarrow{\boldsymbol{\sigma}} \widetilde{\mathbf{E}}_w \tag{A.4}$$

where n_c is the number density of the cold electrons. By inspection of Equation (A.3), the conductvity tensor is

$$\overleftrightarrow{\boldsymbol{\sigma}} = \omega_p^2 \epsilon_0 \begin{bmatrix} \dfrac{-i\omega}{\omega_c^2 - \omega^2} & \dfrac{-\omega_c}{\omega_c^2 - \omega^2} \\ \dfrac{\omega_c}{\omega_c^2 - \omega^2} & \dfrac{-i\omega}{\omega_c^2 - \omega^2} \end{bmatrix} \tag{A.5}$$

where the plasma frequency ω_{p} is defined by

$$\omega_{\mathrm{p}}^2 = \frac{n_{\mathrm{c}} q^2}{m_{\mathrm{e}} \epsilon_0}.$$

With $\mathbf{k}$ in the z direction and the definition of the dielectric tensor in terms of the conductivity tensor, the homogeneous equation becomes

$$\left[\left(1 - \frac{c^2 k^2}{\omega^2} \right) \overset{\leftrightarrow}{\mathbf{I}} - \frac{\overset{\leftrightarrow}{\boldsymbol{\sigma}}}{i \omega \epsilon_0} \right] \widetilde{\mathbf{E}}_{\mathrm{w}} = 0 \tag{A.6}$$

where $\overset{\leftrightarrow}{\mathbf{I}}$ is the identity matrix. Inserting the value of the conductivity tensor from Equation (A.5) into Equation (A.6), the matrix equation

$$\begin{bmatrix} 1 - \dfrac{c^2 k^2}{\omega^2} - \dfrac{\omega_{\mathrm{p}}^2}{\omega_{\mathrm{c}}^2 - \omega^2} & \dfrac{-\omega_{\mathrm{c}} \omega_{\mathrm{p}}^2}{\omega_{\mathrm{c}}^2 - \omega^2} \\ \dfrac{\omega_{\mathrm{c}} \omega_{\mathrm{p}}^2}{\omega_{\mathrm{c}}^2 - \omega^2} & 1 - \dfrac{c^2 k^2}{\omega^2} - \dfrac{\omega_{\mathrm{p}}^2}{\omega_{\mathrm{c}}^2 - \omega^2} \end{bmatrix} \widetilde{\mathbf{E}}_{\mathrm{w}} = 0$$

is obtained. Taking the determinant of the matrix to find the non-trivial solution yields a relation for k and ω,

$$1 - \frac{c^2 k^2}{\omega^2} + \frac{\omega_{\mathrm{p}}^2}{\omega \left(\omega_{\mathrm{c}} \pm \omega \right)} = 0$$

and noting that in the range of frequencies of interest (i.e. $\omega < \omega_{\mathrm{c}}$) a real-valued solution for the index of refraction ck/ω requires the selection of the minus sign in the denominator of the last term. Furthermore, for typical magnetospheric conditions, the magnitudes of the last two terms are significantly greater than unity, giving the whistler dispersion relation,

$$D(k, \omega) \simeq \frac{c^2 k^2}{\omega^2} - \frac{\omega_{\mathrm{p}}^2}{\omega \left(\omega_{\mathrm{c}} - \omega \right)} = 0 \tag{A.7}$$

A.2 The Energetic Electron Dispersion Relation

The addition of a plasma component with super-thermal energy significantly complicates the deriviation of the dispersion relation. Specifically, the application of the Lorentz force is no longer generic to every particle in the sense that $\mathbf{v}_0 \neq 0$ for every particle. Thus, the derivation is performed using Vlasov's equation, which governs the interaction between electromagnetic fields and a disribution of collisionless particles. In particular, the energetic electron distribution function F is decomposed into a static component, F_0, and a component that varies due to the influence of the whistler mode wave, F_{w}, such that

$$F(\mathbf{v}) = F_0(\mathbf{v}) + F_{\mathrm{w}}(\mathbf{v}).$$

Using the linearized fields from Equation (A.1), where the contribution of the wave magnetic field is no longer assumed to be negligible, the Vlasov equation can be expanded into zeroth- and first-order equations. That is the Vlasov equation

$$\frac{\partial F}{\partial t} + \mathbf{v} \cdot \boldsymbol{\nabla} F - \frac{q}{m_{\mathrm{e}}} \left(\widetilde{\mathbf{E}} + \mathbf{v} \times \widetilde{\mathbf{B}} \right) \cdot \boldsymbol{\nabla}_{\mathbf{v}} F = 0$$

becomes

$$\widetilde{\mathbf{v}} \times \mathbf{B}_0 \cdot \boldsymbol{\nabla}_{\mathbf{v}} F_0 = 0 \tag{A.8}$$

and

$$\frac{\partial F_{\mathrm{w}}}{\partial t} + \mathbf{v} \cdot \boldsymbol{\nabla} F_{\mathrm{w}} - \frac{q}{m_{\mathrm{e}}} \left(\mathbf{v} \times \mathbf{B}_0 \right) \cdot \boldsymbol{\nabla}_{\mathbf{v}} F_{\mathrm{w}} - \frac{q}{m_{\mathrm{e}}} \left(\widetilde{\mathbf{E}}_{\mathrm{w}} + \mathbf{v} \times \widetilde{\mathbf{B}}_{\mathrm{w}} \right) \cdot \boldsymbol{\nabla}_{\mathbf{v}} F_0 = 0 \tag{A.9}$$

where $\mathbf{v}$ represents the electron velocity vector. Noting that $\mathbf{B}_0 = B_0 \hat{\mathbf{z}}$, and expressing the electron velocity $\mathbf{v}$ in terms of cylindrical velocity space coordinates $(v_{\|}, v_{\perp}, \zeta)$, where $v_{\|}$ is the electron velocity antiparallel to the magnetic field, $v_{\perp}$ is the electron velocity perpendicular to the magnetic field, and ζ is the gyrophase angle between the electron's perpendicular velocity vector and the wave's magnetic field vector, Equation (A.8) becomes

$$B_0 \frac{\partial F_0}{\partial \zeta} = 0 \tag{A.10}$$

which simply states that the static distribution function must be uniform in gyrophase. Assuming that spatial and temporal variations can be written in the form $e^{-i(\omega t - kz)}$, such that $\mathbf{k} = k\hat{\mathbf{z}}$, Equation (A.9) can be reduced to

$$\omega_{\mathrm{c}}\frac{\partial F_{\mathrm{w}}}{\partial \zeta} - i(kv_{\parallel} + \omega)F_{\mathrm{w}} = \frac{q}{m_{\mathrm{e}}}\left[\left(1 + \frac{kv_{\parallel}}{\omega}\right)\widetilde{\mathbf{E}}_{\mathrm{w}} - \frac{k}{\omega}\left(v_{\perp}E_{\mathrm{w}}\sin\zeta\right)\hat{\mathbf{z}}\right]\cdot\boldsymbol{\nabla}_{\mathbf{v}}F_0$$

$$= \frac{q}{m_{\mathrm{e}}}\frac{kE_{\mathrm{w}}}{\omega}\left[-\left(\frac{\omega}{k} + v_{\parallel}\right)\frac{1}{\sin\zeta}\frac{\partial F_0}{\partial v_{\perp}} + v_{\perp}\sin\zeta\frac{\partial F_0}{\partial v_{\parallel}}\right] \quad (A.11)$$

And the solution for Equation (A.11) can be shown to be (using the more general equation (9.3.12) of *Gurnett and Bhattacharjee* [2005] with a coordinate transform into the frame of the whistler mode wave)

$$F_{\mathrm{w}} = \frac{qn_{\mathrm{c}}}{m_{\mathrm{e}}\omega_{\mathrm{c}}}\frac{E_{\mathrm{w}}}{2}\left[\frac{-\sin\zeta}{\left(\frac{kv_{\parallel} + \omega}{\omega_{\mathrm{c}}}\right) - 1}\right]\left\{\frac{\partial F_0}{\partial v_{\perp}} - \frac{k}{\omega}\left(v_{\perp}\frac{\partial F_0}{\partial v_{\parallel}} - v_{\parallel}\frac{\partial F_0}{\partial v_{\perp}}\right)\right\} \quad (A.12)$$

where n_{c} is the cold electron density, which is approximately the total electron density. The current can be calculated from the distribution function using

$$\mathbf{J} = -q\int_{-\infty}^{\infty}\int_{0}^{\infty}\int_{0}^{2\pi}\hat{\mathbf{v}}_{\perp}v_{\perp}^2 F_{\mathrm{w}}\,dv_{\parallel}\,dv_{\perp}\,d\zeta \quad (A.13)$$

where $\hat{\mathbf{v}}_{\perp}$ is the unit vector in the direction of the electron perpendicular velocity. Equation (A.13) implies that a current in the energetic electron plasma exists due to the presence of the whistler mode wave. In particular, the ζ-dependence of the distribution function represents a bunching of the electrons that acts to damp or amplify the wave, the magnitude of which depends on the amplitude of the wave field and gradients in the unperturbed distribution. Another point to note is the singularity in the distribution function when $\omega_{\mathrm{c}} = kv_{\parallel} + \omega$, which corresponds to the cyclotron resonance condition (See Section 3.1.2).

The current integral and the definition of the conductivity tensor (Equation (A.4)) are again used in the homogeneous equation to derive the following expression for the

dispersion relation, [*Gurnett and Bhattacharjee*, 2005]

$$D(k,\omega) = 1 - \frac{c^2 k^2}{\omega^2} - \frac{\omega_{\mathrm{p}}^2}{\omega^2 n_{\mathrm{c}}} \int_{-\infty}^{\infty} \int_0^{\infty} \frac{\left(\frac{\omega}{k} + v_{\parallel}\right) \frac{\partial F_0}{\partial v_{\perp}} - v_{\perp} \frac{\partial F_0}{\partial v_{\parallel}}}{v_{\parallel} - \frac{\omega_{\mathrm{c}} - \omega}{k}} \pi v_{\perp}^2 \, dv_{\perp} \, dv_{\parallel} \quad \text{(A.14)}$$

To calculate the whistler dispersion relation in the presence of a population of energetic electrons, F_0 is divided into two components, $F_0 = F_{\mathrm{c}} + F_{\mathrm{h}}$, where F_{c} represents the dominant distribution of cold electrons, defined by

$$F_{\mathrm{c}} = \frac{n_{\mathrm{c}}}{2\pi v_{\perp}} \delta(v_{\perp}) \delta(v_{\parallel})$$

where δ is the Dirac delta function, and F_{h} is the arbirary energetic distribution of interest. Under these conditions and noting that the unity term is negligible, Equation (A.14) reduces to

$$D(k,\omega) = \frac{c^2 k^2}{\omega^2} - \frac{\omega_{\mathrm{p}}^2}{\omega(\omega_{\mathrm{c}} - \omega)} - \frac{\omega_{\mathrm{p}}^2}{\omega^2 n_{\mathrm{c}}} \int_{-\infty}^{\infty} \int_0^{\infty} \frac{\left(\frac{\omega}{k} + v_{\parallel}\right) \frac{\partial F_{\mathrm{h}}}{\partial v_{\perp}} - v_{\perp} \frac{\partial F_{\mathrm{h}}}{\partial v_{\parallel}}}{v_{\parallel} - \frac{\omega_{\mathrm{c}} - \omega}{k}} \pi v_{\perp}^2 \, dv_{\perp} \, dv_{\parallel}$$

$$\text{(A.15)}$$

The evaluation of the intergral in Equation (A.15) requires a contour integration, which yields a real component that is small by a factor of $n_{\mathrm{h}}/n_{\mathrm{c}}$ and an imaginary component, which gives the full expression for the whistler dispersion relation,

$$D(k,\omega) = \frac{c^2 k^2}{\omega^2} - \frac{\omega_{\mathrm{p}}^2}{\omega(\omega_{\mathrm{c}} - \omega)} - i\pi \frac{\omega_{\mathrm{p}}^2}{\omega^2 n_{\mathrm{c}}} \int_0^{\infty} \left[\left(\frac{\omega}{k} + v_{\parallel}\right) \frac{\partial F_{\mathrm{h}}}{\partial v_{\perp}} - v_{\perp} \frac{\partial F_{\mathrm{h}}}{\partial v_{\parallel}} \right] \pi v_{\perp}^2 \, dv_{\perp} \Bigg|_{v_{\parallel} = v_{\mathrm{res}}}$$

$$\text{(A.16)}$$

where

$$v_{\mathrm{res}} = \frac{\omega_{\mathrm{c}} - \omega}{k}$$

corresponds to the singular point in the integral in Equation (A.15). Due to the assumption that the cold electron density is much greater than the hot electron density, the real part of the dispersion relation and the propagation characteristics of the whistler mode wave are unchanged from Equation (A.7).

Appendix B

Single Particle Equations of Motion

This appendix derives the equations of motion for the interaction between a whistler mode wave packet and a single counter streaming electron. The equations of motion are first derived assuming a homogeneous static magnetic field, then the contribution of magnetic field gradients is included. The coordinate system used is defined in Figure 3.4, which is repeated here as Figure B.1.

B.1   Homogeneous Static Magnetic Field

In the case of the homogeneous static magnetic field, the equations of motion are derived directly from the Loretz force.

$$m_e \frac{d\mathbf{v}}{dt} = -q \left[\mathbf{E}_w + \mathbf{v} \times (\mathbf{B}_0 + \mathbf{B}_w) \right] \tag{B.1}$$

where m_e is the rest mass of the electron, q is the fundamental charge, $\mathbf{E}_{rmw}$ and $\mathbf{B}_{rmw}$ are the wave electric and magnetic fields, respectively, $\mathbf{v}$ is the electron velocity vector, and $\mathbf{B}_0$ is the static magnetic field vector. Noting that $|\mathbf{E}_w| = \omega B_w / k$, where $B_w = |\mathbf{B}_w|$, and $\omega_c = q |\mathbf{B}_0| / m_e$, the evaluation of the cross product terms in Equation (B.1) and separation into velocity and spatial components yields [*Dysthe*, 1971; *Inan*, 1977;

Figure B.1: The coordinate system used to model the interaction. The whistler mode wave is propagating along the magnetic field line in the $+z$ direction, which corresponds to the magnetic field direction, while the electron is streaming along the field line in the $-z$ direction. The electron velocity vector is decomposed into components perpendicular and parallel to the magnetic field, with a gyrophase angle described between the perpendicular velocity vector and the wave magnetic field vector.

Carlson et al., 1990]

$$\frac{dz}{dt} = -v_{||} \tag{B.2}$$

$$\frac{dv_{||}}{dt} = -\frac{q}{m_e} B_{\text{w}} v_\perp \sin \zeta \tag{B.3}$$

$$\frac{dv_\perp}{dt} = \frac{q}{m_e} B_{\text{w}} \left(\frac{\omega}{k} + v_{||} \right) \sin \zeta \tag{B.4}$$

$$\frac{d\zeta}{dt} = \left[\omega_{\text{c}} - \omega - k v_{||} \right] + \frac{q}{m_e} B_{\text{w}} \frac{\left(\frac{\omega}{k} + v_{||} \right)}{v_\perp} \cos \zeta \tag{B.5}$$

B.2   Inhomogeneous Static Magnetic Field

Assuming now that the static magnetic field can vary along the direction of the magnetic field, i.e.

$$\frac{\partial B_{0z}}{\partial z} \neq 0,$$

additional magnetic field components must be introduced. Neglecting the azimuthal varation in the magnetic field,

$$\boldsymbol{\nabla} \cdot \mathbf{B}_0 = \frac{1}{r}\frac{\partial}{\partial r}\left(rB_{0r}\right) + \frac{\partial B_{0z}}{\partial z} = 0 \tag{B.6}$$

and considering the spatial region of a gyroradius, assuming that the radial variation in the magnetic field gradient along the direction of the magnetic field is negligible within a gyroradius, the radial component of the magnetic field as a function of radial position is

$$B_{0r} = \frac{1}{r}\int r\frac{\partial B_{0z}}{\partial z}dr = \frac{r}{2}\frac{\partial B_{0z}}{\partial z} \tag{B.7}$$

The variation in electron parallel velocity is then

$$\frac{dv_{\parallel}}{dt} = -\frac{q}{m_{\mathrm{e}}}\left(\mathbf{v}_{\perp} \times B_{0r}\hat{\boldsymbol{r}}\right) = -\frac{q}{m_{\mathrm{e}}}v_{\perp}\frac{r}{2}\frac{\partial B_{0z}}{\partial z} \tag{B.8}$$

where $\hat{\boldsymbol{r}}$ is the unit vector in the radial direction. Determining the value of $dv_{\parallel}/dt$ at the gyroradius,

$$r = \frac{m_{\mathrm{e}}v_{\perp}}{qB_0}$$

and neglecting the small radial component of the static magnetic field (i.e., $B_0 \simeq B_{0z}$) yields

$$\frac{dv_{\parallel}}{dt} = -\frac{v_{\perp}^2}{2B_0}\frac{\partial B_0}{\partial z} = -\frac{v_{\perp}^2}{2\omega_{\mathrm{c}}}\frac{\partial \omega_{\mathrm{c}}}{\partial z} \tag{B.9}$$

The variation in the perpendicular velocity of the electron can be found by noting that the magnetic force cannot change the elctron energy; therefore the total energy

of the electron remains constant throughout the interaction. Thus,

$$\frac{dv}{dt} = 0 = \frac{d}{dt}\left(v_{||}^2 + v_{\perp}^2\right)^{1/2}$$

$$0 = \frac{1}{2v}\left(2v_{||}\frac{dv_{||}}{dt} + 2v_{\perp}\frac{dv_{\perp}}{dt}\right)$$

$$\frac{dv_{\perp}}{dt} = -\frac{v_{||}}{v_{\perp}}\frac{dv_{||}}{dt} = \frac{v_{||}v_{\perp}}{2\omega_{\mathrm{c}}}\frac{\partial\omega_{\mathrm{c}}}{\partial z} \tag{B.10}$$

Inserting Equation (B.9) into Equation (B.3) and Equation (B.10) into Equation (B.4) yields the electron equations of motion, including the effect of the inhomogeneous magnetic field,

$$\frac{dz}{dt} = -v_{||} \tag{B.11}$$

$$\frac{dv_{||}}{dt} = -\frac{q}{m_e}B_{\mathrm{w}}v_{\perp}\sin\zeta - \frac{v_{\perp}^2}{2\omega_{\mathrm{c}}}\frac{\partial\omega_{\mathrm{c}}}{\partial z} \tag{B.12}$$

$$\frac{dv_{\perp}}{dt} = \frac{q}{m_e}B_{\mathrm{w}}\left(\frac{\omega}{k} + v_{||}\right)\sin\zeta + \frac{v_{||}v_{\perp}}{2\omega_{\mathrm{c}}}\frac{\partial\omega_{\mathrm{c}}}{\partial z} \tag{B.13}$$

$$\frac{d\zeta}{dt} = \omega_c - \omega - kv_{||} + \frac{q}{m_e}B_{\mathrm{w}}\frac{\left(\frac{\omega}{k} + v_{||}\right)}{v_{\perp}}\cos\zeta \tag{B.14}$$

Appendix C

Derivation of the Wave Update

This appendix provides a detailed derivation of the wave evolution equation for whistler mode waves in the presence of a resonant current. The discussion follows the derivation provided by *Nunn* [1974], which was modified for the MKS system and clarified by *Omura and Matsumoto* [1982].

The derivation begins with Faraday's law and Ampere's law, where the displacement current is neglected due to the slow-wave nature of the medium,

$$i\frac{\partial \widetilde{\mathbf{E}}_{\mathrm{w}}}{\partial z} = -\frac{\partial \widetilde{\mathbf{B}}_{\mathrm{w}}}{\partial t} \tag{C.1}$$

$$i\frac{\partial \widetilde{\mathbf{B}}_{\mathrm{w}}}{\partial z} = \mu_0 \left(\widetilde{\mathbf{J}}_{\mathrm{c}} + \widetilde{\mathbf{J}}_{\mathrm{r}} \right) \tag{C.2}$$

where the current has been divided into a term represnting the current carried by the whistler mode wave in the cold electrons $\widetilde{\mathbf{J}}_{\mathrm{c}}$ and the resonant current created by the interaction with the near-resonant electrons J_{r}.

From Equation (A.2) and Equation (A.4), an expression for the cold current in terms of the wave electric field can be obtained,

$$\frac{\partial \widetilde{\mathbf{J}}_{\mathrm{c}}}{\partial t} + i\omega_c \widetilde{\mathbf{J}}_{\mathrm{c}} = \omega_{\mathrm{p}}^2 \epsilon_0 \widetilde{\mathbf{E}}_{\mathrm{w}} \tag{C.3}$$

Combining Equation (C.1), Equation (C.2) and Equation (C.3) to eliminate $\widetilde{\mathbf{E}}_{\mathrm{w}}$

and $\widetilde{\mathbf{J}}_\mathrm{c}$, and assuming that the variation of the inhomogeneous medium is small on the scales of a wavelength, i.e.

$$\left|\frac{\partial \omega_\mathrm{c}}{\partial z}\right| \ll k\left(\omega_\mathrm{c} - \omega\right)$$

$$\left|\frac{\partial \omega_\mathrm{p}}{\partial z}\right| \ll k\omega_\mathrm{p}^2$$

yields

$$\left[\frac{\partial}{\partial z}\left(\frac{\partial}{\partial t} + i\omega_\mathrm{c}\right)\frac{\partial}{\partial z} - \frac{\omega_\mathrm{p}^2}{c^2}\frac{\partial}{\partial t}\right]\widetilde{\mathbf{B}}_\mathrm{w} = -i\mu_0\left(\frac{\partial}{\partial t} + i\omega_\mathrm{c}\right)\frac{\partial}{\partial z}\widetilde{\mathbf{J}}_\mathrm{r} \qquad \text{(C.4)}$$

The wave magnetic field and the resonant current are then exressed as

$$\widetilde{\mathbf{B}}_\mathrm{w} = B_\mathrm{w}e^{i\left(\phi - \omega t + \int_{z_0}^{z} k\,dz'\right)}$$

$$\widetilde{\mathbf{J}}_\mathrm{r} = J_\mathrm{r}e^{i\left(\psi + \phi - \omega t + \int_{z_0}^{z} k\,dz'\right)}$$

where ϕ is the phase variation of the wave field with the rapid phase variation due to the temporal varation of the wave ω and the spatial variation of the wave k removed. That is, ϕ represents the additional phase accumulated due to the instability. The term ψ represents the phase of the current relative to the wave phase, or the angle between the current vector and the wave magnetic field vector. Expanding the right side and left sides of Equation (C.4) yields

$$-i\mu_0\left(\frac{\partial}{\partial t} + i\omega_\mathrm{c}\right)\frac{\partial}{\partial z}\widetilde{\mathbf{J}}_\mathrm{r} = -i\mu_0 J_\mathrm{r}e^{i\left(\psi + \phi - \omega t + \int_{z_0}^{z} k\,dz'\right)}.$$

$$\left\{-\left(\omega_\mathrm{c} - \omega + \frac{\partial\phi}{\partial t} + \frac{\partial\psi}{\partial t}\right)\left(k + \frac{\partial\phi}{\partial z} + \frac{\partial\psi}{\partial z}\right) + \frac{1}{J_\mathrm{r}}\frac{\partial^2 J_\mathrm{r}}{\partial t\partial z}\right.$$

$$+ i\left[\left(\frac{\partial^2\phi}{\partial t\partial z} + \frac{\partial^2\psi}{\partial t\partial z}\right) + \frac{1}{J_\mathrm{r}}\frac{\partial J_\mathrm{r}}{\partial t}\left(k + \frac{\partial\phi}{\partial z} + \frac{\partial\psi}{\partial z}\right)\right.$$

$$\left.\left. + \frac{1}{J_\mathrm{r}}\frac{\partial J_\mathrm{r}}{\partial z}\left(\omega_\mathrm{c} - \omega + \frac{\partial\phi}{\partial t} + \frac{\partial\psi}{\partial t}\right)\right]\right\} \qquad \text{(C.5)}$$

and

$$
\left[\frac{\partial}{\partial z} \left(\frac{\partial}{\partial t} + i\omega_{\mathrm{c}} \right) \frac{\partial}{\partial z} - \frac{\omega_{\mathrm{p}}^2}{c^2} \frac{\partial}{\partial t} \right] \widetilde{\mathbf{B}}_{\mathrm{w}} = e^{i\left(\phi - \omega t + \int_{z_0}^{z} k dz' \right)}
$$
$$
\left(- \frac{\partial B_{\mathrm{w}}}{\partial t} \left[\left(k + \frac{\partial \phi}{\partial z} \right)^2 + \frac{\omega_{\mathrm{p}}^2}{c^2} \right] - 2 \frac{\partial B_{\mathrm{w}}}{\partial z} \left(\omega_{\mathrm{c}} - \omega + \frac{\partial \phi}{\partial t} \right) \left(k + \frac{\partial \phi}{\partial z} \right) \right.
$$
$$
+ B_{\mathrm{w}} \left[-2 \left(k + \frac{\partial \phi}{\partial z} \right) \frac{\partial^2 \phi}{\partial t \partial z} - \left(\omega_{\mathrm{c}} - \omega + \frac{\partial \phi}{\partial t} \right) \frac{\partial^2 \phi}{\partial z^2} \right]
$$
$$
- 2 \frac{\partial B_{\mathrm{w}}}{\partial t} \frac{\partial^2 \phi}{\partial z^2} - 2 \frac{\partial B_{\mathrm{w}}}{\partial z} \frac{\partial^2 \phi}{\partial t \partial z} + \frac{\partial^3 B_{\mathrm{w}}}{\partial t \partial z^2}
$$
$$
+ i \left\{ B_{\mathrm{w}} \left[- \left(k + \frac{\partial \phi}{\partial z} \right)^2 \left(\omega_{\mathrm{c}} - \omega + \frac{\partial \phi}{\partial t} \right) + \frac{\omega_{\mathrm{p}}^2}{c^2} \left(-\omega + \frac{\partial \phi}{\partial t} \right) - \frac{\partial^3 \phi}{\partial t \partial z^2} \right] \right.
$$
$$
\left. \left. + 2 \left(k + \frac{\partial \phi}{\partial z} \right) \frac{\partial^2 B_{\mathrm{w}}}{\partial t \partial z} + \left(\omega_{\mathrm{c}} - \omega + \frac{\partial \phi}{\partial t} \right) \frac{\partial^2 B_{\mathrm{w}}}{\partial z^2} \right\} \right)
$$

(C.6)

respectively. Combining Equation (C.5) and Equation (C.6), cancelling the phase

varation term and expanding the $J_\mathrm{r} e^{i\psi}$ term,

$$
\begin{aligned}
-\frac{\partial B_\mathrm{w}}{\partial t} &\left[\left(k + \frac{\partial \phi}{\partial z} \right)^2 + \frac{\omega_\mathrm{p}^2}{c^2} \right] - 2 \frac{\partial B_\mathrm{w}}{\partial z} \left(\omega_\mathrm{c} - \omega + \frac{\partial \phi}{\partial t} \right) \left(k + \frac{\partial \phi}{\partial z} \right) \\
+ B_\mathrm{w} &\left[-2 \left(k + \frac{\partial \phi}{\partial z} \right) \frac{\partial^2 \phi}{\partial t \partial z} - \left(\omega_\mathrm{c} - \omega + \frac{\partial \phi}{\partial t} \right) \frac{\partial^2 \phi}{\partial z^2} \right] \\
&- 2 \frac{\partial B_\mathrm{w}}{\partial t} \frac{\partial^2 \phi}{\partial z^2} - 2 \frac{\partial B_\mathrm{w}}{\partial z} \frac{\partial^2 \phi}{\partial t \partial z} + \frac{\partial^3 B_\mathrm{w}}{\partial t \partial z^2} \\
+ i \Bigg\{ B_\mathrm{w} &\left[- \left(k + \frac{\partial \phi}{\partial z} \right)^2 \left(\omega_\mathrm{c} - \omega + \frac{\partial \phi}{\partial t} \right) - \frac{\omega_\mathrm{p}^2}{c^2} \left(-\omega + \frac{\partial \phi}{\partial t} \right) - \frac{\partial^3 \phi}{\partial t \partial z^2} \right] \\
&+ 2 \left(k + \frac{\partial \phi}{\partial z} \right) \frac{\partial^2 B_\mathrm{w}}{\partial t \partial z} + \left(\omega_\mathrm{c} - \omega + \frac{\partial \phi}{\partial t} \right) \frac{\partial^2 B_\mathrm{w}}{\partial z^2} \Bigg\} \\
= \mu_0 J_\mathrm{r} &\left(\sin \psi - i \cos \psi \right) \cdot \\
&\Bigg\{ - \left(\omega_\mathrm{c} - \omega + \frac{\partial \phi}{\partial t} + \frac{\partial \psi}{\partial t} \right) \left(k + \frac{\partial \phi}{\partial z} + \frac{\partial \psi}{\partial z} \right) + \frac{1}{J_\mathrm{r}} \frac{\partial^2 J_\mathrm{r}}{\partial t \partial z} \\
&+ i \left[\left(\frac{\partial^2 \phi}{\partial t \partial z} + \frac{\partial^2 \psi}{\partial t \partial z} \right) + \frac{1}{J_\mathrm{r}} \frac{\partial J_\mathrm{r}}{\partial t} \left(k + \frac{\partial \phi}{\partial z} + \frac{\partial \psi}{\partial z} \right) \right. \\
&\left. + \frac{1}{J_\mathrm{r}} \frac{\partial J_\mathrm{r}}{\partial z} \left(\omega_\mathrm{c} - \omega + \frac{\partial \phi}{\partial t} + \frac{\partial \psi}{\partial t} \right) \right] \Bigg\} \\
= \mu_0 J_\mathrm{r} &\left(\sin \psi \left[- \left(\omega_\mathrm{c} - \omega + \frac{\partial \phi}{\partial t} + \frac{\partial \psi}{\partial t} \right) \left(k + \frac{\partial \phi}{\partial z} + \frac{\partial \psi}{\partial z} \right) + \frac{1}{J_\mathrm{r}} \frac{\partial^2 J_\mathrm{r}}{\partial t \partial z} \right] \right. \\
&+ \cos \psi \left[\left(\frac{\partial^2 \phi}{\partial t \partial z} + \frac{\partial^2 \psi}{\partial t \partial z} \right) + \frac{1}{J_\mathrm{r}} \frac{\partial J_\mathrm{r}}{\partial t} \left(k + \frac{\partial \phi}{\partial z} + \frac{\partial \psi}{\partial z} \right) \right. \\
&\left. + \frac{1}{J_\mathrm{r}} \frac{\partial J_\mathrm{r}}{\partial z} \left(\omega_\mathrm{c} - \omega + \frac{\partial \phi}{\partial t} + \frac{\partial \psi}{\partial t} \right) \right] \\
&+ i \Bigg\{ \cos \psi \left[\left(\omega_\mathrm{c} - \omega + \frac{\partial \phi}{\partial t} + \frac{\partial \psi}{\partial t} \right) \left(k + \frac{\partial \phi}{\partial z} + \frac{\partial \psi}{\partial z} \right) - \frac{1}{J_\mathrm{r}} \frac{\partial^2 J_\mathrm{r}}{\partial t \partial z} \right] \\
&\sin \psi \left[\left(\frac{\partial^2 \phi}{\partial t \partial z} + \frac{\partial^2 \psi}{\partial t \partial z} \right) + \frac{1}{J_\mathrm{r}} \frac{\partial J_\mathrm{r}}{\partial t} \left(k + \frac{\partial \phi}{\partial z} + \frac{\partial \psi}{\partial z} \right) \right. \\
&\left. \left. + \frac{1}{J_\mathrm{r}} \frac{\partial J_\mathrm{r}}{\partial z} \left(\omega_\mathrm{c} - \omega + \frac{\partial \phi}{\partial t} + \frac{\partial \psi}{\partial t} \right) \right] \Bigg\} \right)
\end{aligned}
\tag{C.7}
$$

Taking the real part of Equation (C.7),

$$
\begin{aligned}
-\frac{\partial B_{\mathrm{w}}}{\partial t} & \left[\left(k + \frac{\partial \phi}{\partial z} \right)^2 + \frac{\omega_{\mathrm{p}}^2}{c^2} \right] - 2 \frac{\partial B_{\mathrm{w}}}{\partial z} \left(\omega_{\mathrm{c}} - \omega + \frac{\partial \phi}{\partial t} \right) \left(k + \frac{\partial \phi}{\partial z} \right) \\
+ B_{\mathrm{w}} & \left[-2 \left(k + \frac{\partial \phi}{\partial z} \right) \frac{\partial^2 \phi}{\partial t \partial z} - \left(\omega_{\mathrm{c}} - \omega + \frac{\partial \phi}{\partial t} \right) \frac{\partial^2 \phi}{\partial z^2} \right] \\
& - 2 \frac{\partial B_{\mathrm{w}}}{\partial t} \frac{\partial^2 \phi}{\partial z^2} - 2 \frac{\partial B_{\mathrm{w}}}{\partial z} \frac{\partial^2 \phi}{\partial t \partial z} + \frac{\partial^3 B_{\mathrm{w}}}{\partial t \partial z^2} \\
= \mu_0 J_{\mathrm{r}} & \left\{ \sin \psi \left[- \left(\omega_{\mathrm{c}} - \omega + \frac{\partial \phi}{\partial t} + \frac{\partial \psi}{\partial t} \right) \left(k + \frac{\partial \phi}{\partial z} + \frac{\partial \psi}{\partial z} \right) + \frac{1}{J_{\mathrm{r}}} \frac{\partial^2 J_{\mathrm{r}}}{\partial t \partial z} \right] \right. \\
& + \cos \psi \left[\left(\frac{\partial^2 \phi}{\partial t \partial z} + \frac{\partial^2 \psi}{\partial t \partial z} \right) + \frac{1}{J_{\mathrm{r}}} \frac{\partial J_{\mathrm{r}}}{\partial t} \left(k + \frac{\partial \phi}{\partial z} + \frac{\partial \psi}{\partial z} \right) \right. \\
& \left. \left. + \frac{1}{J_{\mathrm{r}}} \frac{\partial J_{\mathrm{r}}}{\partial z} \left(\omega_{\mathrm{c}} - \omega + \frac{\partial \phi}{\partial t} + \frac{\partial \psi}{\partial t} \right) \right] \right\}
\end{aligned}
\tag{C.8}
$$

At this point in the derivation, *Nunn* [1974] assumes that the wave is monochromatic. More explicitly, *Omura and Matsumoto* [1982] assume that the first order derivatives of J_{r} and ψ are negligible as well as all second and third order derivatives and the first order derivitives of ϕ on the right side of Equation (C.8), which gives

$$
\begin{aligned}
-\frac{\partial B_{\mathrm{w}}}{\partial t} \left(k^2 + \frac{\omega_{\mathrm{p}}^2}{c^2} \right) - 2k \left(\omega_{\mathrm{c}} - \omega \right) \frac{\partial B_{\mathrm{w}}}{\partial z} &= -\frac{\mu_0}{2} J_{\mathrm{r}} \sin \psi \left[2k \left(\omega_{\mathrm{c}} - \omega \right) \right] \\
\frac{\partial B_{\mathrm{w}}}{\partial t} + \frac{2k \left(\omega_{\mathrm{c}} - \omega \right)}{k^2 + \frac{\omega_{\mathrm{p}}^2}{c^2}} \frac{\partial B_{\mathrm{w}}}{\partial z} &= -\frac{\mu_0}{2} J_{\mathrm{E}} \frac{2k \left(\omega_{\mathrm{c}} - \omega \right)}{k^2 + \frac{\omega_{\mathrm{p}}^2}{c^2}} \\
\frac{\partial B_{\mathrm{w}}}{\partial t} + v_{\mathrm{g}} \frac{\partial B_{\mathrm{w}}}{\partial z} &= -\frac{\mu_0}{2} J_{\mathrm{E}} v_{\mathrm{g}}
\end{aligned}
\tag{C.9}
$$

where $J_{\mathrm{E}} = -J_{\mathrm{r}} \sin \psi$ is the component of the resonant current in the direction of the wave electric field, and v_{g} is the wave group velocity (See Equation (3.3)). Returning

to Equation (C.7) and taking the imaginary part,

$$
\begin{aligned}
B_{\mathrm{w}} & \left[-\left(k + \frac{\partial \phi}{\partial z}\right)^2 \left(\omega_{\mathrm{c}} - \omega + \frac{\partial \phi}{\partial t}\right) - \frac{\omega_{\mathrm{p}}^2}{c^2}\left(-\omega + \frac{\partial \phi}{\partial t}\right) - \frac{\partial^3 \phi}{\partial t \partial z^2} \right] \\
& + 2\left(k + \frac{\partial \phi}{\partial z}\right)\frac{\partial^2 B_{\mathrm{w}}}{\partial t \partial z} + \left(\omega_{\mathrm{c}} - \omega + \frac{\partial \phi}{\partial t}\right)\frac{\partial^2 B_{\mathrm{w}}}{\partial z^2} \\
= \mu_0 J_{\mathrm{r}} & \left\{ \cos\psi \left[\left(\omega_{\mathrm{c}} - \omega + \frac{\partial \phi}{\partial t} + \frac{\partial \psi}{\partial t}\right)\left(k + \frac{\partial \phi}{\partial z} + \frac{\partial \psi}{\partial z}\right) - \frac{1}{J_{\mathrm{r}}}\frac{\partial^2 J_{\mathrm{r}}}{\partial t \partial z} \right] \right. \\
& + \sin\psi \left[\left(\frac{\partial^2 \phi}{\partial t \partial z} + \frac{\partial^2 \psi}{\partial t \partial z}\right) + \frac{1}{J_{\mathrm{r}}}\frac{\partial J_{\mathrm{r}}}{\partial t}\left(k + \frac{\partial \phi}{\partial z} + \frac{\partial \psi}{\partial z}\right) \right. \\
& \left. \left. + \frac{1}{J_{\mathrm{r}}}\frac{\partial J_{\mathrm{r}}}{\partial z}\left(\omega_{\mathrm{c}} - \omega + \frac{\partial \phi}{\partial t} + \frac{\partial \psi}{\partial t}\right) \right] \right\}
\end{aligned}
\tag{C.10}
$$

expanding the first term,

$$
\begin{aligned}
B_{\mathrm{w}} & \left[\left(k^2 + \frac{\omega_{\mathrm{p}}^2}{c^2}\right)\omega - \left(k^2 + \frac{\omega_{\mathrm{p}}^2}{c^2}\right)\frac{\partial \phi}{\partial t} - k^2 \omega_{\mathrm{c}} - 2k\left(\omega_{\mathrm{c}} - \omega\right)\frac{\partial \phi}{\partial z} \right. \\
& \left. - 2k\frac{\partial \phi}{\partial z}\frac{\partial \phi}{\partial t} - \left(\frac{\partial \phi}{\partial z}\right)^2\left(\omega_{\mathrm{c}} - \omega + \frac{\partial \phi}{\partial t}\right) - \frac{\partial^3 \phi}{\partial t \partial z^2} \right] \\
& + 2\left(k + \frac{\partial \phi}{\partial z}\right)\frac{\partial^2 B_{\mathrm{w}}}{\partial t \partial z} + \left(\omega_{\mathrm{c}} - \omega + \frac{\partial \phi}{\partial t}\right)\frac{\partial^2 B_{\mathrm{w}}}{\partial z^2} \\
= \mu_0 J_{\mathrm{r}} & \left\{ \cos\psi \left[\left(\omega_{\mathrm{c}} - \omega + \frac{\partial \phi}{\partial t} + \frac{\partial \psi}{\partial t}\right)\left(k + \frac{\partial \phi}{\partial z} + \frac{\partial \psi}{\partial z}\right) - \frac{1}{J_{\mathrm{r}}}\frac{\partial^2 J_{\mathrm{r}}}{\partial t \partial z} \right] \right. \\
& + \sin\psi \left[\left(\frac{\partial^2 \phi}{\partial t \partial z} + \frac{\partial^2 \psi}{\partial t \partial z}\right) + \frac{1}{J_{\mathrm{r}}}\frac{\partial J_{\mathrm{r}}}{\partial t}\left(k + \frac{\partial \phi}{\partial z} + \frac{\partial \psi}{\partial z}\right) \right. \\
& \left. \left. + \frac{1}{J_{\mathrm{r}}}\frac{\partial J_{\mathrm{r}}}{\partial z}\left(\omega_{\mathrm{c}} - \omega + \frac{\partial \phi}{\partial t} + \frac{\partial \psi}{\partial t}\right) \right] \right\}
\end{aligned}
\tag{C.11}
$$

and applying the "monochromatic" assumption

$$
\omega - \frac{\partial \phi}{\partial t} - k\frac{k\omega_{\mathrm{c}}}{k^2 + \frac{\omega_{\mathrm{p}}^2}{c^2}} - \frac{2k\left(\omega_{\mathrm{c}} - \omega\right)}{k^2 + \frac{\omega_{\mathrm{p}}^2}{c^2}}\frac{\partial \phi}{\partial z} = \frac{\mu_0}{2}\frac{J_{\mathrm{r}}\cos\psi}{B_{\mathrm{w}}}\frac{2k\left(\omega_{\mathrm{c}} - \omega\right)}{k^2 + \frac{\omega_{\mathrm{p}}^2}{c^2}}
$$

$$
\omega - kv_{\mathrm{p}} - \frac{\partial \phi}{\partial t} - v_{\mathrm{g}}\frac{\partial \phi}{\partial z} = \frac{\mu_0}{2}\frac{J_{\mathrm{r}}\cos\psi}{B_{\mathrm{w}}}v_{\mathrm{g}}
$$

$$
\frac{\partial \phi}{\partial t} + v_{\mathrm{g}}\frac{\partial \phi}{\partial z} = -\frac{\mu_0}{2}\frac{J_{\mathrm{B}}}{B_{\mathrm{w}}}v_{\mathrm{g}}
\tag{C.12}
$$

where $J_{\mathrm{B}} = J_{\mathrm{r}} \cos \psi$ is the component of the resonant current in the direction of the wave magnetic field, and $v_{\mathrm{p}} = \omega/k$ is the wave phase velocity (See Equation (3.2)). Together, Equation (C.9) and Equation (C.12) describe the influence of a resonant current on the evolution of a parallel propagating, monochromatic, whistler mode wave.

Appendix D

Liouville's Theorem

Liouville's theorem states simply that the phase space density along a dynamical path is constant. That is,

$$F(t + \Delta t, \mathbf{r} + \Delta \mathbf{r}, \mathbf{v} + \Delta \mathbf{v}) = F(t, \mathbf{r}, \mathbf{v}) \tag{D.1}$$

where the value of the distribution function at time $t + \Delta t$ is directly related to the value of the distribution function at time t, and $\Delta \mathbf{r}$ and $\Delta \mathbf{v}$ represent the change in position and velocity in time Δt, repectively for a particle that starts at position $\mathbf{r}$ and $\mathbf{v}$ at time t. Liouville's theorem can be easily demonstrated in one dimension. The discussion here follows that of *Walt* [1994].

Consider a box in (r, v) space with dimensions Δr and Δv. The amount of particles entering an leaving that box in unit time can be expressed as

$$\frac{\partial F}{\partial t} \Delta r \Delta v = F(r, v') \Delta r \frac{\partial v'}{\partial t} \bigg|_{v'=v} - F(r, v') \Delta r \frac{\partial v'}{\partial t} \bigg|_{v'=v+\Delta v} +$$
$$F(r', v) \Delta v \frac{\partial r'}{\partial t} \bigg|_{r'=r} - F(r', v) \Delta v \frac{\partial r'}{\partial t} \bigg|_{r'=r+\Delta r} \tag{D.2}$$

where the terms on the left side of Equation (D.2) represent the particles entering

and leaving from each side of the box. Dividing through by $\Delta r \Delta v$,

$$
\frac{\partial F}{\partial t} = \frac{1}{\Delta v} \left[F(r, v') \frac{\partial v'}{\partial t} \bigg|_{v'=v} - F(r, v') \frac{\partial v'}{\partial t} \bigg|_{v'=v+\Delta v} \right] +
$$
$$
\frac{1}{\Delta r} \left[F(r', v) \frac{\partial r'}{\partial t} \bigg|_{r'=r} - F(r', v) \frac{\partial r'}{\partial t} \bigg|_{r'=r+\Delta r} \right] \quad \text{(D.3)}
$$

taking Equation (D.3) in the limit $\Delta r, \Delta q \to 0$,

$$
\frac{\partial F}{\partial t} = -\frac{\partial}{\partial v} \left(F \frac{\partial v}{\partial t} \right) - \frac{\partial}{\partial r} \left(F \frac{\partial r}{\partial t} \right) \quad \text{(D.4)}
$$

and with Hamiltons equations,

$$
\frac{\partial v}{\partial t} = \frac{\partial H}{\partial r} \qquad \frac{\partial r}{\partial t} = -\frac{\partial H}{\partial q}
$$

Equation (D.4) becomes

$$
\frac{\partial F}{\partial t} = -\frac{\partial v}{\partial t}\frac{\partial F}{\partial v} - F\frac{\partial^2 H}{\partial r \partial v} - \frac{\partial r}{\partial t}\frac{\partial F}{\partial r} + F\frac{\partial^2 H}{\partial r \partial v}
$$
$$
\frac{\partial F}{\partial t} = -\frac{\partial v}{\partial t}\frac{\partial F}{\partial v} - \frac{\partial r}{\partial t}\frac{\partial F}{\partial r} \quad \text{(D.5)}
$$

The absolute derivative of F in terms of r and v is

$$
\frac{dF}{dt} = \frac{\partial F}{\partial t} + \frac{\partial v}{\partial t}\frac{\partial F}{\partial v} + \frac{\partial r}{\partial t}\frac{\partial F}{\partial r} = 0 \quad \text{(D.6)}
$$

where the result from Equation (D.5) has been used. Thus, the value of the phase space density remains constant as the dynamical path of the particles within the box is followed, which is a restatement of Liouville's theorem, which holds as long as the individual particle motions in the system can be described by a Hamiltonian.

Bibliography

Angerami, J. J., Whistler duct properties deduced from VLF observations made with the ogo 3 satellite near the magnetic equator, *J. Geophys. Res.*, *75*, 6115, 1970.

Anonymous, Proposal evaluation forms for NSF proposal DPP-8022282, 1980.

Anonymous, Proposal evaluation forms for NSF proposal AEO-8318508, 1983a.

Anonymous, Proposal evaluation forms for NSF proposal DPP-8317092, 1983b.

Anonymous, Proposal evaluation forms for NSF proposal DPP-8613783, 1987.

Anonymous, Proposal evaluation forms for NSF proposal DPP 89-18326, 1989.

Ashour-Abdalla, M., Amplification of whistler waves in the magnetosphere, *Planet. Sp. Sci.*, *20*, 639, 1972.

Bell, T. F., High amplitude VLF transmitter signals and associated sidebands observed near the magnetic equatorial plane on the ISEE-1 satellite, *J. Geophys. Res.*, *90*, 2792, 1985.

Bell, T. F., and O. Buneman, Plasma instability in the whistler mode caused by a gyrating electron stream, *The Physical Review*, *133*, A1300, 1964.

Bell, T. F., and R. A. Helliwell, Pulsation phenomena observed in long duration VLF whistler-mode signals, *J. Geophys. Res.*, *76*, 8414, 1971.

Bell, T. F., and U. S. Inan, Satellite observation of nonducted signals from the Siple transmitter, *Antarctic J.*, *X*, 211, 1975.

Bell, T. F., and U. S. Inan, Transient nonlinear pitch angle scattering of energetic electrons by coherent VLF wave packets in the magnetosphere, *J. Geophys. Res.*, *86*, 9047, 1981.

Bell, T. F., and H. D. Ngo, Electrostatic lower hybrid waves excited by electromagnetic whistler mode waves scattering from planar magnetic-field-aligned plasma density irregularities, *J. Geophys. Res.*, *95*, 149, 1990.

Bell, T. F., U. S. Inan, I. Kimura, H. Matsumoto, T. Mukai, and K. Hashimoto, EXOS-B/Siple Station VLF wave-particle interaction experiments: 2. transmitter signals and associated emissions, *J. Geophys. Res.*, *88*, 295, 1983a.

Bell, T. F., H. G. James, U. S. Inan, and J. P. Katsufrakis, The apparent spectral broadening of VLF transmitter signals during transionospheric propagation, *J. Geophys. Res.*, *88*, 4813, 1983b.

Bell, T. F., U. S. Inan, R. A. Helliwell, and J. D. Scudder, Simultaneous triggered VLF emissions and energetic electron distributions observed on POLAR with PWI and HYDRA, *Geophys. Res. Lett.*, *27*, 165, 2000.

Bezrukikh, V. V., G. A. Kotova, L. A. Lezhen, J. Lemaire, V. Pierrard, and Y. I. Venediktov, Dynamics of temperature and density of cold protons of the earth's plasmasphere measured by the auroral probe/alpha-3 experiment data during geomagnetic disturbances, *Cosmic Research*, *41*, 392, 2003.

Bittencourt, J. A., *Fundamentals of Plasma Physics*, Springer-Verlag, New York, 2004.

Brice, N., An explanation of triggered very-low-frequency emissions, *J. Geophys. Res.*, *68*, 4626, 1963.

Brice, N., Fundamentals of VLF emission generation mechanisms, *J. Geophys. Res.*, *69*, 4515, 1964.

Brinca, A. L., Whistler side-band growth due to nonlinear wave-particle interaction, *J. Geophys. Res.*, *77*, 3508, 1972.

Brinca, A. L., Whistler modulational instability, *J. Geophys. Res.*, *78*, 181, 1973.

Burke, W. J., A. G. Rubin, D. A. Hardy, and E. G. Holeman, Banded electron structures in the plasmasphere, *J. Geophys. Res.*, *100*, 7759, 1995.

Carlson, C. R., R. A. Helliwell, and D. L. Carpenter, Variable frequency VLF signals in the magnetosphere: associated phenomena and plasma diagnostics, *J. Geophys. Res.*, *90*, 1507, 1985.

Carlson, C. R., R. A. Helliwell, and U. S. Inan, Space-time evolution of whistler mode wave growth in the magnetosphere, *J. Geophys. Res.*, *95*, 15,073, 1990.

Carpenter, D. L., Whistler evidence of a 'knee' in the magnetospheric ionization density profile, *J. Geophys. Res.*, *68*, 1675, 1963.

Carpenter, D. L., Whistler studies of the plasmapause in the magnetosphere, 1: Temporal variations in the position of the knee and some evidence of plasma motions near the knee, *J. Geophys. Res.*, *71*, 693, 1966.

Carpenter, D. L., Ducted whistler-mode propagation in the magnetosphere; a half-gyrofrequency upper intensity cutoff and some associated wave growth phenomena, *J. Geophys. Res.*, *73*, 2919, 1968.

Carpenter, D. L., Fast fluctuations in the arrival bearing of magnetospherically propagating signals from the Siple, Antarctica VLF transmitter, *J. Geophys. Res.*, *85*, 4157, 1980.

Carpenter, D. L., Active and passive VLF experiments at Siple Station, 1980-1981, *Antarctic J.*, *XVI*, 211, 1981.

Carpenter, D. L., Letter to Dr. Benson Fogle, NSF, to extend support for grant DPP 80-22540, Personal Letter, 1982a.

Carpenter, D. L., Letter to Dr. Benson Fogle, NSF, to extend support for grant DPP 80-22282, Personal Letter, 1982b.

Carpenter, D. L., Siple operations plan 1983, 1983.

Carpenter, D. L., Material on operations guidelines, 1986a.

Carpenter, D. L., Email to J. Logan 12 December, 1986b.

Carpenter, D. L., Notes on transmitter operations at siple, Personal Notes, 1987.

Carpenter, D. L., and R. R. Anderson, An ISEE/whistler model of equatorial electron density in the magnetosphere, *J. Geophys. Res.*, *97*, 1097, 1992.

Carpenter, D. L., and Z. T. Bao, Occurence properties of ducted whistler-mode signals from the new VLF transmitter at Siple Station, Antarctica, *J. Geophys. Res.*, *88*, 7051, 1983.

Carpenter, D. L., and T. R. Miller, Ducted magnetospheric propagation of signals from the Siple, Antarctica VLF transmitter, *J. Geophys. Res.*, *81*, 2692, 1976.

Carpenter, D. L., and T. R. Miller, Rare ground-based observations of Siple VLF transmitter signals outside the plasmapause, *J. Geophys. Res.*, *88*, 10,227, 1983.

Carpenter, D. L., V. S. Sonwalkar, R. A. Helliwell, M. Walt, U. S. Inan, M. Ikeda, and D. L. Claude, Probing properties of the magnetospheric hot plasma distribution by whistler mode wave injection at multiple frequencies: Evidence of spatial as well as temporal growth, *J. Geophys. Res.*, *102*, 14,355, 1997.

Chang, D. C. D., and R. A. Helliwell, Emission triggering in the magnetosphere by controlled interruption of coherent VLF signals, *J. Geophys. Res.*, *84*, 7170, 1979.

Chang, D. C. D., R. A. Helliwell, and T. F. Bell, Side-band mutual interactions in the magnetosphere, *J. Geophys. Res.*, *85*, 1703, 1980.

Chiang, E., Visit to Siple Station, Memorandum, 1987.

Cornilleau-Wehrlin, N., and R. Gendrin, VLF transmitter-induced quiet bands: a qualitative interpretation, *J. Geophys. Res.*, *84*, 882, 1979.

Courant, R., K. O. Friedrichs, and H. Lewy, On the partial differential equations of mathematical physics, *IBM J. Res. Dev.*, *11*, 215, 1967.

Das, A. C., A mechanism for VLF emissions, *J. Geophys. Res.*, *73*, 7457, 1968.

Dowden, R. L., A. D. McKay, L. E. S. Amon, H. C. Koons, and M. H. Dazey, Linear and nonlinear amplification in the magnetosphere during a 6.6-khz transmission, *J. Geophys. Res.*, *83*, 169, 1978.

Dysthe, K. B., Some studies of triggered whistler emissions, *J. Geophys. Res.*, *76*, 6915, 1971.

Golkowski, M. A., U. S. Inan, A. R. Gibby, and M. Cohen, Magnetospheric amplification and emission triggering by ELF/VLF waves injected by the 3.6 MW HAARP ionospheric heater, *J. Geophys. Res.*, *113*, A013,157, 2008.

Gurnett, D. A., and A. Bhattacharjee, *Introduction to Plasma Physics*, Cambridge University Press, 2005.

Helliwell, R. A., *Whistlers and related ionospheric phenomena*, Stanford University Press, 1965.

Helliwell, R. A., A theory of discrete vlf emissions from the magnetosphere, *J. Geophys. Res.*, *72*, 4773, 1967.

Helliwell, R. A., A new program of magnetospheric research in Antarctica, NSF Proposal, stanford Proposal No. RL 10-71, 1970.

Helliwell, R. A., Active very low frequency experiments on the magnetosphere from Siple Station, Antarctica, *Phil. Trans. R. Soc. Lond.*, *279*, 213, 1977.

Helliwell, R. A., Effects of power line radiation into the magnetosphere, in *Wave instabilities in space plasma*, edited by P. J. Palmadesso and K. Papadopolos, p. 27, D. Reidel Publishing Co., 1979a.

Helliwell, R. A., Siple Station experiments on wave-particle interactions in the magnetosphere, in *Wave instabilities in space plasmas*, edited by P. J. Palmadesso and K. Papadopolos, p. 191, D. Reidel Publishing Co., 1979b.

Helliwell, R. A., Very low frequency wave investigations of the magnetosphere at roberval, quebec, NSF Proposal, NSF Proposal No. DPP 80-22540, 1981a.

Helliwell, R. A., Letter to Dr. Benson Fogle, NSF, to extend support for grant DPP 80-22540, Personal Letter, 1981b.

Helliwell, R. A., Active and passive very low frequency wave particle experiments on the magnetosphere from Siple Station, Antarctica, NSF Proposal, NSF Proposal No. DPP 80-22282, 1981c.

Helliwell, R. A., Letter to Dr. Benson Fogle, NSF, to provide background on a proposal to modify the VLF antenna at Siple Station, Personal Letter, 1981d.

Helliwell, R. A., Controlled stimulation of VLF emissions from Siple Station, Antarctica, *Radio Sci.*, *18*, 801, 1983a.

Helliwell, R. A., Wave-particle experimentson the magnetosphere and ionosphere from Siple Station, Antarctica, NSF Proposal, NSF Proposal No. DPP-8317092, 1983b.

Helliwell, R. A., VLF wave injections from the ground, in *Active Experiments in Space Symposium*, 1983c.

Helliwell, R. A., Letter to Dr. Benson Fogle, NSF, providing revised budgets and summaries of research activities, Personal Letter, 1984.

Helliwell, R. A., Letter to Dr. John Lynch, NSF, to request continuing support under NSF grant DPP 83-17092, Personal Letter, 1985.

Helliwell, R. A., VLF wave-particle interaction experiments on the magnetosphere and ionosphere from Siple Station, Antarctica and lake mistissini, quebec, NSF Proposal, NSF Proposal No. DPP-8613783, 1986.

Helliwell, R. A., Final project report for NSF award DPP 83-18508, VLF wave-particle experiments in the magnetosphere and ionosphere from roberval, quebec, Final Project Report, 1987a.

Helliwell, R. A., Letter to J. T. Lynch, NSF, concerning upcoming Siple operations, Personal Letter, 1987b.

Helliwell, R. A., Letter to J. T. Lynch, NSF, requesting a third year of support for grant DPP 86013783, 1987c.

Helliwell, R. A., VLF wave stimulation experiments in the magnetosphere from Siple Station, Antarctica, *Rev. Geophys.*, *26*, 551, 1988a.

Helliwell, R. A., Letter to J. T. Lynch, NSF, regarding 88/89 summer operations, Personal Letter, 1988b.

Helliwell, R. A., Letter to I. Kimura, Kyoto University, regarding Siple/EXOS-D collaboration, Personal Letter, 1988c.

Helliwell, R. A., Letter to J. T. Lynch requesting a third year of support for grant DPP 86-13783, Personal Letter, 1989a.

Helliwell, R. A., Investigations of mechanisms and effects of wave-particle interactions using data from Siple Station, Antarctica VLF wave-injection experiments, NSF Proposal, NSF Proposal No. DPP 89-18326, 1989b.

Helliwell, R. A., Triggering of whistler mode emissions by the band-limited impulse associated with amplified VLF signals from Siple Station, Antarctica, *Geophys. Res. Lett.*, *27*, 1455, 2000.

Helliwell, R. A., and T. L. Crystal, A feedback model of cyclotron interaction between whistler-mode waves and energetic electrons in the magnetosphere, *J. Geophys. Res.*, *78*, 7357, 1973.

Helliwell, R. A., and J. P. Katsufrakis, VLF wave injection into the magnetosphere from Siple Station, Antarctica, *J. Geophys. Res.*, *79*, 2511, 1974.

Helliwell, R. A., and J. P. Katsufrakis, Controlled wave-particle interaction experiments, in *Upper atmospere research in Antarctica*, edited by L. J. Lanzerotti and C. G. Park, p. 100, Antarctic Research Service, 1978.

Helliwell, R. A., J. P. Katsufrakis, M. Trimpi, and N. Brice, Artificially stimulated very-low-frequency radiation from the ionosphere, *J. Geophys. Res.*, *69*, 2391, 1964.

Helliwell, R. A., J. P. Katsufrakis, T. F. Bell, and R. Raghuram, VLF line radiation in the Earth's magnetosphere and its association with power system radiation, *J. Geophys. Res.*, *80*, 4249, 1975.

Helliwell, R. A., D. L. Carpenter, and T. R. Miller, Power theshold for growth of coherent VLF signals in the magnetosphere, *J. Geophys. Res.*, *85*, 3360, 1980.

Helliwell, R. A., U. S. Inan, J. P. Katsufrakis, and D. L. Carpenter, Beat excitation of whistler mode sidebands using the Siple VLF transmitter, *J. Geophys. Res.*, *91*, 143, 1986.

Helms, W. J., H. M. Swarm, and D. K. Reynolds, An experimental study of the polar lower ionosphere using very low frequency waves, *Tech. Rep. 123*, University of Washington, Seattle, Washington, 1968.

Holmes and Narver, Inc., *B.4.2-2AS(8178) Operations and Management Manual for Siple Station*, Holmes and Narver, Inc., 1978.

Inan, U. S., Non-linear gyroresonant interactions of energetic particles and coherent VLF waves in the magnetosphere, Ph.D. thesis, Stanford University, 1977.

Inan, U. S., Gyroresonant pitch angle scattering by coherent and incoherent whistler mode waves in the magnetosphere, *J. Geophys. Res.*, *92*, 127, 1987.

Inan, U. S., and T. F. Bell, Spectral broadening of VLF transimitter signals observed on DE-1: a quasi-electrostatic phenomenon?, *J. Geophys. Res.*, *90*, 1771, 1985.

Inan, U. S., and R. A. Helliwell, De-1 observation of VLF transmitter signals and wave-particle interactions in the magnetosphere, *Geophys. Res. Lett.*, *9*, 917, 1982.

Inan, U. S., T. F. Bell, D. L. Carpenter, and R. R. Anderson, Explorer 45 and Imp 6 observations in the magnetosphere of injected waves from the Siple Station VLF transmitter, *J. Geophys. Res.*, *82*, 1177, 1977.

Inan, U. S., T. F. Bell, and R. A. Helliwell, Nonlinear pitch angle scattering of energetic electrons by coherent VLF waves in the magnetosphere, *J. Geophys. Res.*, *83*, 3235, 1978.

Inan, U. S., M. Platino, T. F. Bell, D. A. Gurnett, and J. S. Pickett, Cluster measurements of rapidly moving sources of ELF/VLF chorus, *J. Geophys. Res.*, *109*, A05,214, 2004a.

Inan, U. S., et al., Multi-hop whistler-mode ELF/VLF signals and triggered emissions excited by the HAARP HF heater, *Geophys. Res. Lett.*, *31*, L24,805, 2004b.

Karpman, V. I., J. N. Isotomin, and S. D. R., Nonlinear theory of a quasi-monochromatic whistler mode packet in inhomogeneous plasma, *Plas. Phys.*, *16*, 685, 1974.

Katoh, Y., and Y. Omura, A study of generation mechanism of VLF triggered emission by self-consistent particle code, *J. Geophys. Res.*, *111*, A12,207, 2006.

Kennel, C. F., and H. E. Petschek, Limit on stably trapped particle fluxes, *J. Geophys. Res.*, *71*, 1, 1966.

Kimura, I., Triggering of VLF magnetospheric noise by a low-power ($\sim$100 watts) transmitter, *J. Geophys. Res.*, *73*, 445, 1968.

Kimura, I., H. Matsumoto, T. Mukai, K. Hashimoto, T. F. Bell, U. S. Inan, R. A. Helliwell, and J. P. Katsufrakis, EXOS-B/Siple Station VLF wave-particle interaction experiments: 1. general description and wave-particle correlations, *J. Geophys. Res.*, *88*, 282, 1983.

Kintner, P. M., R. Brittain, M. C. Kelley, D. L. Carpenter, and M. J. Rycroft, In situ measurements of transionospheric VLF wave injection, *J. Geophys. Res.*, *88*, 7065, 1983.

Koons, H. C., and M. H. Dazey, Transportable VLF transmitter, in *ELF–VLF Radio Wave Propagation*, edited by J. A. Holtet, p. 413, D. Reidel Publishing Co., 1974.

Lampe, M., G. Joyce, W. M. Manheimer, A. Streltsov, and G. Ganguli, Quasineutral particle simulation technique for whistlers, *J. Comp. Phys.*, *214*, 284, 2006.

Lanzerotti, L. J., Studies of geomagnetic pulsations, in *Upper Atmospheric Research in Antarctica*, edited by L. J. Lanzerotti and C. G. Park, p. 130, Antarctic Research Service, 1978.

Lasch, S., Unique features of VLF noise triggered in the magnetosphere by Morse code dots from NAA, *J. Geophys. Res.*, *74*, 1856, 1969.

Leavitt, M. K., D. L. Carpenter, M. T. Seely, R. R. Paddin, and J. H. Doolittle, Initial results from a tracking receiver direction finder for whistler mode signals, *J. Geophys. Res.*, *83*, 1601, 1978.

Logan, J., Email to J. P. Katsufrakis 23 June, Electronic Mail, 1986a.

Logan, J., Email to D. L. Carpenter, R. A. Helliwell, U. S. Inan, T. Wolf, M. Schafer, and VLF, 13 August, Electronic Mail, 1986b.

Lynch, J. T., Letter to R. A. Helliwell, regarding future Siple operations, Personal Letter, 1987.

Matsumoto, H., and I. Kimura, Linear and nonlinear instabilities and VLF emissions in the magnetosphere, *Planet. Sp. Sci.*, *19*, 567, 1971.

Matsumoto, H., K. Hashimoto, and I. Kimura, Dependence of coherent nonlinear whistler interaction on wave amplitude, *J. Geophys. Res.*, *85*, 644, 1980.

Matthews, D. L., Siple Station magnetospheric physics campaign, *Antarctic J.*, *XVI*, 202, 1981.

McPherson, D. A., H. C. Koons, M. H. Dazey, R. L. Dowden, L. E. S. Amon, and N. R. Thompson, Conjugate magnetospheric transmissions of VLF from Alaska to New Zealand, *J. Geophys. Res.*, *79*, 1555, 1974.

Mielke, T. A., C. J. Elkins, R. A. Helliwell, and U. S. Inan, Excitation of whistler mode signals via injection of polarized VLF waves with the Siple transmitter, *Radio Sci.*, *21*, 31, 1992.

Miller, T. R., Siple reception stats 1973-78 (no 1976), Informal Report, 1979.

Nunn, D., A self-consistent theory of triggered VLF emissions, *Planet. Sp. Sci.*, *22*, 349, 1974.

Nunn, D., The numerical simulation of VLF nonlinear wave-particle interactions in collision-free plasmas using the Vlasov hybrid simulation technique, *Comp. Phys. Comm.*, *60*, 1, 1990.

Nunn, D., A novel technique for the numerical simulation of hot collision-free plasma; Vlasov Hybrid Simulation, *J. Comp. Phys.*, *108*, 180, 1993.

Nunn, D., Y. Omura, H. Matsumoto, I. Nagano, and S. Yagitani, The numerical simulation of VLF chorus and discrete emissions observed on the Geotail satellite using a Vlasov code, *J. Geophys. Res.*, *102*, 27,083, 1997.

Nunn, D., A. G. Demenkov, V. Y. Trakhtengerts, and M. J. Rycroft, VLF emission triggering by a highly anisotropic energetic electron plasma, *Ann. Geophys.*, *21*, 481, 2003.

Nunn, D., M. J. Rycroft, and V. Y. Trakhtengerts, A parametric study of the numerical simulations of triggered VLF emissions, *Ann. Geophys.*, *23*, 1, 2005.

Omura, Y., and H. Matsumoto, Computer simulations of basic processes of coherent whistler mode wave-particle interactions in the magnetosphere, *J. Geophys. Res.*, *87*, 4435, 1982.

Omura, Y., and D. Summers, Computer simulations of relativistic whistler-mode wave-particle interactions, *Phys. Plas.*, *11*, 3530, 2004.

Omura, Y., and D. Summers, Dynamics of high-energy electrons interacting with whistler mode chorus emissions in the magnetosphere, *J. Geophys. Res.*, *111*, A09,222, 2006.

Omura, Y., D. Nunn, H. Matsumoto, and M. J. Rycroft, A review of observational, theoretical and numerical studies of VLF triggered emissions, *J. Atm. Terr. Phys.*, *53*, 351, 1991.

Park, C. G., Methods of determining electron concentrations in the magnetosphere from nose whistlers, *Tech. Rep. 3454-1*, Stanford Electronics Laboratories, Stanford, California, 1972.

Park, C. G., Generation of whistler-mode sidebands in the magnetosphere, *J. Geophys. Res.*, *86*, 2286, 1981.

Park, C. G., and D. C. D. Chang, Transmitter simulation of power line radiation effects in the magnetosphere, *Geophys. Res. Lett.*, *5*, 861, 1978.

Paschal, E. W., Whistler precursors on a VLF transmitter signals, *J. Geophys. Res.*, *95*, 225, 1990.

Paschal, E. W., Email to the author, 4 May, Electronic mail, 2008a.

Paschal, E. W., Email to the author 8 May, Electronic Mail, 2008b.

Paschal, E. W., and R. A. Helliwell, Phase measurement of whistler mode signals from the Siple VLF transmitter, *J. Geophys. Res.*, *89*, 1667, 1984.

Pasmanik, D. L., D. A. G., D. Nunn, V. Y. Trakhtengerts, and M. J. Rycroft, Cyclotron amplification of whistler-mode waves: a parametric study relevant to discrete VLF emissions in the Earth's magnetosphere, *J. Geophys. Res.*, *107*, 1162, 2002.

Platino, M., U. S. Inan, T. F. Bell, J. S. Pickett, E. J. Kennedy, J. G. Trotignon, R. J. L., and P. Canu, Cluster observations of ELF/VLF signals generated by modulated heating of the lower ionosphere with the HAARP HF transimitter, *Ann. Geophys.*, *22*, 2643, 2004.

Platino, M., U. S. Inan, T. F. Bell, J. S. Pickett, and P. Canu, Rapidly moving sources of upper-band ELF/VLF chorus near the magnetic equator, *J. Geophys. Res.*, *111*, A09,218, 2006.

Raghuram, R., R. L. Smith, and T. F. Bell, VLF antarctic antenna: Impedence and efficiency, *IEEE Trans. on Ant. and Prop.*, *AP-22*, 334, 1974.

Raghuram, R., T. F. Bell, R. A. Helliwell, and J. P. Katsufrakis, A quiet band produced by VLF transmitter signals in the magnetosphere, *Geophys. Res. Lett.*, *4*, 199, 1977a.

Raghuram, R., T. F. Bell, R. A. Helliwell, and J. P. Katsufrakis, Echo-induced suppressions of coherent VLF transmitter signals in the magnetosphere, *J. Geophys. Res.*, *82*, 2787, 1977b.

Rastani, K., U. S. Inan, and R. A. Helliwell, De 1 observations of Siple transmitter signals and associated sidebands, *J. Geophys. Res.*, *90*, 4128, 1985.

Rosenberg, T. J., and J. R. Barcus, Energetic particle precipitation from the magnetosphere, in *Upper Atmospheric Research in Antarctica*, edited by L. J. Lanzerotti and C. G. Park, p. 42, Antarctic Research Service, 1978.

Rosenberg, T. J., R. A. Helliwell, and J. P. Katsufrakis, Electron perecipitation associated with discrete very-low-frequency emissions, *J. Geophys. Res.*, *76*, 8445, 1971.

Roux, A., and R. Pellat, A theory of triggered emissions, *J. Geophys. Res.*, *83*, 1433, 1978.

Sa, L. A. D., A wave-particle-wave interaction mechanism as a cause of VLF triggered emissions, *J. Geophys. Res.*, *95*, 12,277, 1990.

Sa, L. A. D., and R. A. Helliwell, Structure of VLF whistler mode sideband waves in the magnetosphere, *J. Geophys. Res.*, *93*, 1987, 1988.

Sazhin, S. S., H. Hayakawa, and K. Bullough, Whistler diagnostics of magnetospheric parameters: a review, *Ann. Geophys.*, *10*, 293, 1992.

Serra, F. M., VLF two-wave-electron interactions in the magnetosphere, *Planet. Sp. Sci.*, *32*, 985, 1984.

Smith, A. J., and D. Nunn, Numerical simulation of VLF risers, fallers, and hooks observed in Antarctica, *J. Geophys. Res.*, *103*, 6771, 1998.

Smith, R. L., Propagation characteristics of whistlers trapped in field-aligned columns of enhanced ionization, *J. Geophys. Res.*, *66*, 3699, 1961.

Sonwalkar, V. S., and U. S. Inan, Measurements of Siple transmitter signals on the de 1 satellite: wave normal direction and antenna effective length, *J. Geophys. Res.*, *91*, 154, 1986.

Sonwalkar, V. S., T. F. Bell, R. A. Helliwell, and U. S. Inan, Direct multiple path magnetospheric propagation: a fundamental property of nonducted VLF waves, *J. Geophys. Res.*, *89*, 2823, 1984.

Sonwalkar, V. S., D. L. Carpenter, R. A. Helliwell, M. Walt, U. S. Inan, D. L. Claude, and M. Ikeda, Properties of the magnetospheric hot plasma distribution deduced from whistler mode wave injection at 2400 hz: Ground-based detection of azimuthal structure in magnetospheric hot plasmas, *J. Geophys. Res.*, *102*, 14,363, 1997.

Stiles, G. S., and R. A. Helliwell, Frequency-time behavior of artificially stimulated VLF emissions, *J. Geophys. Res.*, *80*, 608, 1975.

Stiles, G. S., and R. A. Helliwell, Stimulated growth of coherent VLF waves in the magnetosphere, *J. Geophys. Res.*, *82*, 1523, 1977.

Sudan, R. N., and E. Ott, Theory of triggered VLF emissions, *J. Geophys. Res.*, *76*, 4463, 1971.

Trabucco, W. J., Letter to U. S. Inan from McMurdo Station, Personal Letter, 1987a.

Trabucco, W. J., Siple Station report, Internal Memorandum, 1987b.

Trabucco, W. J., Notes, Personal Notes, c.1990.

Trakhtengerts, V. Y., A. G. Demenkov, D. L. Pasmanik, E. E. Titora, B. V. Kozelov, D. Nunn, and M. J. Rycroft, Highly anisotropic distributions of energetic electrons and triggered VLF emissions, *Geophys. Res. Lett.*, *28*, 2577, 2001.

Walt, M., *Introduction to Geomagnetically Trapped Radiation*, Cambridge University Press, 1994.

Wolf, T., Email to J. Logan, 7 November, Electronic Mail, 1986.