Annales Geophysicae Manuscript NT.
(will be inserted by hand later)

On recent developments in E–region irregularity simulations and a summary of related theory

P. Janhunen

Finnish Meteorological Institute, Department of Geophysics

Received: 22 June 1994 / Accepted: 17 February 1995

Abstract. Theoretical and simulation approaches to E–region irregularities (gradient drift and Farley-Buneman instabilities) are reviewed, with some account on relevant observations also. As a new material, a hybrid linear dispersion relation is derived and presented. The most important problem that the straightforward theories cannot explain is the saturation of the phase velocity to the ion acoustic speed (C_s saturation). This phenomenon is well–known from the equatorial electrojet radar observations. Recent particle simulations yield an interesting new explanation for the C_s saturation, named flow angle stabilization: The phase velocity is actually not C_s saturated, but the flow angle distribution of the spatial power spectrum is highly asymmetric. The asymmetry is such that the most intense waves propagate at the $\mathbf{k} \cdot \mathbf{E} < 0$ edge of the linearly unstable sector, and thus the phase velocity of the most intense waves is close to C_s. Depending on the level of larger scale turbulence, the radar observes varying degrees of C_s saturation. If the larger scale turbulence level is high (equatorial electrojet case), the local flow angle fluctuates and there are always subregions inside scattering volume that have their local flow angle favourable for the detection of the most intense waves. Under these conditions the spectra show C_s saturation. If the larger scale turbulence level is lower, there will not always be enough mixing of the flow angle to see the most intense waves at all. In these cases the mean Doppler shift will be proportional to the electric field, i.e. it will obey the linear theory.

1 Introduction

Coherent VHF scatter radars are widely used in monitoring the ionospheric electric field. They can follow the two-dimensional electric field pattern in large areas, typically several hundred kilometers, with good time resolution. This property makes these radars rather unique among the observational methods of ionospheric plasma.

However, coherent E–region radars do not measure the electric field directly, but instead they measure the spectral properties of plasma waves driven unstable by the electric field. In order to estimate the electric field inside radar scattering volume, we should therefore know quantitatively what kind of waves will develop in given electric field conditions. The generated waves will also depend on plasma parameters such as ion-neutral and electron-neutral collision frequencies and large-scale electron density gradient.

Farley (1963) and Buneman (1963) showed that the plasma is unstable against a specific type of modified two-stream instability, later called the Farley–Buneman instability, if the ambient electric field $\mathbf{E}$ is strong enough. There are many factors affecting the growth rate, the main trend being that the larger the ion-electron velocity difference $|\mathbf{V}_{de} - \mathbf{V}_{di}|$, the stronger the instability. Regardless of velocity difference, the Farley–Buneman instability is unlikely everywhere below about 95 km, because of collisional damping. At high altitudes (more than 120 km), both ions and electrons $\mathbf{E}\times\mathbf{B}$ drift, and the instability is quenched because of the lack of velocity difference. Therefore the lower E–region, about 95 to 115 km altitude range, is the primary place where Farley–Buneman unstable waves may be found in both equatorial and auroral electrojets.

It is now widely accepted that the Farley-Buneman and the gradient drift instability are the instabilities mainly or even totally responsible for meter-scale waves seen by VHF coherent scatter radars. The linear theory of these instabilities is well known, but the nonlinear saturation of Farley-Buneman waves has remained an open question.

The structure of the paper is the following. We first go through the analytic two-fluid linear theory, discussing the form of the unstable region as a function of several parameters. Then follows a brief introduction to kinetic theory, and a derivation of a hybrid fluid-kinetic dispersion relation. The hybrid relation is compared with linear theory in several plots. Radar and rocket observations as well as laboratory investigations are treated to the extent needed in subsequent sections. Fluid and particle simulations are then reviewed. Some previously unpublished two-dimensional fluid simulation results are

Correspondence to: Pekka Janhunen

also presented. Then follows a section on various proposed nonlinear models to explain the observations and simulation results. The paper ends with a concluding section on current status and future directions.

This paper has been adapted from the PhD thesis of the author (Janhunen, 1994).

2 Linear theory

In a mathematical treatment of plasma instabilities it is usually assumed that the medium is initially homogeneous, allowing for a possible linear trend. We shall call such a medium laminar. If a very small sinusoidal wave perturbation is superposed to the laminar background, it is possible to compute the phase velocity and growth rate of the plane wave. The wave amplitude time dependence is $\exp(-i\omega t)$, where ω is the complex angular frequency, $\omega = \omega_r + i\gamma$, where $\omega_r = \mathrm{Re}\,\omega$, $v_{ph} = \omega_r/|\mathbf{k}|$ is the phase speed, and $\gamma = \mathrm{Im}\,\omega$ is the growth rate. The complex variable ω and the real wavenumber vector $\mathbf{k}$ are related by the linear disperion relation. In the fluid approximation, the dispersion relation is a polynomial in ω, in the kinetic model it is a more complicated transcendental equation. Usually the dispersion equation will have multiple roots; it is then the ω with largest imaginary part (growth rate) which is of most interest.

This standard linearized approach is frequently used since it presents only modest computational difficulty and because it provides a classification of wave modes (e.g., Sudan et al., 1973; Fejer and Kelley, 1980; Farley, 1985). However, as soon as the growth rate $\gamma(\mathbf{k}) > 0$ for any $\mathbf{k}$, the associated mode grows exponentially until it is saturated by nonlinear processes. After this, the assumptions of the linear formalism break down since the perturbations are no longer small. This pitfall concerns the system as a whole, not just the linearly unstable modes. The point is that the linearized theory can only be used to test whether the system is quiescent or unstable, but, strictly speaking, not to derive its properties if it happens to be unstable.

2.1 Two-fluid approach

The fluid model of a plasma is formally based on the first few moments of the Boltzmann equation (e.g., Krall and Trivelpiece, 1973; Ichimaru, 1973; Schunk, 1977). In the two-fluid approach the electrons and the ions are treated as separate fluids. The neutral atmosphere is usually not explicitly modelled, since we are generally interested in time scales much shorter than it takes to move or heat the neutral gas significantly by plasma collisions.

Applied to the E–region the continuity and momentum equations for charged fluids are in the rest frame of neutrals (e.g., Sudan et al., 1973; Fejer et al., 1984b)

$$\frac{\partial n_s}{\partial t} = -\nabla\cdot(n_s\mathbf{v}_s) - L_s + Q_s \tag{1}$$

$$\frac{\partial \mathbf{v}_s}{\partial t} + \mathbf{v}_s\cdot\nabla\mathbf{v}_s = \frac{q_s}{m_s}(\mathbf{E} + \mathbf{v}_s\times\mathbf{B}) - \frac{k_B T_s}{m_s}\frac{\nabla n_s}{n_s} - \nu_s\mathbf{v}_s \tag{2}$$

where s labels the particle species, n_s is the number density, $\mathbf{v}_s$ is the flow velocity, L_s is the loss rate, Q_s is the production rate, q_s is the charge of sth species, m_s is the mass of sth species, $\mathbf{E}$ is the electric field, $\mathbf{B}$ is the (external) magnetic field, k_B is Boltzmann's constant, T_s is the temperature of the sth fluid (both species are assumed isothermal) and ν_s is the sth particle-neutral collision frequency. In our case, $s = i, e$.

In addition, we need to specify a relation linking the electrostatic electric field $\mathbf{E} = -\nabla\phi$ to the other quantities. In principle, Poisson's equation

$$-\nabla^2\phi = \frac{1}{\epsilon_0}\sum_s q_s n_s \tag{3}$$

can always be used. A typical value for the electron Debye length is, however, less than one centimeter. As we are interested in wavelengths of the order of one meter or longer, the plasma is quasineutral in a good approximation, and instead of Poisson's equation we can just use the relation $n_e = n_i = n$. The continuity equations for both species then imply $\nabla\cdot(n\mathbf{v}_e) = \nabla\cdot(n\mathbf{v}_i)$, which can also be thought to follow from the current density continuity equation, $\nabla\cdot\mathbf{j} = 0$. It is also a good approximation to neglect the time derivative terms in the electron momentum equation, because we are dealing with frequencies much smaller than the electron cyclotron frequency. Neglecting the recombination and production terms is usually justified as well (Fejer et al., 1984b), and if we are in the lower E–region, we can assume that ions are unmagnetized. Carrying out all these approximations, we are left with the following system of equations:

$$\frac{\partial n}{\partial t} = -\nabla\cdot(n\mathbf{v}_i) \tag{4}$$

$$\frac{\partial \mathbf{v}_i}{\partial t} + \mathbf{v}_i\cdot\nabla\mathbf{v}_i = \frac{q_i}{m_i}\mathbf{E} - \frac{k_B T_i}{m_i}\frac{\nabla n}{n} - \nu_i\mathbf{v}_i \tag{5}$$

$$\frac{1}{1+\psi_e^2}\left[\frac{\psi_e}{B}\nabla\cdot(n\nabla\phi') - \frac{1}{B^2}\mathbf{B}\cdot\nabla n\times\nabla\phi'\right] = \nabla\cdot(n\mathbf{v}_i). \tag{6}$$

We have introduced the notation $\psi_s = \nu_s/\Omega_s$ ($s = i, e$) and

$$\phi' = \phi - \frac{1}{e}k_B T_e \log n \tag{7}$$

where $\mathbf{E} = -\nabla\phi$ and ϕ' is an effective potential. We have also assumed that no quantities depend on the field–aligned coordinate.

We use a coordinate system with the x axis parallel to the electric field $\mathbf{E}$, y parallel to $\mathbf{E}\times\mathbf{B}$, and z antiparallel to the geomagnetic field $\mathbf{B}$ (Fig. 1). The angle between $\mathbf{k}$ vector and the y axis is called the flow angle θ. The angle between $\mathbf{k}$ and the $\mathbf{B}$-perpendicular plane is called the aspect angle α. In Eq. (6) above we have for simplicity assumed zero aspect angle, however, in what follows we quote linearized results that contain the aspect angle dependence. For a version of Eq. (6) with parallel effects included, see for example Janhunen (1992), equation (33). The Cartesian components k_x, k_y, k_z of $\mathbf{k}$ are obtained from

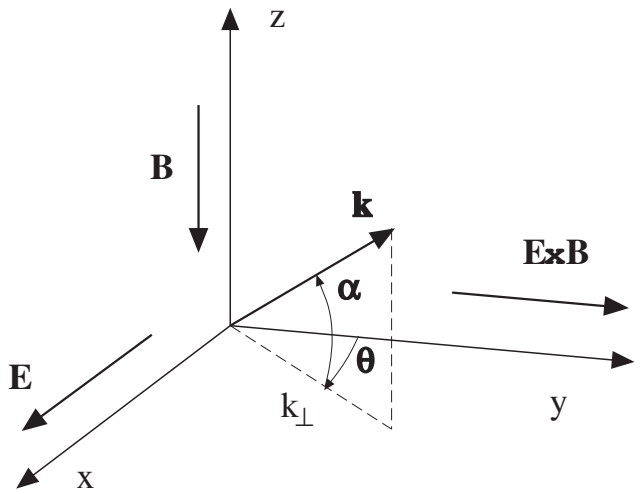

Fig. 1. This coordinate system is employed throughout, unless otherwise stated. The neutrals are at rest in the coordinate system.

$$
\begin{aligned}
k_x &= k_\perp \sin\theta \\
k_y &= k_\perp \cos\theta \\
k_z &= k_\parallel = k \sin\alpha \\
k_\perp &= k \cos\alpha .
\end{aligned} \tag{8}
$$

The linearization of the fluid equations yields the dispersion relation

$$
\omega - k_y V_d = \left(\frac{\psi}{\nu_i} - \frac{i k_y}{\mathbf{k}^2 L \Omega_i} \right) \left[\omega(i\omega - \nu_i) - i k^2 C_s^2 \right] , \tag{9}
$$

where

$$
\psi = \frac{\nu_e \nu_i}{\Omega_e \Omega_i} \left(1 + \frac{\Omega_e^2}{\nu_e^2} \sin^2\alpha \right) \tag{10}
$$

(Sudan et al., 1973; Fejer et al., 1984b). Here L is the zeroth order electron density gradient scale length in the x direction, i.e. $n(x) = n_0(1 + x/L)$. The ion acoustic velocity C_s is given by $C_s = \sqrt{[k_B(T_e + T_i)/m_i]}$. Recall that we are in this section all the time assuming isothermality, and this relation for C_s holds only under that assumption. In order to obtain a simple tractable solution of Eq. (9) for $\omega = \omega_r + i\gamma$ ($\gamma = \mathrm{Im}\,\omega, \omega_r = \mathrm{Re}\,\omega$), we need to assume that L is large in some sense. Let us first obtain an expression for ω_r for marginally stable ($\gamma = 0$) waves. Taking the real part of Eq. (9) while assuming a real ω yields a quadratic equation for ω. If the discriminant is expanded to first order in $V_d/L\Omega_i$, we obtain for the real angular frequency

$$
\omega_r = \frac{k_y}{1+\psi} \left(V_d - \frac{C_s^2}{L\Omega_i} \right) . \tag{11}
$$

A typical value for both V_d and C_s is 450 m/s. Typical Ω_i is 150 s^{-1}. Therefore, if L is larger than 500 m, we can accurately use the expression

$$
\omega_r = \frac{k_y V_d}{1+\psi} = \frac{k V_d \cos\theta}{1+\psi} \tag{12}
$$

for the real angular frequency.

Assuming $|\gamma| \ll |\omega_r|$, we also readily obtain a solution for the growth rate γ:

$$
\gamma = \frac{1}{1+\psi} \left[\frac{\psi}{\nu_i} \left(\omega_r^2 - \mathbf{k}^2 C_s^2 \right) + \frac{\omega_r \nu_i k_y}{\mathbf{k}^2 L \Omega_i} \right] \tag{13}
$$

(Fejer et al., 1984b). The symbol ω_r in Eq. (13) is to be substituted from Eq. (12).

It should be noted that above threshold Eq. (13) produces the maximum growth rate at nonzero aspect angle α, as was noted by Wang and Tsunoda (1975) and also by Fejer et al.(1984b). In Fig. (2) we plot the growth rate γ as a function of aspect angle. Both fluid and a more accurate hybrid dispersion relation (see below) are shown, and it is seen that the fluid result is qualitatively correct at 2 meter wavelength and 800 m/s drift. The growth rate is a symmetric function of the aspect angle. Notice, however, that although the growth rate peaks for nonzero aspect angle if the growth rate is positive, zero aspect angle waves first become unstable when the electric field exceeds the threshold.

The assumption $|\gamma| \ll |\omega_r|$ is often justified in practice. In particular, it is by definition ideally justified if we are looking for the boundary of the unstable region, $\{\mathbf{k} | \gamma(\mathbf{k}) = 0\}$. The curve $\gamma = 0$ in (k_y, k_x) plane can be solved analytically. The magnitude k of the vector $\mathbf{k}$ as a function of the flow angle θ is obtained as

$$
k^2 = \frac{\left(\frac{1+\psi}{\psi_e} \right) \left(\frac{\nu_i}{V_d L} \right)}{\left[\frac{(1+\psi) C_s}{V_d \cos\theta} \right]^2 - 1} . \tag{14}
$$

Eq. (14) is equivalent to

$$
\cos^2\theta = \frac{\left[(1+\psi) C_s / V_d \right]^2}{1 + \frac{1+\psi}{\psi_e} \frac{\nu_i}{V_d k^2 L}} , \tag{15}
$$

from which it is easily seen that in the limit $L \to \infty$ the unstable region becomes a sector (i.e., θ becomes independent of k).

The smallest drift velocity V_d for which γ vanishes when $L = \infty$, or the corresponding electric field $E = V_d B$, is called the Farley–Buneman threshold. In fluid theory the threshold velocity is given by $V_d^{thr} = (1 + \psi) C_s$. If $V_d < (1 + \psi) C_s$, all waves are damped, unless there is a density gradient with $\mathbf{E} \cdot \nabla n > 0$. In the last case the unstable region is a "droplet" concentrated on large-scale (several tens of meters) waves (Fig. 3, topright panel).

When V_d exceeds the threshold, the droplet quickly "opens up" to an instability sector. Within the sector, the growth rate increases without limit as k increases. This is not, of course, physically correct, but is an artifact of the fluid theory. Fig. (3) shows the instability region for three different V_d and L values. In the first column L is negative and thus the density gradient is a stabilizing one ("unfavourable"). The instability region is completely absent except in the bottom plot. The middle column displays the $L = \infty$ case, in which case the instability regions are sectors. The effect of a destabilizing L ("favourable") is seen in the rightmost column,

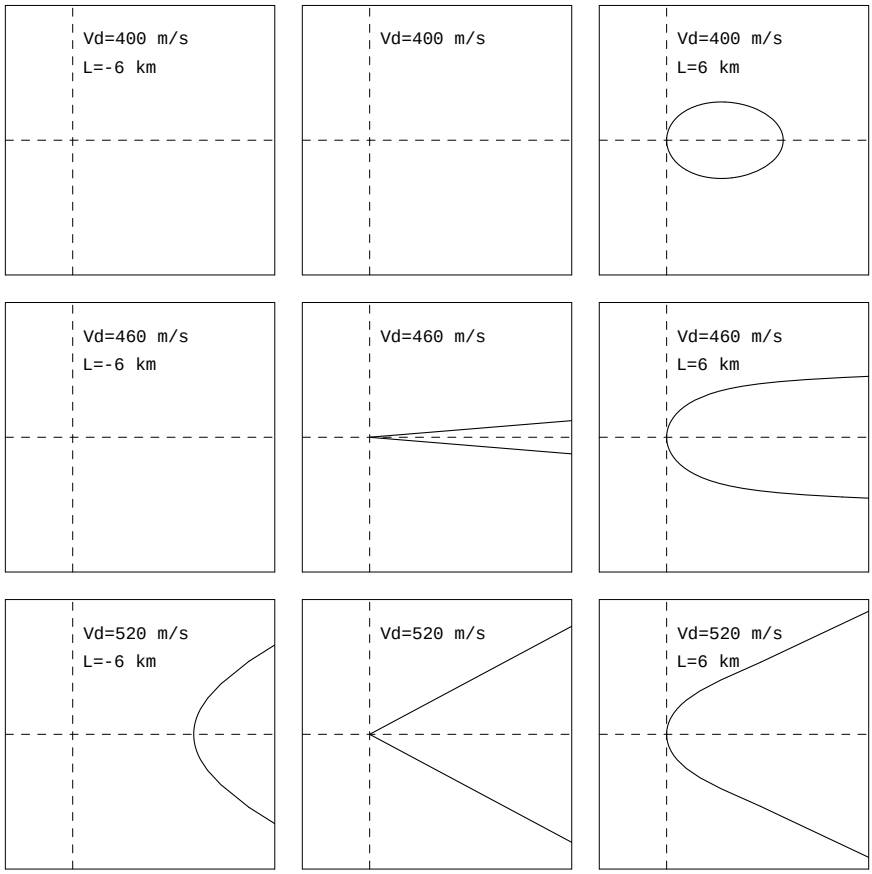

Fig. 3. Unstable region in **k** plane for different combinations of V_d and L. The vertical axis is parallel to the electric field, and the horizontal axis is parallel to the $\mathbf{E}\times\mathbf{B}$ direction. In the middle column the value of L is ∞.

where the region transforms from a droplet shape to a distorted sector as V_d grows.

Although the linear theory growth rate should not be taken literally whenever it is positive, its magnitude nevertheless gives some indication of how the nonlinear system will tend to behave. If the growth rate is large in some region of **k** space, the corresponding modes (Fourier components in **k** space) will probably be present in the nonlinearly saturated spectrum, too. Fig. (4) shows how the droplet opens up to a sector and how the growth rate soon becomes large inside the sector. Considering that the electric field changes only 15 % in Fig. (4), the name "threshold" is by no means misleading.

2.2 Kinetic theory approach

The starting point for kinetic calculations is the Boltzmann equation for both species $s = i, e$:

$$\frac{\partial f_s}{\partial t} + \mathbf{v}_s \cdot \nabla f_s + \frac{q_s}{m_s}(\mathbf{E} + \mathbf{v}\times\mathbf{B}) \cdot \nabla_v f_s = \left(\frac{\partial f_s}{\partial t}\right)_c . \quad (16)$$

The linearization of Eq. (16) requires the specification of the unperturbed distribution functions for electrons and ions. Furthermore, it requires a simple tractable form for the neutral collision term $(\partial f_s/\partial t)_c$, otherwise the problem seems too difficult to handle even by numerical means.

The original kinetic theory by Farley (1963) already contained the most important ingredients. It employed the Bhatnagar-Gross-Krook (BGK) collision term:

$$\left(\frac{\partial f_s}{\partial t}\right)_c = -\nu_s \left[f_{s1}(\mathbf{r},\mathbf{v},t) - \frac{N_{s1}(\mathbf{r},t)}{N_{s0}} f_{s0}(\mathbf{v},t) \right] \quad (17)$$

(Bhatnagar et al., 1954), and considered only fully perpendicular waves without ambient electron density gradient. Here,

$$N_{s0,1}(\mathbf{r},t) = \int d^3\mathbf{v} f_{s0,1}(\mathbf{r},\mathbf{v},t). \quad (18)$$

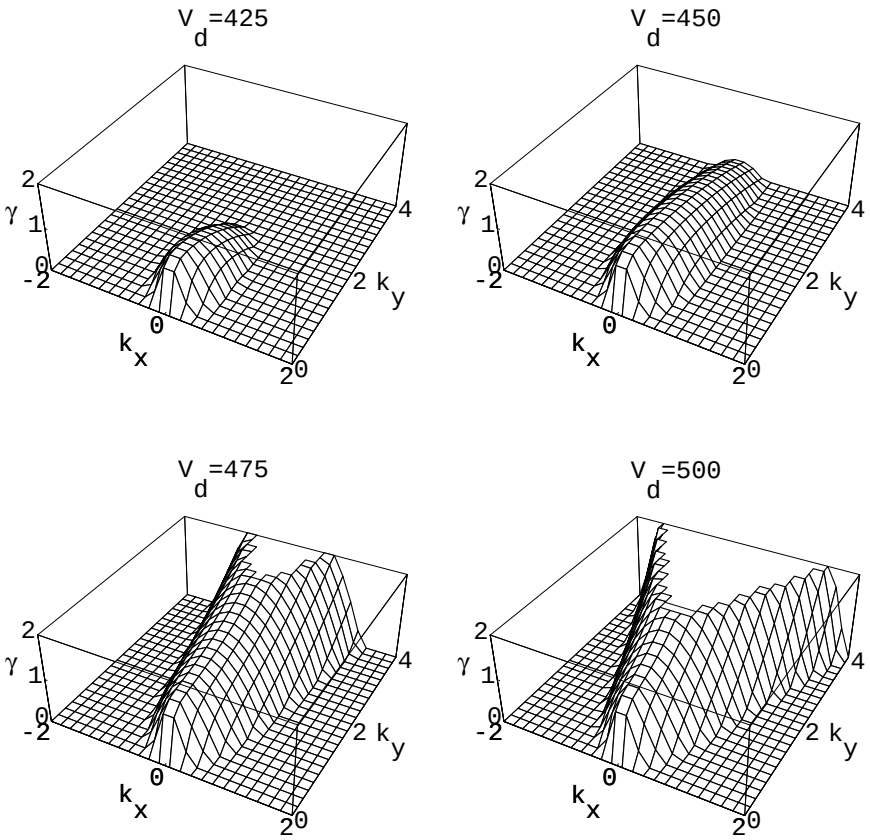

Fig. 4. Fluid theory growth rate in **k** plane as a series of 3D plots for four drift values ranging from 425 m/s to 500 m/s. Negative growth rates have been replaced with zero. The gradient scale length L is 6 km, and other variables have their typical E-region values.

New formulations and extensions have thereafter been constructed. Schmidt and Gary (1973) showed that the electrons can practically always be handled as a fluid, but the ions must be treated kinetically unless the wavelength is much larger than the ion mean free path. More specifically, the electrons are fluidlike for wavelengths much larger than the electron Larmor radius and Debye length, which are both generally less than one centimeter. For the ions, on the other hand, the Landau damping is important for wavelengths less than about half a meter.

Parallel propagation effects were studied by Ossakow et al.(1975). They found that the fastest growing modes have small but finite $k_\parallel$. Similar result also follows from fluid theory (Wang and Tsunoda, 1975; Fejer et al., 1984b). The case of nonequal electron and ion temperatures was elaborated by Schlegel and St.-Maurice (1982) and Schlegel (1983).

The collision term (Eq. 17) is the weakest point of the kinetic theory. It is a simple relaxation model, which has been corrected to conserve particle density locally. It drives the whole distribution function toward an equilibrium in a common time scale. Therefore, the mean velocity and temperature, for instance, relax in the same time scale. This is qualitatively correct for the ions since the ion and neutral masses are comparable, but not for the electrons, for which the thermalization time scale should be much longer (by factor of m_n/m_e, if only elastic collision are considered) than the momentum transfer time scale. In effect, the BGK collision model forces the fluctuations to be isothermal (Bhatnagar et al., 1954; Farley, 1963). In reality, inelastic electron-neutral collisions are much more important than elastic ones, and thus the error with BGK is not as large as m_n/m_e, but it still remains fairly large.

Stubbe (1989) presented an improved collision term, which involved separate time scales for momentum transfer ($1/\nu_s^{(m)}$), isotropization ($1/\nu_s^{(r)}$), and energy transfer ($1/\nu_s^{(e)}$), $s = i, e$. As a relatively good approximation, $\nu_s^{(r)} = \nu_s^{(m)}$, $s = i, e$, and $\nu_i^{(e)} = \nu_i^{(m)}$, that is, ions have only one time scale, but the electron energization time differs from the momentum transfer and isotropization time. Unfortunately, Stubbe's full expressions are complicated to such an extent that, to our knowledge, no one has thus far solved the resulting dispersion relation numerically.

In summary, a satisfactory and tractable kinetic theory for partially ionized gases is still an open problem. In the ionospheric E–region this implies that one is forced to use approximate fluid, kinetic, or hybrid models of the plasma, and one must therefore evaluate the validity of the assumptions separately in each case.

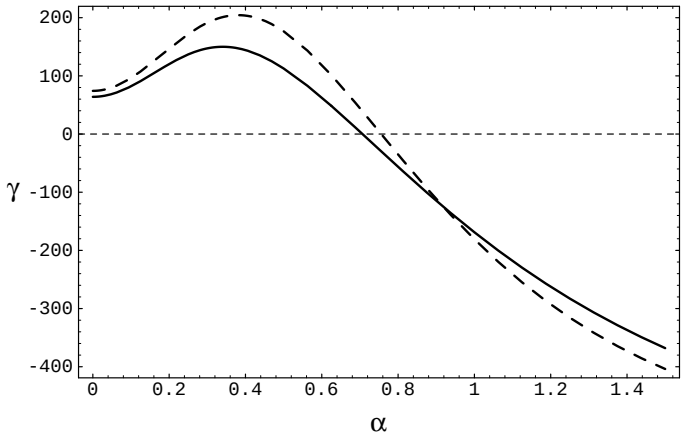

Fig. 2. Linear growth rate γ (s^{-1}) as a function of aspect angle α (degree) for zero flow angle, 800 m/s electron drift, and 2 m wavelength. Dashed curve is the fluid result, Eq. (13), solid curve is a hybrid dispersion relation, Eq. (36). The hybrid dispersion relation is more accurate than the fluid model and it is introduced in a later subsection below. The values T_i = 230 K, T_e = 350K, $m_i = 5.18{\times}10^{-26}$ kg, C_s = 390 m/s and V_d^{thr} = 410 m/s have been used in this and other figures, unless otherwise stated. These parameter values are for the auroral region.

2.3 Hybrid dispersion relation

In this subsection we derive and present a hybrid dispersion relation, where the electron gas is treated as a fluid in the 5-moment approximation, while the ions are kinetic and unmagnetized. In the 5-moment approximation the continuity and momentum equations are accompanied by an energy equation with isotropic pressure. The resulting dispersion relation is suitable for numerical computations, and various analytical limits are easily derived.

Using the fluid model for the electrons is a good approximation in almost all cases (Schmidt and Gary, 1973; Stubbe 1989; Farley and Providakes, 1989). Only for large aspect angles there is a real need for full kinetic treatment (Stubbe, 1989). Using the 5-moment approximation, we retain the temperature fluctuations, and thus relax the most important limitation of the standard isothermal theory.

The electron fluid equations are (Schunk, 1977):

$$\frac{\partial n_e}{\partial t} + \nabla \cdot (n_e \mathbf{v}_e) = 0 \tag{19}$$

$$n_e m_e \left(\frac{\partial \mathbf{v}_e}{\partial t} + \mathbf{v}_e \cdot \nabla \mathbf{v}_e \right) + \nabla p_e - n_e q_e (\mathbf{E} + \mathbf{v}_e \times \mathbf{B}) = -n_e m_e \Phi \nu_{en}^{(m)} \mathbf{v}_e \tag{20}$$

$$\left(\frac{\partial}{\partial t} + \mathbf{v}_e \cdot \nabla \right) p_e + \frac{5}{3} p_e \nabla \cdot \mathbf{v}_e = \frac{n_e m_e \nu_{en}^{(m)}}{m_e + m_n} \left[2k_B (T_n - T_e) \Psi + \frac{2}{3} m_n \mathbf{v}_e^2 \Phi \right]. \tag{21}$$

Here Φ and Ψ are dimensionless correction factors (Schunk, 1977). By taking the limit $\Psi \to \infty$, we recover the usual isothermal results. By setting Ψ larger than unity but finite, we can model electron-neutral inelastic collisions. Stubbe (1989) effectively used $\Psi \approx (10^4 K)/T_{e0}$, making Ψ about 30. Inelastic collisions are ignored if Ψ is set to unity. Notice that Ψ and Φ have nothing to do with the ψ parameter (10) or electric potential ϕ.

After linearizing Eqs. (19) to (21) and replacing $\partial/\partial t$ by $-i\omega$ and ∇ by $i\mathbf{k}$, we eliminate the pressure and velocity perturbations, starting from the higher moment equations and proceeding toward the continuity equation (19). As a result, the electron density perturbation n_1^e is expressed as linearly proportional to the assumed electric potential perturbation ϕ_1. We write this relationship in the form

$$\frac{n_1^e}{n_0} = -\frac{q_e}{k_B T_{0e}} F_e \phi_1 \tag{22}$$

where F_e is a dimensionless function having the form

$$F_e = i \frac{\Delta_e}{\tilde{\omega}_e + i \dfrac{\Delta_e}{1 - i \frac{\tilde{\omega}_e}{2\Psi \nu_{en}^{(m)}} (\frac{m_n}{m_e} + 1)}} \tag{23}$$

where

$$\Delta_e = \left(\mathbf{k} - i \frac{\nabla n_0}{n_0} \right) \cdot \left[D_e^{\perp} (\mathbf{k} + \frac{1}{\tilde{\nu}} \mathbf{k} \times \Omega_e) + D_e^{\parallel} k_{\parallel} \hat{\Omega}_e \right] \tag{24}$$

$$D_e^{\perp} = \frac{k_B T_{0e} \tilde{\nu}}{m_e (\Omega_e^2 + \delta \cdot \mathbf{k} + \tilde{\nu}^2)} \tag{25}$$

$$D_e^{\parallel} = \frac{k_B T_{0e}}{m_e \tilde{\nu}} \left(\frac{\Omega_e^2}{\Omega_e^2 + \delta \cdot \mathbf{k} + \tilde{\nu}^2} \right) \tag{26}$$

$$\delta = \tilde{\nu} \mathbf{w} + \Omega_e \times \mathbf{w} + \frac{1}{\tilde{\nu}} (\mathbf{w} \cdot \Omega_e) \Omega_e \tag{27}$$

$$\mathbf{w} = \frac{\frac{5}{3} i \frac{k_B T_{0e}}{m_e} \mathbf{k} + \frac{\nabla p_{0e}}{n_0 m_e} - \frac{4}{3} \frac{m_n}{m_n + m_e} \Phi \nu_{en}^{(m)} \mathbf{v}_{e0}}{\tilde{\omega}_e + 2i \frac{m_e}{m_n + m_e} \Psi \nu_{en}^{(m)}} \tag{28}$$

$$\tilde{\nu} = \Phi \nu_{en}^{(m)} - i \tilde{\omega}_e \tag{29}$$

$$\tilde{\omega}_e = \omega_e - \mathbf{k} \cdot \mathbf{v}_{0e}. \tag{30}$$

The notation is similar to Robinson (1992) wherever feasible. The quantities $D_e^{\perp}$ and $D_e^{\parallel}$ can be interpreted as electron diffusion coefficients in the perpendicular and parallel directions, respectively (Robinson, 1992). No approximations, other than linearity, have been made in the above equations.

The usual isothermal results in the absence of density gradients can be recovered by taking the limit $\Psi \to \infty$, ignoring electron inertia, setting Φ to unity, and assuming $\nabla n_0 = 0 = \nabla p_{0e}$. Then $\tilde{\nu} = \nu_{en}^{(m)}$, $\mathbf{w} \to 0$, $\delta \to 0$, and the result reduces to that of Robinson (1992). (Notice that Robinson's equations (3) and (5) contain an erroneous factor of imaginary unit i.)

For the ions, we write the density-potential relationship in the form

$$\frac{n_1^i}{n_0} = -\frac{q_i}{k_B T_{0i}} F_i \phi_1, \tag{31}$$

in analogy with electrons (cf. Eq. (22)). The expression for F_i in the unmagnetized BGK approximation becomes (e.g., Ossakow et al., 1975):

$$F_i = \frac{1 + zZ(z)}{1 + \frac{i\nu_{in}}{k}\sqrt{a}Z(z)}, \tag{32}$$

where

$$z = \frac{(\omega - \mathbf{k}\cdot\mathbf{v}_{0i} + i\nu_{in})\sqrt{a}}{k}, \tag{33}$$

$$a = \frac{m_i}{2k_B T_{0i}}. \tag{34}$$

and $Z(z)$ is the plasma dispersion function for any complex z,

$$Z(z) = \frac{1}{\sqrt{\pi}} \int_{-\infty+i\gamma}^{\infty+i\gamma} du \frac{e^{-u^2}}{u - z} \qquad (\gamma < \mathrm{Im}\ z) \tag{35}$$

(Fried and Conte, 1961).

The dispersion relation is then obtained from Gauss' law as

$$F_e + \frac{T_{0e}}{T_{0i}} Z_i^2 F_i + (k\lambda_{De})^2 = 0 \tag{36}$$

where

$$\lambda_{De} = \sqrt{\frac{\epsilon_0 k_B T_{0e}}{n_0 e^2}} \tag{37}$$

is the electron Debye length, and Z_i is the ion charge number, usually $Z_i = 1$. It is usually the case that our wavelength is much larger than the Debye length; we might then drop the Debye length term from the dispersion equation, but as this is not a major simplification we will not do it.

Let us present a few applications of the hybrid dispersion relation Eq. (36), as compared with the fluid relation Eq. (13). Fig. (2) already showed a comparison as a function of aspect angle. Fig. (5) compares the hybrid and fluid equations by plotting the growth rate γ versus the electron drift V_d. The fluid theory gives a higher growth rate and lower threshold than the hybrid theory, but the difference in threshold is only about 5 % for the one-meter wavelength considered. In phase velocity the difference between fluid and hybrid curves in similar conditions as Fig. (5) would be almost invisible.

In the plots of this section we use the following parameters, unless otherwise stated: ion temperature T_i = 230 K, electron temperature T_e = 350K, ion-neutral collision frequency ν_i = 2500 Hz, electron-neutral collision frequency ν_e = 25 kHz, $\Psi = \infty$ (isothermal electrons), electron density $n_0 = 10^{11}$ m^{-3}. These parameters are suitable for the auroral region.

Fig. (6) displays the relative growth rate γ/ω_r as a function of wavelength for 500 m/s electron drift for both hybrid and fluid equations. The fluid curve (dashed line) rises without limit as the wavelength is made smaller,

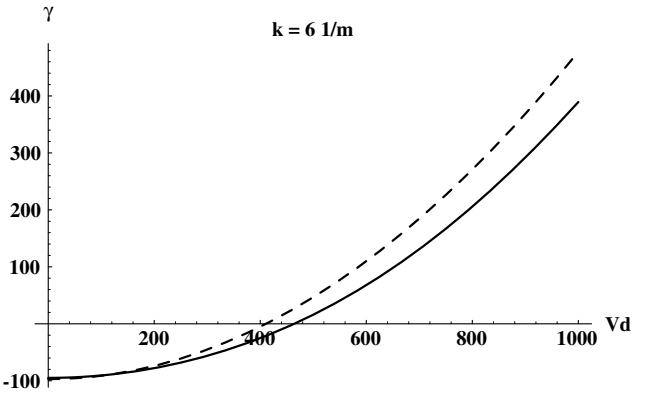

Fig. 5. Hybrid theory (solid line) and fluid theory (dashed line) growth rate as a function of the electron drift (m/s). The wavelength is ≈ 1 m.

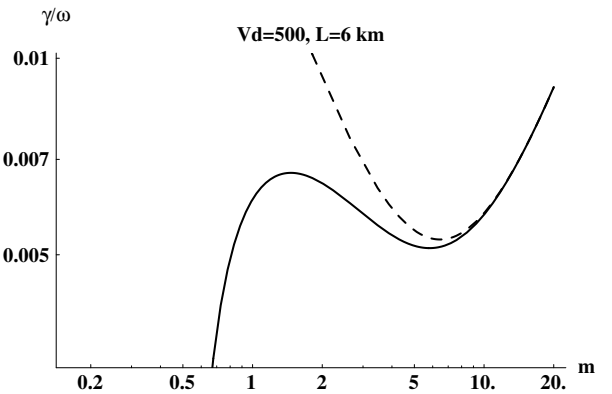

Fig. 6. Relative growth rate γ/ω_r as a function of wavelength for hybrid theory (solid line) and fluid theory (dashed line). The electron drift is 500 m/s. Notice logarithmic scale in both axes.

and therefore only part of the fluid curve is shown. We have also used a 6 km density gradient that enhances the instability; therefore the relative growth rate rises for more than 10 meter wavelengths. The maximum relative growth rate occurs at about 2 meter waves in this case. The difference between fluid and hybrid curves seems drastic, because the growth rate is all the time very small (500 m/s is only slightly above the threshold in this case). Fig. (7) is similar to Fig. (6), except that the electron drift is now 600 m/s. This causes the largest rel-

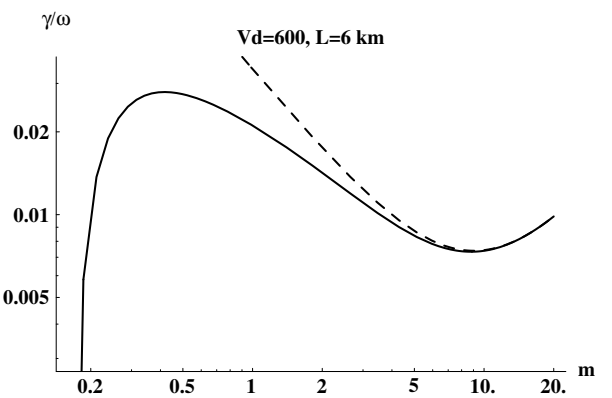

Fig. 7. Same as Fig. (6) but for 600 m/s electron drift.

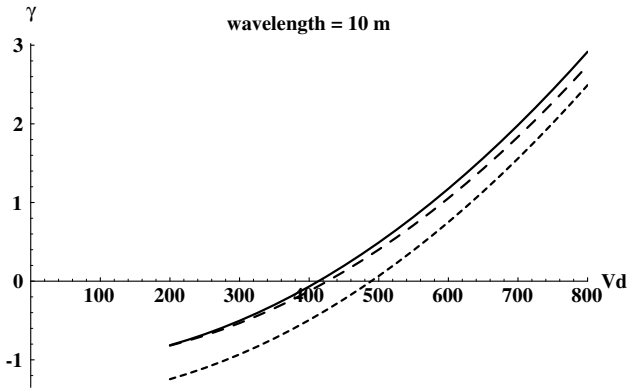

Fig. 9. Hybrid theory growth rate as a function of the electron drift for three different values of Ψ: solid line has $\Psi = \infty$ (isothermal), dashed line has $\Psi = 30$, and dotted line has $\Psi = 1$ (adiabatic). The wavelength is 10 m.

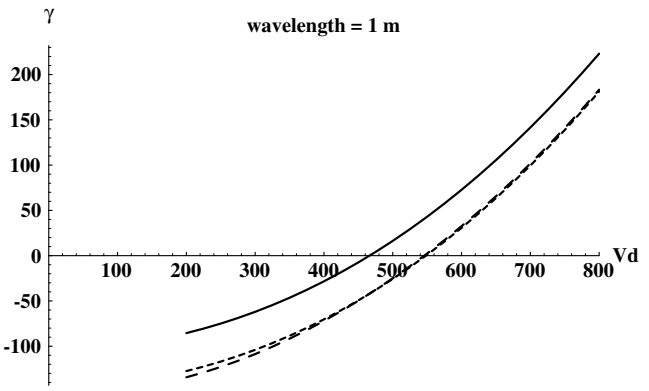

Fig. 10. Same as Fig. (9) but for one meter wavelength.

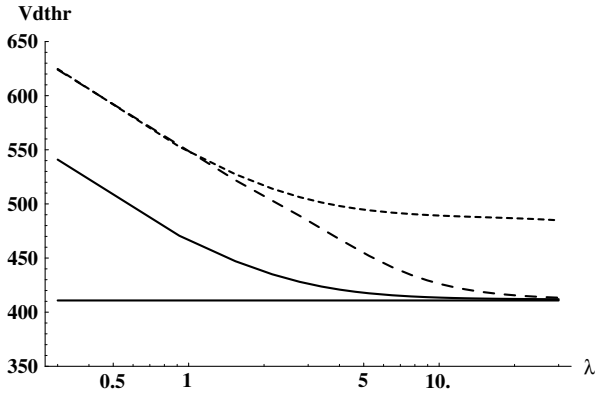

Fig. 11. The threshold electron drift (m/s) as a function of wavelength (m) in hybrid theory: solid line has $\Psi = \infty$ (isothermal electrons), dashed line has $\Psi = 30$ and dotted line has $\Psi = 1$ (adiabatic electrons). In addition, the straight horizontal line marks the fluid theory result. Notice logarithmic scale in horizontal axis.

ative growth rate to appear at half a meter wavelength for the hybrid relation.

Fig. (8) shows the hybrid theory relative growth rate as a function of the aspect angle α and the wavenumber k. The growth rate is an even function of the aspect angle. It is clearly seen that small-wavelength waves are more field-aligned than long-wavelength waves according to linear theory. For this k range, the maximum growth rate $\max_\alpha(\gamma/\omega_r)$ increases with wavenumber k, but there are points where γ/ω_r decreases with k, if the aspect angle α is kept fixed.

In all hybrid relation figures thus far, we have assumed that the electrons are isothermal by taking the limit $\Psi \to \infty$. On the other hand, the choice $\Psi = 1$ would correspond to adiabatic electrons. Fig. (9) displays the hybrid theory growth rate γ as a function of the electron drift for three values of Ψ: the solid curve corresponds to $\Psi = \infty$, the dashed curve has $\Psi = 30$, and the dotted curve means adiabatic electrons with $\Psi = 1$. The difference in threshold between isothermal and adiabatic electrons is about 20 %, and the case $\Psi = 30$ seems to be close to isothermal. We can compare this result with simple fluid theory estimation (e.g., Farley and Providakes, 1989): For adiabatic electrons, the ion acoustic speed squared is $C_s^2 = k_B((5/3)T_e + T_i)/m_i$. For isothermal electrons, the corresponding expression is, of course, $C_s^2 = k_B(T_e + T_i)/m_i$. In our case the temperature ratio was assumed 1.52, and this yields $C_s^{(\mathrm{adiab.})}/C_s^{(\mathrm{isoth.})} = 1.18$, in agreement with Fig. (9). Fig. (10) is similar to Fig. (9), but now the wavelength is one meter. A qualitative change is that now $\Psi = 30$ is no longer nearly isothermal, but nearly adiabatic.

The threshold drift velocity V_d^{thr} as a function of wavelength is plotted in Fig. (11) for several values of Ψ. The fluid theory equivalent (straight solid horizontal line) is also shown. At large wavelengths the fluid theory and isothermal hybrid theory agree (solid lines), but for smaller wavelengths the hybrid theory threshold is clearly larger. At one meter wavelength the difference is already about 100 m/s. The adiabatic hybrid theory (dotted line) curve is for all wavelengths at about 100 m/s higher than the corresponding isothermal curve. Probably the most realistic curve is the dashed one, which has $\Psi = 30$. For long wavelengths it agrees with the fluid and isothemal hybrid theories, but it rises steeper than the isothermal hybrid curve for smaller wavelengths. The conclusion is that the usual isothermal kinetic theory probably underestimates the threshold for small wavelengths. However, the practical significance of this is not clear; it may be that nonlinear effects are in reality can mask this temperature fluctuation effect.

In summary: our hybrid dispersion relation given by (36), (23) and (32) is valid when $\nu_{in} \gg \Omega_i$, and $\lambda \gg \max(\lambda_{De}, a_e)$ where a_e is the electron Larmor radius. These conditions are usually met for radar wavelengths in the lower and middle E–region. In the higher E–region the ion magnetization should be included, but this would yield to more complicated expressions which are also much less convenient to handle numerically.

3 Phenomenological characteristics

3.1 Radar observations

There are excellent reviews of the observed properties of the E–region waves, both in equatorial (Farley, 1985; Fejer and Kelley, 1980) and auroral (Haldoupis, 1989)

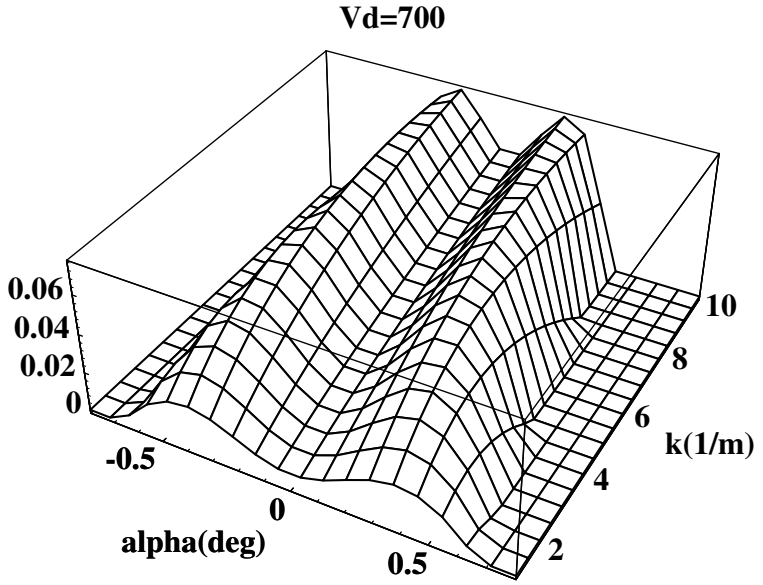

Fig. 8. Hybrid theory relative growth rate γ/ω_r as a function of aspect angle α and wavenumber k.

electrojets. Therefore we just briefly mention those properties which are in our view the most important.

In the equatorial electrojet, the effect of the electron density gradient obeys the general picture obtainable from linear fluid theory. A gradient with $\mathbf{E} \cdot \nabla n > 0$ produces broad spectra (called type 2) even if the electric field is below the threshold. On the other hand, a gradient with $\mathbf{E} \cdot \nabla n < 0$ can shut off the instability. The intensity of type 2 spectra at 50 MHz is roughly proportional to V_d^2, this is in agreement with fluid simulations (McDonald et al., 1974; Ferch and Sudan, 1977; Keskinen et al., 1979).

The existence of a threshold is also clearly seen in the data. Whenever the drift is large enough, narrow and intense spectra (called type 1) suddenly appear. This corresponds to the "opening up" of the instability droplet (Fig. 3). Type 1 spectra can usually be seen also outside the linearly unstable sector, even for flow angle 90°, as soon as the drift exceeds the threshold.

The most peculiar experimental finding in the equatorial spectra is that the phase velocity of type 1 waves appears to be limited by the ion acoustic velocity C_s (e.g., Farley and Balsley, 1973). This is in contrast with linear theory which states that the phase velocity should be proportional to the electric field also above the threshold. This phenomenon is called C_s saturation in this work.

In the auroral region, the spectra are more variable, and the C_s saturation is more obscure, although it seems that peaks close to C_s appear more frequently than they should on purely statistical grounds (e.g., Haldoupis et al., 1993). Despite large amounts of auroral coherent scatter data, surprisingly little is known about the geophysical contexts of the spectra.

Fig. (12) is a sketch of the spectral types 1, 2, 3 and 4. Type 1 waves are generally believed to be the 'normal' Farley–Buneman waves in both equatorial and auroral regions. Ion cyclotron waves were first considered as a possible explanation for type 3 spectra (Fejer et al., 1984a), but it has later been shown by interferometric measurements that this explanation is incorrect (Sahr, 1990; Sahr et al., 1991). Thus the type 3 problem is open again (e.g., Haldoupis et al., 1992). The origin of type 4 spectra is also unknown; elevated electron temperature is one possible mechanism (Fejer et al., 1986; Farley and Providakes, 1989).

Electron temperature measurements of the E–region using incoherent scatter radars led to the discovery of anomalous heating (Schlegel and St.-Maurice, 1981). Theories trying to explain the observed high temperatures were thereafter built (e.g., St.-Maurice et al., 1981; StMaurice and Laher, 1985), based on the idea that the Farley–Buneman waves could nonlinearly heat the electrons. It has, however, also been suggested that the classical frictional heating might be enough and no anomalous heating mechanism is required to explain the observations (D'Angelo and Merlino, 1987).

3.2 Rocket observations

Sounding rockets provide for *in situ* measurements of the E–region ionosphere. Although rocket campaigns are expensive and the data produced are spatially and temporarily limited, they are essential in confirming and calibrating radar measurements and theories.

The early studies (e.g., Kelley and Mozer, 1973; Pfaff et al., 1984) confirmed that the Farley–Buneman waves exist and that their basic properties such as altitude distribution, electron density turbulence level and phase velocity are roughly in agreement with theoretical estimates. The most frequently quoted turbulence levels ($\Delta n/n$) have been about 2–3 %, and less than 10 % in almost all cases. The variability may be partly due to instrumentation effects.

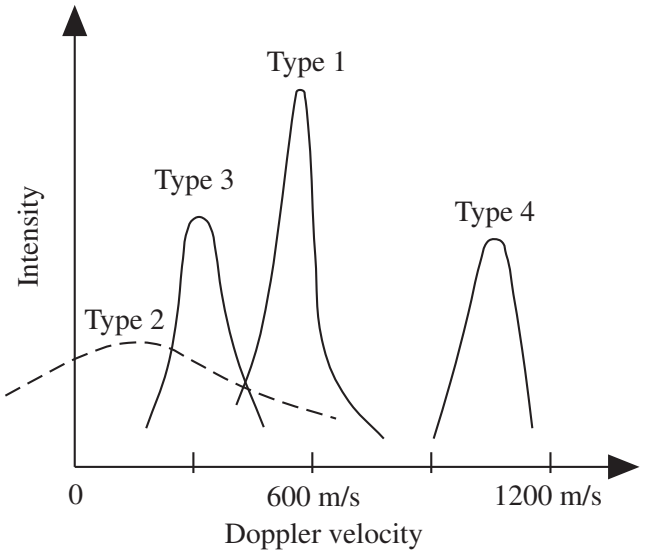

Fig. 12. Schematic representation of the radar spectral types 1, 2, 3 and 4. Type 1 is the most common, and its phase velocity is close to C_s. Type 4 phase velocity is about twice as large as for type 1. Type 3 phase velocity is close to 250 m/s, or half the type 1 Doppler shift. Type 2 spectra have small average Doppler velocity and they are wide. (Adapted from Fig. 7 of Haldoupis et al. (1993)).

Some of the later measurements have been able to dwell into the more difficult questions of phase velocities and angular distributions, which are the key to testing the various theories. A work by Pfaff et al.(1987) indicated that the phase velocity does not saturate to C_s but rather is proportional to the electron drift as the linear theory suggests. We shall return to this interesting finding later in the paper. In some studies (Schlegel, 1992), a measurement of the flow angle distribution of the intensity spectrum has been attempted.

3.3 Laboratory investigations

Farley–Buneman type instabilities have been studied in laboratory conditions using cylindrical plasma columns with radial driving electrostatic fields (e.g., D'Angelo et al., 1974). The plasma $\mathbf{E}\times\mathbf{B}$ rotates in the cylinder. The most severe limitation of these studies is their non-cartesian geometry.

The early experiments (D'Angelo et al., 1974) confirmed the basic linear theory characteristics: the existence of a threshold and the overall $\mathbf{E}\times\mathbf{B}$ directed motion of plasma structures. Kustom et al.(1985) found a steeply falling $k^{-3.5}$ spatial power spectrum and concluded that there is far too little power at short wavelengths (10–30 cm in the ionosphere) to provide any anomalous wave heating. John and Saxena (1975) found the spectral exponent to be in the range 2..3.5, and the phase velocity to be between C_s and the linear theory prediction.

Despite the limited accuracy and scope of laboratory investigations, they nevertheless provide an independent approach to the Farley–Buneman problem. However, nowadays and in the future numerical simulations are likely to supersede laboratory experiments, since the former do not suffer from the cylindrical geometry restriction.

4 Simulations

4.1 Fluid simulations

A numerical solution of the two-fluid model, Eqs. (4) through (6), using finite-difference techniques does not seem too difficult at first sight. The pioneering works (McDonald et al.; 1974, Ferch and Sudan, 1977; Keskinen et al., 1979) indeed solved the full set of equations and obtained results that supported the nonlinear cascade process that Sudan et al.(1973) proposed to explain type 2 waves. The simulations gave rise to theoretical developments (Sudan and Keskinen, 1977), which completed the current understanding of gradient drift turbulence and type 2 spectra.

An example of a two-fluid model simulation and the associated cascade process is shown in Fig. (13) and Fig. (14). The system was initialized with a sinusoidal perturbation of 3 % amplitude, the electric field was 30 % of the Farley–Buneman threshold, and a density gradient was arranged such that the sinusoidal wave was weakly unstable according to Eq. (13). Fig. (13) displays the electron density at time t_1, which corresponds to five wave travel times through the simulation box, which is equipped with periodic boundary conditions. The electric field is upward and the waves travel from left to right. The simulation box size is 100 m $\times$ 100 m. At time t_1, the initially sinusoidal perturbation has started to erode, and fingerlike structures characteristic to gradient drift instability have started to develop. Fig. (14) is at time t_2, which is twice time t_1, that is, it corresponds to ten wave travel times across the box, or about 10 seconds physical time, or a few linear e-folding times of the initial wave. (It is not reasonable to define the physical quantities more accurately since the scaling from code units to physical units depends on C_s and the ion Larmor radius, for instance.) At t_2, the turbulence is already almost fully developed and nearly isotropic. Almost all traces of the initial sinusoidal wave have disappeared.

The success of fluid simulations in explaining type 2 spectra awakened optimism: maybe these simulations can be run for drift velocities approaching and exceeding the Farley–Buneman threshold. McDonald et al.(1974) wrote: "We have retained ion inertia ... so that we can later investigate the transition from type 2 instabilities to type 1 as the electron drift velocity is increased in stages until it exceeds the acoustic speed. The results of this investigation will be described in detail elsewhere." Fejer and Kelley (1980) were even more optimistic: "The inclusion of the ion inertia, and the use of large drift velocities, may explain eventually the physical processes associated with both type 1 and type 2 irregularities." If the ion inertia is neglected the term proportional to ω_r^2 disappears in Eq. (13), and thus the Farley–Buneman instability is suppressed. Indeed, Newman and Ott (1981) were able

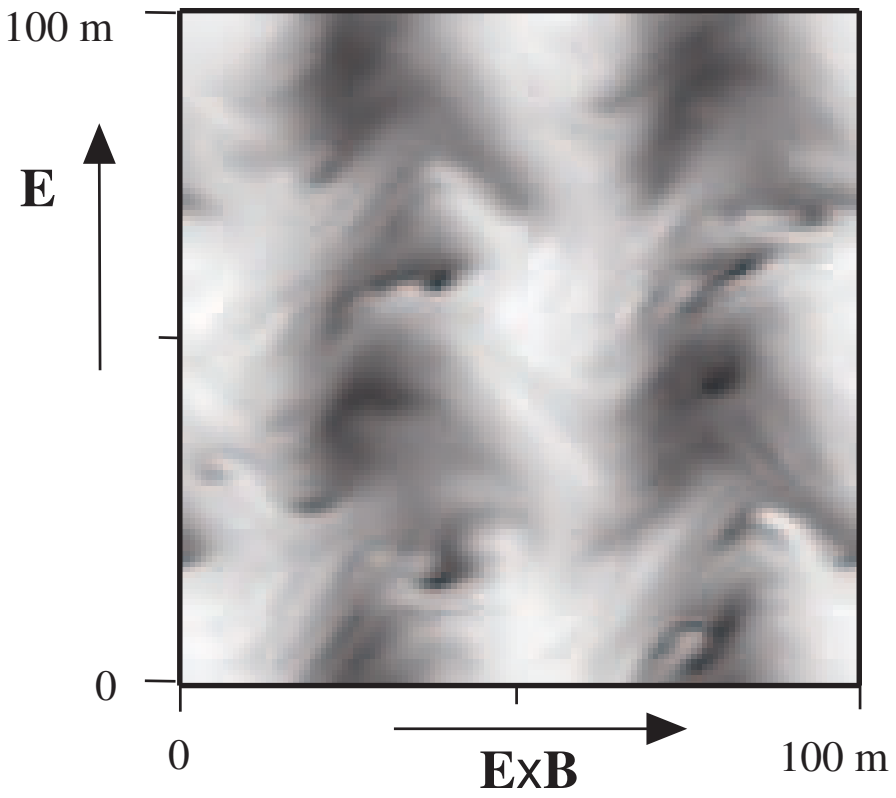

Fig. 13. 2D fluid simulation electron density after five through-the-box wave travel times. Dark and light shades correspond to low and high electron densities, respectively. The maximum n_e perturbation is about 5 %. The ion inertial term was not included in the equations.

to extend fluid simulations to electric field just exceeding the threshold by adding diffusive damping terms to suppress short wavelength modes. However, as their prescribed electric field was only 1.06 times the threshold at maximum, it was difficult to compare the phase velocities with linear theory and C_s, because both were almost equal. They performed some runs with higher electric field, but in these runs it was the current density, not the electric field that was kept constant, and consequently the interpretation of these runs was difficult.

Unfortunately, the initial optimism was not justified, and the problem of Farley–Buneman waves is far more difficult than then realized. If one tries to run the fluid codes with electric fields approaching or exceeding the threshold, numerical divergence soon results. The reason for this bad behaviour is that the linear theory growth rate increases without limit as the wavelength is decreased in the unstable sector (Fig. 3), and thus the most unstable mode will be the shortest wavelength that the spatial grid can realize. It is hopeless to try to design a stable numerical method for a problem which is so unstable by its very nature.

One might hope to cure this problem by introducing "cutoff" terms in the fluid equations, which damp the smallest scale waves. This has been attempted (Janhunen, 1992–1993, unpublished; Sahr, 1993, private communication), but it has not been too successful. The cutoffs required are quite severe, and even then the code may occassionally and unpredictably explode. Even if the time development before the onset of numerical instability can be trusted, simulation codes should be more tolerant to be really useful.

The most difficult part in fluid simulations is solving the elliptic equation (Eq. 6). The initial investigations (e.g., McDonald et al., 1974) used alternating direction implicit method (ADI, Press et al., 1992), but it is possible that more robust codes would result if the strongly implicit procedure (SIP; Jesshope, 1979) was used. For 3D problems, successive overrelaxation (SOR) has been successfully employed (Janhunen, 1992). However, even the best elliptic solver can not totally remove the convergence problems intrinsic to the fluid approximation Farley–Buneman waves.

The auroral plasma irregularities were studied in Janhunen (1992) using 3D fluid simulations. The investigation showed that the 3D code is numerically stable for larger electric fields than 2D, but the Farley–Buneman threshold could still not be exceeded. In the auroral region, the collision frequencies and thus ψ (Eq. 10) depend on altitude with about 6 km scale height. Therefore, the horizontal and perpendicular phase velocity $v_{ph} \approx V_d/(1+\psi)$ (assuming vertical magnetic field) will also depend on altitude. Theoretically, a typical parallel/perpendicular wavelength ratio is of the order of $\sqrt{m_i/m_e} \approx 240$ (this makes ψ about unity), therefore the parallel component of a wave whose perpendicular wavelength is 10 m will be about 2 km, which is compa-

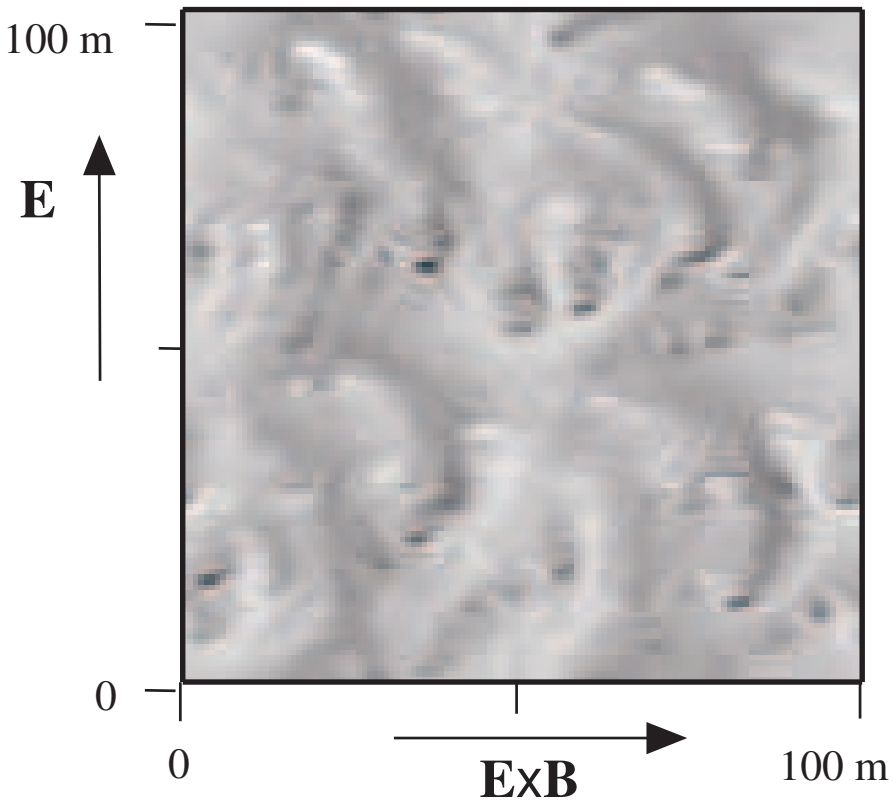

Fig. 14. Similar to Fig. (13), but after 10 wave travel times.

rable to the scale height 6 km. This will have two effects. 1) Because of altitude dependency of the phase velocity, "wavetower tilting" (Janhunen, 1992) results, which leads to nonsymmetric nonzero aspect angle propagation, which increases ψ and causes damping. 2) Because there is coupling to lower altitudes where ψ is larger by more frequent collisions, the wave is more damped than it would be in a homogeneous medium. These were studied using a 3D simulation, however, the results were not completely satisfactory because even the 3D code was not unconditionally stable. The wavetower tilting applies only to auroral latitudes, not to the equatorial electrojet.

4.2 Particle simulations

To get rid of the limitations and numerical instabilities associated with fluid simulations, particle simulations can be used.

In electrostatic particle simulations, the motion of a large (up to a million, say) number of electrons and ions is solved under the self-consistent electrostatic field (e.g., Birdsall and Langdon, 1985).

Traditionally, particle simulation techniques have been used in fully ionized plasmas, while applications to weakly ionized systems have remained rare. Therefore, we shall devote a bit more elaborate discussion on the benefits and difficulties associated with particle simulations in E–region conditions.

In the ionosphere, the neutral collisions are a fundamental part of the system. In Boltzmann equation formalism, they are described by the collision term. In a particle simulation there are more options for including them: we could add "drag" terms to the equations of motion for each particle, or we could use a Monte Carlo type method to randomly collide the particles. We have used the latter type of neutral collisions, which makes it possible and easy to also implement inelastic collisions.

In a fully ionized electrostatic plasma, the sum of plasma kinetic and potential energies is conserved. The particle simulations are initial value problems, where the distribution functions are specified at $t = 0$ and computed at later times. New energy is not usually fed into the system during the run. The results very much depend on the initial conditions, and therefore the interpretation requires great care.

When we add the neutral collisions, the plasma energy is no longer constant. Instead, both ion and electron distribution functions approach thermal equilibrium with the neutrals and themselves, if there is no fresh energy input. The final state is independent of the initial conditions, if the run is sufficiently long. If an external electric field is imposed, energy starts flowing from the electric field through the plasma component to the neutrals. The external field drives unstable plasma waves, which heat the particles. The particles are also heated directly by the external field (frictional heating). The heated plasma particles are cooled by the neutral collisions, until equilibrium is reached. The neutrals also get heated, but in a time scale very much longer than we are interested in, so they can be regarded as an infinite

energy sink with fixed temperature. The final state does not depend on the initial conditions if the run is long enough, but it is a function of the imposed electric field only (and, of course, collision frequencies and other such parameters). This is the main difference between fully ionized and weakly ionized plasma simulations, and it is a great benefit for E–region simulators.

The most pressing difficulty with E–region particle simulations is the smallness of the Debye length, as compared to the wavelengths of interest. The Debye length is usually less than one centimeter, while relevant wavelengths are one meter or larger. Therefore, very large number of particles should be used.

The first particle simulations were performed by Machida and Goertz (1988). These were 2D simulations with relatively small number of particles in the yz plane ($x \parallel \mathbf{E}$, $y \parallel \mathbf{E}\times\mathbf{B}$, $z \parallel \mathbf{B}$). Because of the geometry, only aspect angle dependencies could be studied. In the early stage of the run, the time development was in qualitative agreement with BGK kinetic theory (the fastest growing modes had nonzero aspect angle), but later the largest amplitude waves had more and more perpendicular $\mathbf{k}$ vectors.

Janhunen (1994a) presented fully perpendicular (xy) simulations with much larger number of particles. In this geometry, only flow angle dependencies can be studied. The results showed that the phase velocity was proportional to the electric field, i.e., it followed the linear theory prediction quite well. In other words, no C_s saturation was seen. Also, no anomalous wave heating was found, but the heating present was due to the classical frictional heating only. The spatial power spectrum in (k_x, k_y) plane was not symmetric as a function of the flow angle, but the peak was on $\mathbf{k}$ vectors satisfying $\mathbf{k}\cdot\mathbf{E} < 0$. The phase velocity of the most intense wave was close to C_s for above-threshold electric fields, although in any given direction, the phase velocity obeyed the fluid linear theory prediction.

The results of Janhunen (1994a) were surprising, and, consequently, several consistency checks (listed in the paper) were performed to ensure that the results were of physical and not of numerical origin. The flow angle asymmetry seemed strange, since both isothermal fluid and kinetic linear theories are manifestly symmetric with respect to the flow angle. The final test was the development of a 3D particle simulation, from which both perpendicular-2D and parallel-2D (Machida and Goertz, 1988; Schlegel, 1993, private communication) results could be reproduced.

Fig. (15) shows an example of the 3D particle simulation results. Plotted is a time series of the electron density at intervals of 1000 ω_{pe}^{-1} in two orthogonal planes. In panel a we see the random initial state. After 1000 plasma periods (panel b), small wavelength oblique waves in the yz plane have started to develop (lower panel). Meanwhile, in the xy plane (upper panel), an $\mathbf{E}\times\mathbf{B}$ propagating mode fundamental to the simulation box has developed. Thus far the time evolution has more or less obeyed the linear theory, and the fundamental mode in the xy plane has supersonic phase velocity. In panel c at $t = 2000\omega_{pe}^{-1}$ we see that the oblique waves in the yz plane have inversely cascaded toward longer wavelengths, and in the xy plane the most intense wave is now oblique. This kind of asymmetry in the flow angle is characteristic to these simulations, and the phase velocity of the oblique mode is no longer supersonic but is close to C_s (transonic). Panel d at $t = 3000\omega_{pe}^{-1}$ is very similar to panel c, so the system appears to have reached a quasi-steady state. Notice that we used an artificial mass ratio of $m_i/m_e = 100$ in the simulation, and therefore 3000 electron plasma periods is not as short a time as it may seem at first. Actually 3000 plasma periods corresponds to 30 ion gyroperiods.

The smallness of the simulation box in the xy plane is a problem especially in the 3D simulation. The most intense waves would probably be too large to fit in the simulation box, and therefore the fundamental modes are dominating the spatial spectrum. Things like tendencies for oblique aspect and flow angle propagation can nevertheless be identified, if not quantitatively measured from these simulations.

5 Nonlinear theoretical models

As pointed out above, strictly speaking the linear theory looses its meaning whenever there is an unstable region of $\mathbf{k}$ space. Physically we expect the system to reach a quasi-steady nonlinearly saturated state, where the time or ensemble averaged nonlinear growth rate of all modes is zero or negative (e.g., Primdahl and Bahnsen, 1985). The requirement of a non-positive growth rate can be regarded as a "definition" of quasi-steadiness. The term "nonlinear growth rate" here must be understood phenomenologically: choose a spatial Fourier mode, plot its amplitude as a function of time, and fit a function of the form $\exp(\gamma t)$ to it. Notice that this is *a priori* different from the linear theory growth rate γ, Eq. (13). However, to obtain tractable expressions, it is usual practice to equate the nonlinear growth rate with the linear one by using some reinterpretation of the variables appearing in it. This procedure is not unique, however, and different schemes lead to different nonlinear theories. Which one, if any, is correct can be tested only by experiments and numerical simulations.

5.1 Sudan-Robinson theory

Sudan (1983a and b) proposed that the electron-neutral collision frequency is effectively enhanced by the presence of turbulent small-scale waves. That is, $\nu_e^{\rm eff} = \nu_e + \nu_e^*$, where the anomalous collision frequency ν_e^* is interpreted as emerging from electron-wave collisions. In the perpendicular plane, $\psi = \nu_e\nu_i/\Omega_e\Omega_i$ is thus replaced by $\psi^{\rm eff} = \nu_e^{\rm eff}\nu_i/\Omega_e\Omega_i$. The fluid theory dispersion relation becomes (assuming $L = \infty$)

$$\omega_r = \frac{kV_d\cos\theta}{1+\psi^{\rm eff}} \tag{38}$$

$$\gamma = \frac{\psi^{\rm eff}}{1+\psi^{\rm eff}}\frac{1}{\nu_i}\left[\left(\frac{kV_d\cos\theta}{1+\psi^{\rm eff}}\right)^2 - \mathbf{k}^2C_s^2\right]. \tag{39}$$

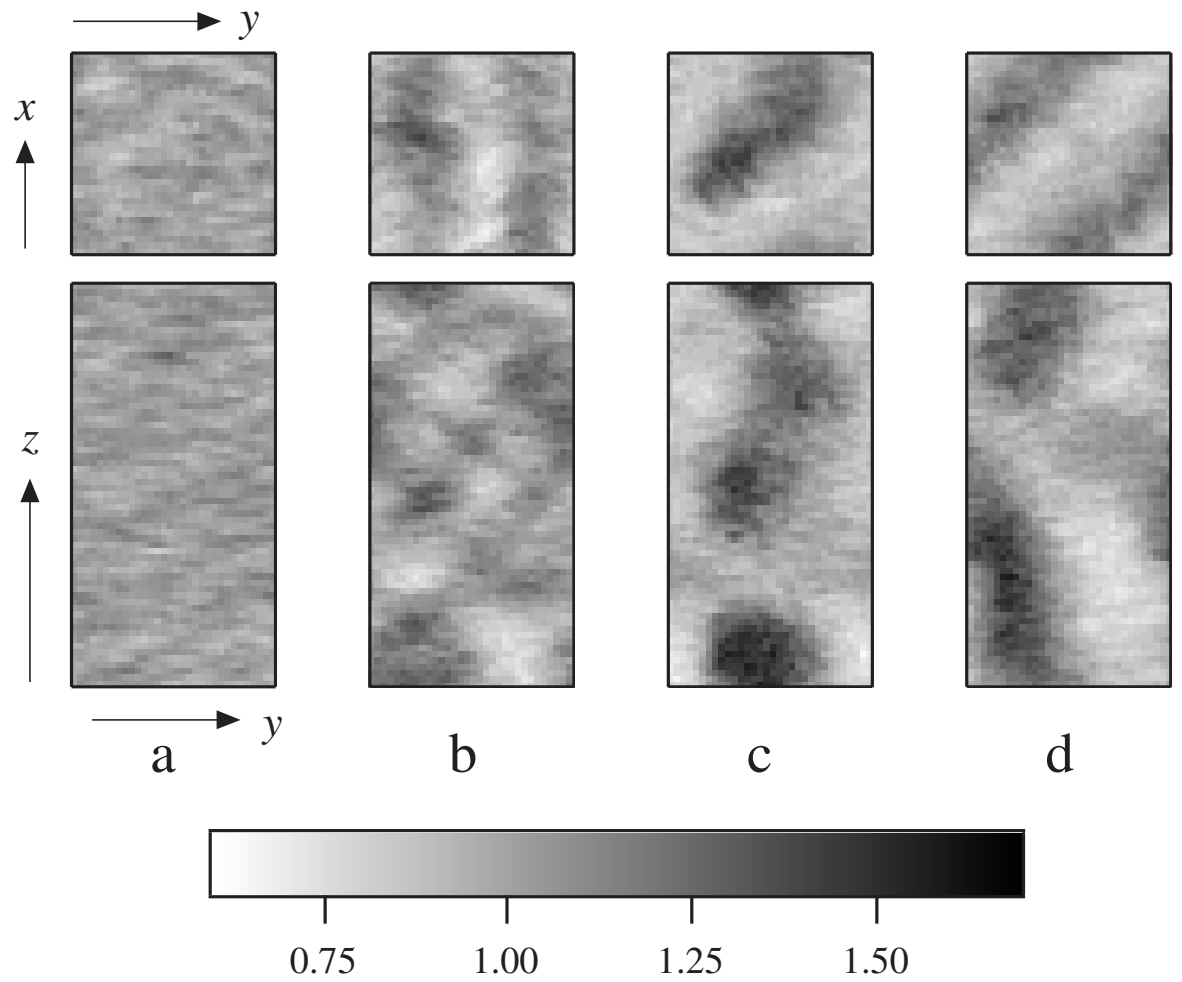

Fig. 15. 3D particle simulation electron density. Columnar panels a, b, c and d display the electron density in the xy (top) and yz planes (bottom) at times 0, 1000, 2000 and 3000 electron plasma periods, respectively. Time averaging has been applied to smooth the panels somewhat. In the top panels, the x coordinate (the electric field direction) is upwards and the y coordinate (the $\mathbf{E}\times\mathbf{B}$ direction) points from left to right. In the bottom elongated panels, the y coordinate again is directed from left to right and the z coordinate is upwards. The magnetic field is in the negative z direction, or downwards in the bottom panels. Since the $\mathbf{E}\times\mathbf{B}$ direction is to the right, the structures also move from left to right, constantly reappearing because of periodic boundary conditions. No averaging with respect to the third coordinate has been performed.

This corrected γ is then interpreted as the true nonlinear growth rate, although it is formally similar to linear theory. The quasi-steadiness condition $\gamma = 0$ then implies

$$V_d \cos\theta = C_s(1 + \psi + \psi^*). \tag{40}$$

If this equation is understood as defining ψ^*, that is

$$\psi^* = \frac{V_d}{C_s}\cos\theta - 1 - \psi, \tag{41}$$

then we see that ψ^*, and thus ν_e^*, will explicitly depend on the flow angle. [Notice that in formulas like this, it is to be understood that ψ^* is never negative: in fact we should write $\psi^* = \max((V_d/C_s)\cos\theta - 1 - \psi, 0)$].

Sudan's theory predicts the following "C_s-limited cosine law" for the phase velocity:

$$v_{ph} = \frac{\omega_r}{k} = \min\left[\frac{V_d\cos\theta}{1+\psi}, C_s\right]. \tag{42}$$

Other predictions are harder to make. Many authors have tried to relate ψ^*, or ν_e^*/ν_e, to the turbulence spectrum, but mutual agreement has not been particularly good.

Robinson (1986) generalized Sudan's mechanism to include self-consistent heating. The mechanism was also applied to the BGK kinetic linear dispersion relation (Robinson and Honary, 1990). The results of Robinson's heating calculations were in good agreement with incoherent scatter radar observations (Williams et al., 1992).

5.2 Aspect angle stabilization

To generalize or modify Sudan's theory, we notice that in order to obtain the experimental result $v_{ph} \approx C_s$ we only need the condition $\gamma = 0$, not necessarily anomalous collisions. The full expression for ψ (Eq. 10) contains the aspect angle. If α were nonzero, then we would have

enhanced ψ also (Kudeki and Farley, 1989). In analogy with Sudan's theory, this would yield

$$\sin^2\alpha = \frac{\nu_e}{\nu_i}\frac{m_e}{m_i}\left(\frac{V_d}{C_s}\cos\theta - 1 - \frac{\nu_e\nu_i}{\Omega_e\Omega_i}\right). \quad (43)$$

The interpretation of this expression is that the most intense waves have nonzero aspect angle and their phase velocity is equal to C_s, if the electric field above the threshold. The radar will see mainly these most intense modes, not the weaker ones propagating exactly perpendicularly, and will consequently report $v_{ph} \approx C_s$ (e.g., Ossakow et al., 1975).

Both aspect angle stabilization theory and Sudan-Robinson theory are formally similar in that both have enhanced ψ. By having a larger ψ, the growth rate can be turned to zero. However, it must be kept in mind that the growth rate is not a monotonically decreasing function of ψ. Instead, the growth rate will have a maximum at some value of ψ, which is larger than the classical perpendicular value but smaller than the anomalous value. Thus, too small anomalous ψ destabilizes rather than stabilizes the waves. This follows directly from Eq. (39), see also Wang and Tsunoda (1975) and Fejer et al.(1984b).

5.3 Flow angle stabilization

As mentioned above, the particle simulations carried out in Janhunen (1994a) produced surprising results: 1) the phase velocity was not saturated by C_s, 2) the flow angle of most intense waves was shifted away from drift direction such that $\mathbf{k}\cdot\mathbf{E} < 0$. These results can be understood in terms of a different kind of nonlinear theory, called flow angle stabilization (Janhunen, 1994b). It is very similar to aspect angle stabilization. In order to satisfy

$$V_d\cos\theta = C_s(1+\psi) \quad (44)$$

we do not assume an enhanced ψ, but simply that the most intense modes have a suitable flow angle θ:

$$\theta = \arccos\left[\frac{C_s}{V_d}(1+\psi)\right]. \quad (45)$$

The interpretation is exactly similar to aspect angle stabilization: the radar will see mainly waves with nonzero θ, and this will lead, in a statistical sense, to $v_{ph} \approx C_s$ whenever the mean drift is above the threshold. For this scheme to work, it must be assumed that there is enough large-scale turbulence so that there is sufficient flow angle mixing. In the equatorial electrojet this seems to be the case (Farley and Balsley, 1973). In the auroral region there is in general less gradient drift turbulence because the perpendicular gradient introduced by the height dependence of the electron density is small. Thus the flow angle stabilization predicts less dramatic C_s saturation than in the equatorial zone. This is in accordance with observations: in the auroral region the C_s saturation is seldom clear, but is usually realized only in a statistical sense.

Flow angle stabilization is conceptually similar to aspect angle stabilization. In both cases, a modification in the angular dependence of the spatial power spectrum is assumed. However, in flow angle stabilization, contrary to aspect angle stabilization, the growth rate decreases monotonically with the angular variable.

6 Current status and future directions

The gradient drift turbulence (type 2 spectra) is well understood. Both simulations (Keskinen et al., 1979) and theory (Sudan and Keskinen, 1977; see also Farley, 1985) are in good agreement with radar observations (Fejer and Kelley, 1980). The theory is not mathematically rigorous, but it makes good physical sense. If required, the simulations could nowadays be performed with good spatial resolution. The spatial power spectrum obeys a power law, and it is almost isotropic with respect to the flow angle. The remaining problem with gradient drift turbulence is how it interacts with the Farley–Buneman waves, but obviously this question can be meaningfully studied only if the Farley–Buneman instability itself is first properly understood.

The Farley–Buneman problem is not yet solved, but a solution may be emerging. The original puzzle with equatorial observations was the C_s saturation (Farley and Balsley, 1973). Throughout the 1970's, many theoretical explanations were proposed (e.g.,Fejer and Kelley, 1980), but no one was really compatible with all observations. Sudan's theory (Sudan, 1983b) explained the phenomenon in a way compatible with observations, but a rigorous mathematical foundation was still missing.

Sudan's theory gained wide popularity, and it was applied to the auroral region by several authors, including Robinson (1986). Perhaps this was a misleading step, since in the auroral region a strict C_s saturation was never really observed. Therefore, it was necessary to modify Sudan's theory to account for phase velocities larger than C_s; also for type 4 waves (Fejer et al., 1986) which have phase velocities about twice C_s. Most authors assumed elevated electron temperatures (high C_s) to explain the high phase velocities.

An alternative explanation is provided by the particle simulations of (Janhunen, 1994a) and the flow angle stabilization concept (Janhunen, 1994b). In this approach, we do not need anomalous collisions, and the existence of C_s saturation depends on the level of larger scale turbulence. The C_s saturation does not mean "real" slowing down of the waves, but only an asymmetric flow angle power distribution. Direct observational support for this new model is still rare, but it is also not incompatible with any of the earlier observations.

We list some observational indications that favor flow angle stabilization. *In situ* rocket measurements (Pfaff et al., 1987) indicated that the phase velocity followed linear theory, and no C_s saturation was seen. However, simultaneous radar observations did show C_s saturation. This is just what flow angle stabilization predicts. (Mattin and Jones, 1987) observed a nonsymmetric flow angle power distribution. Unfortunately, the direction of the electric field in their results is not clear, so we can-

not at this point conclude whether or not their results support flow angle stabilization.

The flow angle stabilization may also explain the high Doppler shift auroral spectra without need to assume very high electron temperatures. Elevated electron temperatures probably exist, but it is not likely that they can alone explain all high Doppler velocity events (Haldoupis and Schlegel, 1990). The flow angle stabilization scheme can produce high Doppler velocities directly, if the flow angle is small and if the scattering volume is not too turbulent. This is probably the primary reason for the overall difference between equatorial and auroral spectra: in the auroral region there is no significant permanent density gradient perpendicular to $\mathbf{B}$, as at the equator, and therefore there is less gradient drift turbulence. Consequently, the spectra obey more the linear theory and are less C_s saturated on the average (Janhunen, 1994b).

When there is auroral particle precipitation, the situation may change drastically since then there may be steep (scale length less than one kilometer) density gradients associated with the auroral arcs. Then there may be a lot of gradient drift turbulence, and C_s saturation is expected. On the other hand, the gradients may be large enough to appreciably raise or lower the Farley–Buneman instability threshold, which may again modify the spectra (Prikryl et al., 1992). This may lead to diverse and irregular spectra during auroral precipitation.

What we need now is experimental verification. To this end, a new combined coherent/incoherent scatter radar campaign would be ideal. The campaign should probably last at least several hundred hours in order to obtain reasonable statistics over all, or almost all, flow angle sectors, and obviously no symmetries should be assumed a priori. The numerical simulations should proceed toward more quantitative estimates of the spatial turbulence spectrum. This could be accomplished by using implicit particle codes or hybrid simulations. The theoretical estimates like those given in Janhunen (1994b) should include the density gradients and electron heating. We also should try to find a theoretical explanation for the sign of the flow angle shift: why is it that the most intense modes always tend to propagate at the $\mathbf{k} \cdot \mathbf{E} < 0$ edge of the unstable sector. Clues to this interesting challenge could be obtained by running several different kinds of numerical simulations, e.g. hybrid codes, and see when the flow angle stabilization disappears. Such work is under development now.

If this program is successful, it may open new possibilities to address the old question of coherent scatter data 'inversion', i.e. the estimation of the electric field, when the temporal power spectrum is known. If the numerical simulations can produce the spatial power spectrum accurately enough, a statistical inversion theoretical approach for the estimation of $\mathbf{E}$ becomes feasible.

References

Bhatnagar, P.L., E.P. Gross, and M. Krook, A model for collision processes in gases, 1, Small amplitude processes in charged and neutral one-component systems, Phys. Rev., 94, 511-525, 1954.

Birdsall, C.K., and A.B. Langdon, Plasma physics via computer simulation, McGraw-Hill Book Company, Singapore, 1985.

Buneman, O., Excitation of field aligned soundwaves by electron streams, Phys. Rev. Lett., 10, 285–287, 1963.

D'Angelo, N., H.L. Pécseli, P.I. Petersen, The Farley instability: A laboratory test, J. Geophys. Res., 79, 4747–4751, 1974.

D'Angelo, N., and R.L. Merlino, Comment on "Are observed broadband plasma wave amplitudes large enough to explain the enhanced electron temperatures of the high-latitude E region?" by J.-P. St.-Maurice and Russ Laher, J. Geophys. Res., 92, 321–322, 1987.

Farley, D.T., A plasma instability resulting in field-aligned irregularities in the ionosphere, J. Geophys. Res., 68, 6083–6097, 1963.

Farley, D.T., and B.B. Balsley, Instabilities in the equatorial electrojet, J. Geophys. Res., 78, 227–239, 1973.

Farley, D.T., Theory of equatorial electrojet plasma waves: new developments and current status, J. Atmos. Terr. Phys., 47, 729–744, 1985.

Farley, D.T., and J. Providakes, The variation with T_e and T_i of the velocity of unstable ionospheric two-stream waves, J. Geophys. Res., 94, 15415–15420, 1989.

Fejer, B., and M.C. Kelley, Ionospheric irregularities, Rev. Geophys. Space Phys., 18 401–454, 1980.

Fejer, B.G., R.W. Reed, D.T. Farley, W.E. Swartz, and M.C. Kelley, Ion cyclotron waves as a possible source of resonant auroral radar echoes, J. Geophys. Res., 89, 187–194, 1984a.

Fejer, B.G., J. Providakes, and D.T. Farley, Theory of plasma waves in the auroral E region, J. Geophys. Res., 89, 7487–7494, 1984b.

Fejer, B.G., J. Providakes, D.T. Farley, and W.E. Swartz, Auroral E region plasma waves and elevated electron temperatures, J. Geophys. Res., 91, 13583–13592, 1986.

Ferch, R.L., and R.N. Sudan, Numerical simulations of Type II gradient drift irregularities in the equatorial electrojet, J. Geophys. Res., 82, 2283–2288, 1977.

Fried, B.D., and S.D. Conte, The plasma dispersion function, Academic Press Inc., New York, 1961.

Haldoupis, C., A review on radio studies of auroral E-region ionospheric irregularities, Ann. Geophys., 7, 329–258, 1989.

Haldoupis, C., and K. Schlegel, Direct comparison of 1-m irregularity phase velocities and ion acoustic speeds in he auroral E-region ionosphere, J. Geophys. Res., 95, 18989–19000, 1990.

Haldoupis, C., K. Schlegel, and E. Nielsen, On type 3 auroral VHF coherent radar backscatter, J. Geophys. Res., 97, 4109–4120, 1992.

Haldoupis, C., J.A. Koehler., G.J. Sofko, D.W. Danskin, M.J. McKibben, and D. André, Preferential phase velocities for type 4 irregularities in the auroral E region plasma, J. Geophys. Res., 98, 6173–6179, 1993.

Ichimaru, S., Basic principles of plasma physics, A statistical approach, W.A. Benjamin, Inc., London, 1973.

Janhunen, P., Three-dimensional stabilization mechanism for the auroral Farley-Buneman instability, J. Atmos. Terr. Phys., 54, 1633–1643, 1992.

Janhunen, P. Numerical simulations of E-region irregularities and ionosphere-magnetosphere coupling, PhD thesis, Finnish Meteorological Institute Contributions, 13, 1994.

Janhunen, P., Perpendicular particle simulation of the E-region Farley-Buneman instability, J. Geophys. Res., in, press, 1994.

Janhunen, P., Implications of flow angle stabilization to E-region coherent spectra, J. Geophys. Res., in, press, 1994.

Jesshope, C.R., SIPSOL, a suite of subprograms for the solution of the linear equations arising from elliptic particle differential equations, Comp. Phys. Comm., 17, 383–391, 1979.

John, P.I., and Y.C. Saxena, Observation of the Farley-Buneman instability in laboratory plasma, Geophys. Res. Lett., 2, 251–254, 1975.

Kelley, M.C., and F.S. Mozer, Electric field and plasma density oscillations due to the high-frequency Hall current two-stream instability in the auroral E region, J. Geophys. Res., 78, 2214–2221, 1973.

Keskinen, M.J., R.N. Sudan, and R.L. Ferch, Temporal and spatial power spectrum studies of numerical simulations of Type II gradient drift irregularities in the equatorial electrojet, J. Geophys. Res., 84, 1419–1430, 1979.

Krall, N.A., and A.W. Trivelpiece, Principles of plasma physics, McGraw-Hill Book Company, New York, 1973.

Kudeki, E., and D.T. Farley, Aspect sensitivity of equatorial electrojet irregularities and theoretical implications, J. Geophys. Res., 94, 426–434, 1989.

Kustom, B., N. D'Angelo, and R.L. Merlino, A laboratory investigation of the high-frequency Farley-Buneman instability, J. Geophys. Res., 90, 1698–1704, 1985.

Machida, S., and C.K. Goertz, Computer simulation of the Farley-Buneman instability and anomalous electron heating in the auroral ionosphere, J. Geophys. Res., 93, 9993–10000, 1988.

Mattin, N., and T.B. Jones, Propagation angle dependence of radar auroral E-region irregularities, J. Atmos. Terr. Phys., 49, 115–121, 1987.

McDonald, B.E., T.P. Coffey, S. Ossakow, and R.N. Sudan, Preliminary report of numerical simulation of type 2 irregularities in the equatorial electrojet, J. Geophys. Res., 79, 2551–2554, 1974.

Newman, A.L., and E. Ott, Nonlinear simulations of type 1 irregularities in the equatorial electrojet, J. Geophys. Res., 86, 6897–6891, 1981.

Ossakow, S.L., K. Papadopoulos, J. Orens, and T. Coffey, Parallel propagation effects on the type 1 electrojet instabilities, J. Geophys. Res., 80, 141–148, 1975.

Pfaff, R.F., M.C. Kelley, B.G. Fejer, E. Kudeki, C.W. Carlson, A. Pedersen, and B. Hausler, Electric field and plasma density measurements in the auroral electrojet, J. Geophys. Res., 89, 236–244, 1984.

Pfaff, R.F., M.C. Kelley, E. Kudeki, B.G. Fejer, and K.D. Baker, Electric field and plasma density measurements in the strongly driven daytime equatorial electrojet 2. Two-stream waves, J. Geophys. Res., 92, 13597–13612, 1987.

Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, Numerical Recipes in C: The art of scientific computing, 2nd edition, Cambridge, 1992.

Prikryl, P., D. André, J.A. Koehler, and G.J. Sofko, Type 4 VHF radio aurora observations during low to moderate geomagnetic activity, Planet. Space Sci., 40, 1371–1389, 1992.

Primdahl, F., and A. Bahnsen, Auroral E-region diagnosis by means of nonlinearly stabilized plasma waves, Ann. Geophys., 3, 57–62, 1985.

Robinson, T.R., Towards a self-consistent non-linear theory of radar auroral backscatter, J. Atmos. Terr. Phys., 48, 417–422, 1986.

Robinson, T.R., and F. Honary, A resonance broadening kinetic theory of the modified two-stream instability: implications for radar auroral backscatter experiments, J. Geophys. Res., 95, 1073–1085, 1990.

Robinson, T.R., The effects of the resonance broadening of Farley-Buneman waves on electron dynamics and heating in the auroral E-region, J. Atmos. Terr. Phys., 54, 749–757, 1992.

Sahr, J.D., Observation and theory of the radar aurora, Ph. D. thesis, Cornell University, New York, 1990.

Sahr, J.D., D.T. Farley, and W.E. Swartz, The altitude of type 3 auroral irregularities: radar interferometer observations and implications, J. Geophys. Res., 96, 17805–17811, 1991.

Schmidt, M.J., and S.P. Gary, Density gradients and the Farley-Buneman instability, J. Geophys. Res., 78, 8261–8265, 1973.

Schlegel, K., and J.-P. St.-Maurice, Anomalous heating of the polar E region by unstable plasma waves, 1, Observations, J. Geophys. Res., 86, 1447–1452, 1981.

Schlegel, K., and J.-P. St-Maurice, Note on the parallel propagation effects of unstable Farley-Buneman waves at high latitudes, Planet. Space Sci., 30, 315–318, 1982.

Schlegel, K., Interpretation of auroral radar experiments using a kinetic theory of the two-stream instability, Radio Sci., 18, 108–118, 1983.

Schlegel, K., Measurement of electron density fluctuations during the ROSE rocket flights, J. Atmos. Terr. Phys., 54, 715–723, 1992.

Schunk, R.W., Mathematical structure of transport equations for multispecies flows, Rev. Geophys. Space Phys., 15 429–445, 1977.

St.-Maurice, J.-P., K. Schlegel, and P.M. Banks, Anomalous heating of the polar E region by unstable plasma waves 2. Theory, J. Geophys. Res., 86, 1453–1462, 1981.

St.-Maurice, J.-P., and R. Laher, Are observed broadband plasma wave amplitudes large enough to explain the enhanced temperatures in the high-latitude E region?, J. Geophys. Res., 90, 2843–2850, 1985.

Stubbe, P., Theory of electrostatic waves in an E region plasma 1. General formulation, J. Geophys. Res., 94, 5303–5315, 1989.

Sudan, R.N., J. Akinrimisi, D.T. Farley, Generation of small-scale irregularities in he equatorial electrojet, J. Geophys. Res., 78, 240–248, 1973.

Sudan, R.N., and M. Keskinen, Theory of strongly turbulent two-dimensional convection of low-pressure plasma, Phys. Rev. Lett., 38, 966–970, 1977.

Sudan, R.N., Unified theory of type I and type II irregularities in the equatorial electrojet, J. Geophys. Res., 88, 4853–4860, 1983a.

Sudan, R.N., Nonlinear theory of type I irregularities in the equatorial electrojet, Geophys. Res. Lett., 10, 983–986, 1983b.

Wang, T.N.C., and R.T. Tsunoda, On a crossed field two-stream plasma instability in the auroral plasma, J. Geophys. Res., 80, 2172–2182, 1975.

Williams, P.J.S., B. Jones, and G.O.L. Jones, The measured relationship between electric field strength and electron temperature in the auroral E-region, J. Atmos. Terr. Phys., 54, 741–748, 1992.

This article was processed by the author using the LaTeX style file *cljour2* from Springer-Verlag.