JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 104, NO. A4, PAGES 6843–6857, APRIL 1, 1999

# Large-amplitude transient growth in the linear evolution of equatorial spread *F* with a sheared zonal flow

J. P. Flaherty and C. E. Seyler

School of Electrical Engineering, Cornell University, Ithaca, New York

L. N. Trefethen

Oxford University Computing Laboratory, Oxford, England

**Abstract.** The occurrence of equatorial spread *F* on the bottomside of the *F* layer is likely the result of a process often referred to as the collisional interchange instability or generalized Rayleigh-Taylor instability. The traditional approach to the analysis of this instability with sheared zonal flow has been to calculate the eigenvalues of the linearized system. It is well documented that the introduction of shear has a stabilizing effect on the eigenvalues and significantly increases the wavelength corresponding to the fastest growing eigenmode. In this paper it is argued that the well-accepted conclusions drawn from eigenvalue analyses are not correct for cases of geophysical interest. A calculation of the $\epsilon$ pseudospectra demonstrates that the system is highly nonnormal and that large-amplitude transients may exist even when all of the eigenmodes are decaying with time. Transient effects are shown to be of fundamental importance to the evolution of the system over a wide range of horizontal wavelengths. This viewpoint is consistent with some aspects of VHF radar observations of *F* region gravitational interchange dynamics and may explain the presence of irregularities which appear to be confined to the bottomside of the *F* layer. The techniques discussed are of general interest to the analysis of nonnormal linear systems.

## 1. Introduction

In the post sunset hours the equatorial ionosphere is an unstable configuration. Gravity together with an eastward ambient electric field in the presence of the density gradient on the bottomside of the *F* layer combine to drive a process known as the generalized Rayleigh-Taylor instability or sometimes the collisional interchange instability. An analogous process exists in the electrojet and barium releases where it is referred to as the gradient drift instability.

It is often the case that instabilities are present in the bottomside of the equatorial *F* region for long periods of time without erupting into the dramatic plume-type structures that have been the focus of a great majority of the spread *F* research [*Woodman and La Hoz*, 1976]. There is no published model for spread *F* which predicts that irregularities develop and then are confined to the bottomside of the layer. It is well known that there is significant shear in the horizontal flow on the bottomside of the *F* region and that this shear is stabilizing. The question which motivates the work presented here is the following: Can the presence of a shear in the flow explain the long-lived irregularities which are observed on the bottomside of the *F* layer after sunset?

Measurements made at the Jicamarca Radio Observatory, located outside Lima, Peru, have provided evidence of significant shear in equatorial *F* region irregularities on the bottomside of the *F* region [*Kudeki et al.*, 1981; *Kelley et al.*, 1986, D. L. Hysell, private communication, 1997]. Shear has also been identified in radar images made with the ALTAIR (air launch ballistic targets demostration) radar [Tsunoda et al., 1981]. While mathematical models for the instability without shear have been studied extensively [*Kelley*, 1989], less has been done on the problem when shear is included. Most of the work on the gravitational interchange instability with shear has focused on calculating the eigenvalues of the linearized system. Eigenvalue analyses may conclude that all of the eigenmodes of the system are decaying with time. This fact does not preclude the possibility of large-amplitude transient growth if the system is far from normal (i.e., the eigenvectors are far from



Paper number 1998JA900178.
0148-0227/99/1998JA900178$09.00

orthogonal). We show that this system is highly nonnormal and that the least stable eigenvalue does not characterize the development of the instability the way it would for a normal system. Consequently, conclusions about the development of this instability based on eigenvalue analysis are significantly in error for a wide range of horizontal wavenumbers.

The effect of the shear flow on this instability was first discussed in the context of studying the evolution of striations in barium releases by *Perkins and Doles*, [1975]. These authors showed that the shear has a stabilizing effect on the eigenvalues of the system. *Huba et al.* [1983], *Guzdar et al.* [1982], and *Satyanarayana et al.* [1984] confirmed these results and calculated the eigenvalues for a wide range of conditions. *Fu et al.* [1986] approached the system of equations as an initial value problem.

They found by numerically integrating the linearized system that transients developed in a manner that was not predicted by an eigenvalue analysis and that transient growth could exist at wavenumbers at which none of the eigenmodes are growing. Their work focused on the regime where the horizontal wavelength is much shorter than the gradient length scale. Additional work of interest on ionospheric interchange instability which included shear was performed by *Huang and Kelley*, [1996] and *Ronchi* [1990].

In this paper we consider a linearized model for the development of equatorial spread $F$ with sheared horizontal flow. We show that for this system, the transient response is not limited to horizontal wavelengths which are much less than the gradient scale length. We quantify the transient response with the use of analytical techniques which have not been previously applied to this problem. The results are compared with the long-time asymptotic behavior of the system, which is equivalent to an eigenvalue analysis. We present a qualitative discussion of the process of stretching and diffusion which is the mechanism by which shear reduces the growth of the instability. We also argue that the qualitative behavior described by the numerical calculations is consistent with features of the VHF radar observations that have not been explained previously.

In section 2 the model is derived and the choices for the shear and density profiles are discussed. In section 3 the results of numerically integrating the system are presented. These results demonstrate that large-amplitude transient growth is the typical response of the system for random initial conditions. This section shows how the transient response dominates the dynamics of the system over a wide range of wavenumbers.

In section 4 the concept of the $\epsilon$ pseudospectrum of the matrix representation of the system is used to provide insight into the behavior of the system. By calculating $\epsilon$ pseudospectra, we show that the system is highly non normal (i.e., the eigenfunctions are highly nonorthogonal). The concept of pseudoresonance and the relation between the large-amplitude transient growth and the nonorthogonality of the eigenfunctions is discussed. The discussion reconciles how the system can grow at a rate faster than any of the normal modes. Further interpretation of results and properties of pseudospectra is given with references which discuss these ideas in detail.

In section 5 we quantify the transient response of the system by calculating the maximum possible growth. First, the maximum possible initial growth rate is calculated as a function of wavenumber and contrasted with the eigenvalues of the system. Then, the maximum possible transient amplification is calculated at several times of interest. This calculation shows that the wavelength at which the transient response is the largest increases as a function of time.

In section 6 we present a qualitative, physically motivated description of how shear acts to reduce the growth rate of the instability. Although the observations presented in this section are similar to those first presented by *Fu et al.* [1986] and later in *Ronchi* [1990], this section is included because it is of fundamental importance to having a physical understanding of the problem and because it illustrates how the shear stabilizes the system in a way that has not been presented previously. We conclude by arguing that many features of the numerical modeling are qualitatively consistent with the behavior of bottomside spread $F$ seen in VHF radar measurements.

## 2. Model

We first derive the system of equations and then discuss the density and shear profiles as well as choices for the parameters of the model. This is followed by a brief discussion of considerations in the numerical calculations.

The coordinate system is defined with $\hat{y}$ in the direction of the magnetic field, $\hat{z}$ is vertically upward (in the direction of the density gradient), and $\hat{x}$ is eastward (which will be referred to as the horizontal or zonal direction). The model is two-dimensional and represents the dynamics of the system in the plane perpendicular to the magnetic field. The derivation of the model is essentially the same in all previously cited work. The system is a set of two coupled partial differential equations. The first describes the time evolution of the density perturbations, and the second relates the perturbation electric field to the density. Quasi-neutrality is assumed so that $\nabla \cdot \mathrm{J} = 0$ is satisfied. The evolution of the density perturbations is determined by the electron continuity equation. The electric field is related to the density perturbations by $\nabla \cdot \mathrm{J} = 0$. If the shear is zero, the equations used here are identical to those presented by *Zargham and Seyler* [1987]. The source of the shear in the flow is the vertical component of the ambient electric field $E_z(z)$.

With the addition of this term, (11) and (12) of *Zargham and Seyler* then become

$$\frac{\partial n}{\partial t} + \hat{y} \times [\nabla\phi - \hat{z}E_z(z)] \cdot \nabla N = D_a \nabla^2 n \quad (1)$$

$$\nabla \cdot [N(\nabla\phi - \hat{z}E_z(z))] = \hat{x}E_0 \cdot \nabla N \quad (2)$$

where $N$ is the plasma density, $\phi$ is the potential associated with the perturbation electric field, $D_a$ is the ambipolar diffusion coefficient, and $E_0$ represents the combined effect of the zonal component of the ambient electric field and gravity. The electric field has been divided by the magnitude of the magnetic field $B_0$ so that it has units of velocity. Equation (1) shows that the evolution of the density perturbations can be described by two components: an advective derivative resulting from the $\mathbf{E} \times \mathbf{B}$ motion of the plasma and ambipolar diffusion. The total density $N$ can be written as

$$N = n_0(z) + n \quad (3)$$

where $n_0(z)$ is the background density which varies only in the vertical coordinate and $n$ is the perturbation density which varies in both the horizontal and vertical directions. Substituting this expression for $N$ into the above equations and linearizing by neglecting products of perturbation quantities and neglecting the diffusion of the background profile $n_0(z)$ (which is small), we get

$$\frac{\partial n}{\partial t} = E_z(z)\frac{\partial n}{\partial x} + n_0'(z)\frac{\partial \phi}{\partial z} + D_a \nabla^2 n \quad (4)$$

$$\nabla^2\phi + \frac{n_0'(z)}{n_0}\frac{\partial \phi}{\partial z} = \frac{1}{n_0(z)}\left\{\frac{\partial}{\partial z}[nE_z(z)] + E_0\frac{\partial n}{\partial x}\right\} \quad (5)$$

where $'$ denotes differentiation with respect to $z$.

The variation of the vertical electric field as a function of altitude satisfies the relation $\nabla \cdot [\hat{z}n_0(z)E_z(z)] = 0$ [*Perkins and Doles*, 1975]. This is equivalent to the statement that the product $n_0(z)E_z(z)$ is a constant. If this were not true, the zero-order terms $n_0(z)$ and $E_z(z)$ would not satisfy $\nabla \cdot \mathrm{J} = 0$. This requirement specifies the shape of the shear profile but does not specify its magnitude. After Fourier transforming in the $\hat{x}$ direction the system can be rewritten as

$$\frac{\partial n}{\partial t} = ik_x[n_0'(z)\phi + E_z(z)n] + D_a\left(\frac{\partial^2}{\partial z^2} - k_x^2\right)n \quad (6)$$

$$\left[\left(\frac{\partial^2}{\partial z^2} - k_x^2\right) + \frac{n_0'(z)}{n_0}\frac{\partial}{\partial z}\right]\phi =$$

$$\frac{1}{n_0}\left[ik_x nE_0 + nE_z'(z) + E_z(z)\frac{\partial n}{\partial z}\right] \quad (7)$$

where $k_x$ is the horizontal wavenumber.

The background density profile was modeled as a single Chapman layer on a pedestal whose height is approximately 1% of the peak density. The Chapman layer is derived by considering the balance between the production of ionization from solar radiation and the recombination with the neutral atmosphere *Budden*, [1988]. Strictly speaking, this model does not apply to the post sunset equatorial ionosphere. However, we believe that this profile is sufficiently representative for our purposes of theoretically investigating the effect of shear on the dynamics of the interchange instability. A plot of the density profile is shown in Figure 1a. Figure 1b shows the corresponding shear profile. As was described above, the shape of the shear profile is specified by the background density, and its amplitude is in general arbitrary. The magnitude of the shear shown here is consi· tent with experimental data [*Kudeki et al.*, 1981] and corresponds to a total change in the horizontal velocity of approximately 100 $\mathrm{ms}^{-1}$ across the bottomside of the $F$ layer. Figure 1c is a plot of $\nabla n_0/n_0$.

After sunset in the equatorial ionosphere, the ambient zonal field is eastward. For a period of variable duration, but typically about two hours, the eastward zonal field increases prior to reversing direction, a phe-

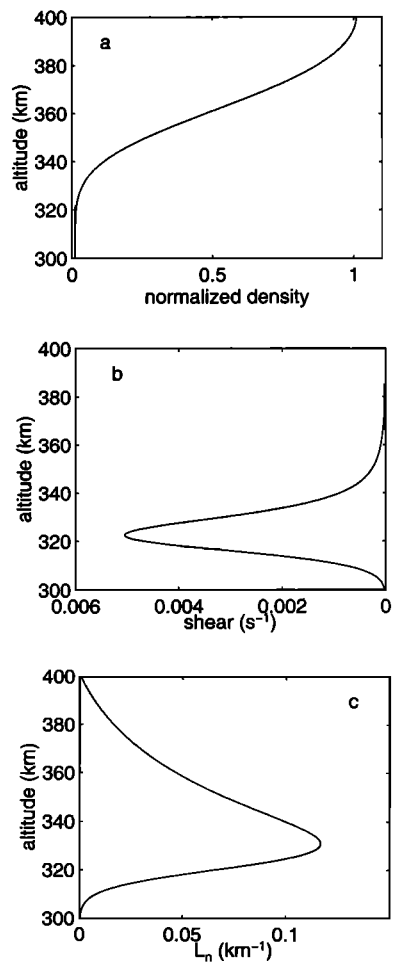


**Figure 1.** Plot of (a) the density, (b) the shear, and (c) the gradient scale length as a function of altitude.

nomenon known as the pre-reversal enhancement [*Kelley*, 1989]. So that we might better concentrate on the fundamental properties of the model, we have used a constant value of the ambient zonal field. The magnitude of the ambient field corresponds to a vertical $\mathbf{E}\times\mathbf{B}$ drift velocity of 15 $ms^{-1}$.

All of the numerical calculations discussed in this paper are based on a matrix representation of this system of the form $\dot{\mathbf{n}} = \mathbf{An}$, where $\mathbf{n}$ is an $m$ element vector and $\mathbf{A}$ is an $m \times m$ matrix. The vector n contains $m$ equally spaced samples of the perturbation density. This system was solved using standard numerical integration techniques.

The matrix representation of the system can be derived by writing (6) and (7) in the form

$$\dot{\mathbf{n}} = \tilde{\mathbf{G}}\phi + \tilde{\mathbf{C}}\mathbf{n} \tag{8}$$

$$\tilde{\mathbf{B}}\phi = \tilde{\mathbf{Q}}\mathbf{n} \tag{9}$$

where $\tilde{\mathbf{B}}, \tilde{\mathbf{Q}}, \tilde{\mathbf{G}}$ and $\tilde{\mathbf{C}}$ are linear operators. Solving for $\phi$ in (9) by inverting $\tilde{\mathbf{B}}$ and substituting into (8), the system becomes

$$\dot{\mathbf{n}} = \tilde{\mathbf{A}}\mathbf{n} \tag{10}$$

Each of these infinite-dimensional operators is approximated by a finite dimensional matrix under the assumption that the boundary conditions are periodic. Techniques for discretizing such operators are described by *Fornberg* [1996]. Discretizing each operator and forming the product $\mathbf{A} = \mathbf{GB}^{-1}\mathbf{Q} + \mathbf{C}$, where $\mathbf{G}, \mathbf{B}^{-1}, \mathbf{Q}$, and $\mathbf{C}$ are $m \times m$ matrices, produces the desired matrix representation for the system.

## 3. Integrating the System of Equations

The system of equations was solved as an initial value problem using random low $k$ perturbations as the initial conditions. In Figures 2 and 3 the magnitude of the density perturbations as a function of time is shown for three horizontal wavelengths (8, 16, and 24 km). In Figures 2a, 2b and 2c there is no shear. In Figures 2d, 2e and 2f there is mild shear. In Figures 3a, 3b and 3c the magnitude of the shear is consistent with published observations at Jicamarca by *Kudeki et al.*, [1981]. In Figures 3d, 3e and 3f the shear is increased from this value by 50% which we refer to as strong shear. In the long time limit the growth rate of the system is equal to the real part of the least stable eigenvalue. A dotted line whose slope is equal to the growth rate of the least stable eigenvalue is overlayed on each plot.

We can see from Figures 2a, 2b and 2c that when the shear is zero, the system is asymptotically unstable at horizontal wavelengths of 8, 16, and 32 km. In addition, when there is no shear the growth rate of the least stable eigenvalue characterizes the behavior of the system very well. With the possible exception of a very brief initial transient period the system grows at this rate.

When sufficient shear is introduced (see Figures 3a, 3b and 3c), the system is asymptotically stable at 8 and 16 km; however, the growth rate of the least stable eigenmode does not provide a good description of the system over the times of interest. With shear it is possible for the system to grow initially much faster than it does at later times.

It is also possible for the system to be asymptotically

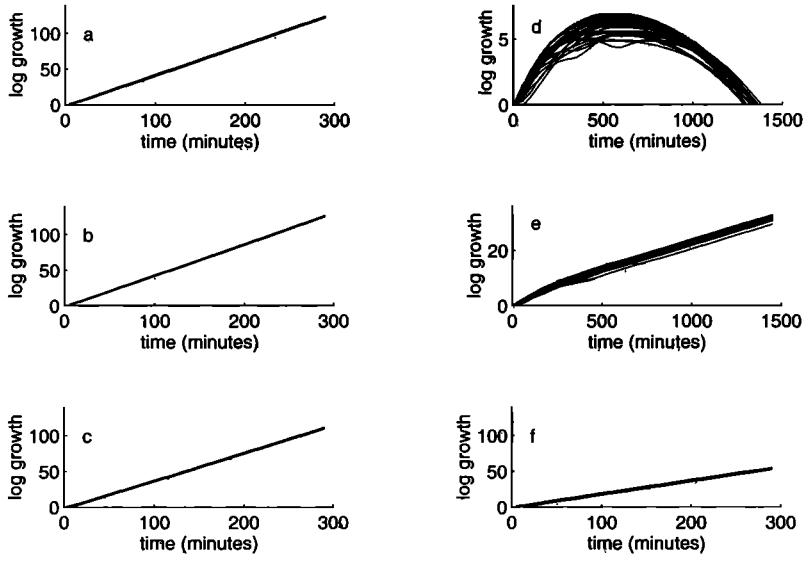


**Figure 2.** Results of numerically integrating the linearized system for (a) an 8 km wavelength without shear, (b) a 16 km wavelength without shear, (c) a 32 km wavelength without shear, (d) an 8 km wavelength with mild shear, (e) a 16 km wavelength with mild shear, and (f) a 32 km wavelength with mild shear. Growth is defined as $(\int |n|^2 dz)^{\frac{1}{2}} / (\int |n(t=0)|^2 dz)^{\frac{1}{2}}$. Each plot displays 25 trajectories. The slope of the dotted line corresponds to the growth rate of the least stable eigenvalue.

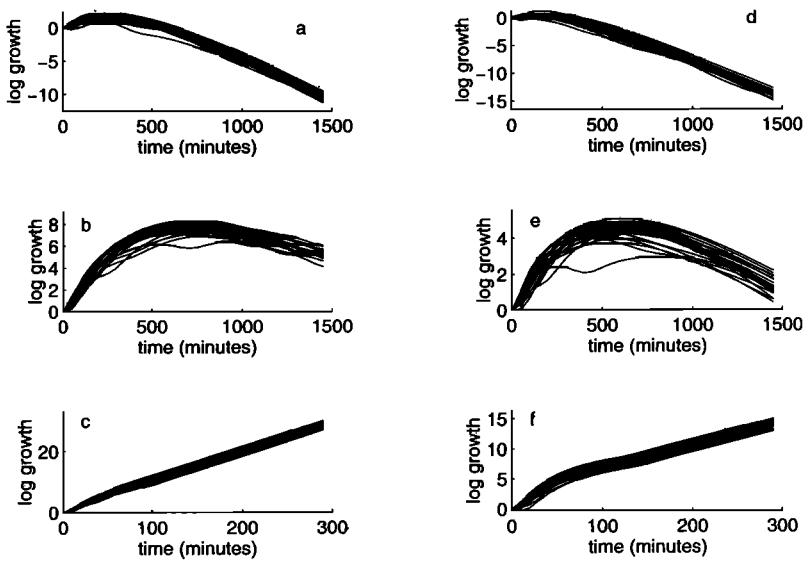


**Figure 3.** Results of numerically integrating the linearized system for (a) an 8 km wavelength with a typical shear , (b) a 16 km wavelength with a typical shear, (c) a 32 km wavelength with a typical shear, (d) an 8 km wavelength with a strong shear, (e) a 16 km wavelength with a strong shear and (f) a 32 km wavelength with a strong shear. Growth is defined as $(\int |n|^2 dz)^{\frac{1}{2}}/(\int |n(t=0)|^2 dz)^{\frac{1}{2}}$. Each plot displays 25 trajectories. The slope of the dotted line corresponds to the growth rate of the least stable eigenvalue.

stable and still exhibit significant growth. If the system is asymptotically stable, this implies that the real parts of the eigenvalues of the linear system are all less than zero: that is, all of the normal modes decay monotonically with time. This result appears to present a contradiction: if all of the normal modes are decaying with time, how can the large amplitude transient growth that is exhibited here be produced? Figure 3b, for example, shows some trajectories exhibiting as much as eight e-foldings (a factor of approximately 3000) of transient growth prior to the growth rate becoming negative and the system beginning to decay.

When the shear is mild, the growth rate at the longest wavelength shown, 32 km (Figure 2f, is close to the growth rate of the least stable eigenmode. Note that compared to the no-shear case (Figure 2c) the shear has reduced the growth rate by approximately a factor of 2. As the shear is increased (Figures 3c and 3f, the growth rate for the 32 km horizontal wavelength is further reduced and the transient response becomes prominent.

At the shortest wavelength, 8 km, even mild shear (Figure 2d) significantly changes the behavior of the system. As the shear is further increased (Figures 3a and 3d), the transient response becomes less significant: the magnitude of the transient amplification is reduced and the duration of the transient phase is decreased.

The 16 km horizontal wavelength is an intermediate case. A modest amount of shear (Figure 3b) creates a noticeable transient effect. As the shear is increased (Figure 3e) the transient response becomes more significant as in the long-wavelength case.

The transient effects are most significant when the normal modes of the system are neither rapidly growing nor strongly attenuated. When the normal modes of the system can grow rapidly, for example as in Figure 2f, the growth rate of the system approaches the growth rate of the least stable eigenmode quickly, and the transient effects are not significant. When the shear is sufficiently strong that all of the normal modes of the system are strongly damped, efor example as in Figure 3d, again the transients are not as significant. It is when the system is neither rapidly growing nor rapidly decaying that the nonnormality of the system and the associated transient amplification are most important.

These results clearly show that the behavior of the system is more complicated than simple exponential growth or decay at the rate of the least stable eigenvalue. To characterize the response of this system, more information is necessary. In section 4 we begin our analysis by considering the pseudospectra of the matrix **A**.

## 4. $\epsilon$ Pseudospectra

### 4.1. Definition and Discussion

The results of section 3 demonstrate that the behavior of the system can be more complicated than simple exponential growth or decay. Consequently, the behavior of the system cannot be explained in terms of

the growth or decay of a single normal mode. For systems in which the eigenvectors are nonorthogonal, the concept of the $\epsilon$ pseudospectrum (pseudospectra) of a matrix can be more meaningful than the eigenvalues. Pseudospectra were developed partially in connection with the study of hydrodynamic stability of fluid flows in pipes and channels. In the sections that follow, we will present computations of the pseudospectra of the matrix $\mathbf{A}$ described above. The development of pseudospectra and many of its properties are discussed by *Trefethen* [1992], *Trefethen et al.* [1993], and in *Trefethen* [1997]. In this section we present a brief description of pseudospectra and state some key results that will be used in interpreting the computed results.

In problems involving shear flow in pipes and channels, a discrepancy existed for over 100 years between theoretical predictions of the onset of turbulence and experimental observations. Since the mid-70s it had been recognized that transient effects had a role in the dynamics of the system [*Landahl*, 1975; *Ellingsen and Palm*, 1975; *Benney and Gustavsson*, 1981; *Gustavsson*, 1986]. At about the same time, *Boberg and Brosa* [1988], *Gustavsson* [1991], *Butler and Farrell* [1992], and *Hennigson et al.*, [1993], made the startling discovery that the system was highly nonnormal and that because of this, extremely large transient growth was possible. These authors showed that small perturbations in the flow could be amplified by several orders of magnitude even when all of the normal modes of the system are decaying. The analysis of this problem was further refined by *Reddy and Hennigson* [1993] and *Trefethen et al.* [1993]. *Trefethen et al.* [1993] reconciled the existence of the large amplitude transients with traditional eigenanalysis by calculating the pseudospectra of the linearized system.

Pseudospectra quantify the amount of nonorthogonality of the eigenfunctions and can also be used to calculate bounds on the transient response of the system. Contour plots of the pseudospectra provide a concise graphical representation of the nature of a nonnormal linear system and clearly illustrate whether large-amplitude transients growth is possible. Pseudospectra are not the only possible way to explain the development of the transients in this type of system. *Butler and Farrell* [1992], for example, discuss the transient response in terms of the growth of optimal perturbations.

We first discretize the operator and then calculate the pseudospectra of its matrix representation. In a manner similar to the approach used in *Reddy et al.* [1993] and in *Reddy and Trefethen* [1994]. The $\epsilon$ pseudospectrum of the matrix $\mathbf{A}$ is defined as

$$\Lambda_\epsilon(\mathbf{A}) = \{ z \in C : \|(z\mathbf{I} - \mathbf{A})^{-1}\| \geq \epsilon^{-1} \} \qquad (11)$$

where $\| \ \|$ indicates the matrix norm, which is defined to be the maximum of $\|\mathbf{Av}\|$ over all vectors $\mathbf{v}$ such that $\|\mathbf{v}\| = 1$, and $\mathbf{I}$ is the identity matrix. All of the calculations presented here will be made using the standard Euclidean norm. When $z$ is equal to an eigenvalue, the quantity $\|(z\mathbf{I} - \mathbf{A})^{-1}\|$ is infinite. This corresponds to the $\epsilon$=0 contour.

An equivalent statement to (11) is that for each point $z \in \Lambda_\epsilon(\mathbf{A})$, there exists a vector $\mathbf{u}$ with $|\mathbf{u}| = 1$ such that $\|(z\mathbf{I} - \mathbf{A})\mathbf{u}\| \leq \epsilon$. This shows that the $\epsilon$-pseudospectrum can be thought of as the set of points $z$ in the complex plane that are in this sense $\epsilon$-close to an eigenvalue. As is discussed in appendix A, for a normal system, a point in the $\epsilon$pseudospectrum must lie within a circle of radius $\epsilon$ in the complex plane from an eigenvalue. Only for a nonnormal system can an $\epsilon-$pseudospectral contour extend farther.

Another definition of pseudospectra that is equivalent to (11) is that the $\epsilon$-pseudospectrum is the set of eigenvalues of some perturbed system $\mathbf{A}+\mathbf{E}$, where $\|\mathbf{E}\| \leq \epsilon$. Pseudospectral contours which extend far from the eigenvalues of the system indicate that the eigenvalues of the system are highly sensitive to small perturbations.

It is useful to consider the response of the forced system [*Trefethen et al.*, 1993]

$$\dot{\mathbf{x}} = \mathbf{Ax} + e^{zt}\mathbf{f} \qquad (12)$$

The solution of this equation is $\mathbf{x} = z(\mathbf{I} - \mathbf{A})^{-1}\mathbf{f}e^{zt}$. The magnitude of $(z\mathbf{I} - \mathbf{A})^{-1}$ quantifies how strongly the system can respond at the complex frequency $z$. In the case of a normal matrix the system resonates only at values of $z$ close to the eigenvalues. When the matrix is nonnormal, the region of $z$ for which the system response is large can extend much further from the eigenvalues. The phenomena is referred to by *Trefethen et al.* [1993] as pseudoresonance. We are considering the response of the unforced system. The transient growth that we observe can be thought of as the growth that results from the excitation of these pseudoresonances.

When the $\epsilon-$pseudospectral contours for small values of $\epsilon$ extend significantly into the unstable right half plane, transient growth will be possible for some initial conditions. A lower bound on the transient growth is given by

$$\max \|e^{\mathbf{A}t}\| \ \geq \ \frac{1}{\epsilon}\alpha_\epsilon(\mathbf{A}) \qquad (13)$$

for any $\epsilon \geq 0$, where $\alpha_\epsilon$ is the $\epsilon$-pseudospectral abscissa, defined as the maximum real part of the set $\Lambda_\epsilon(\mathbf{A})$ [*Trefethen et al.*, 1993]. We can immediately see how close the system is to normal by calculating boundaries of the $\epsilon$ pseudspectra. In the event that significant transient growth is possible, we can calculate a lower bound on the growth. This expression will be used in a specific example in the section 4.2.

Another way to see what is happening is to consider the projection of a small initial perturbation onto the set of eigenvectors. For most of the parameter choices we are considering, the eigenvalues are distinct and the

set of eigenvectors forms a complete basis. The initial amplitude of each normal mode is given by the vector $\mathbf{b}$ satisfying the relation $\mathbf{Xb} = \mathbf{y}$, where $\mathbf{y}$ is a vector containing equally spaced samples of the initial conditions, and $\mathbf{X}$ is a matrix whose columns contain the eigenvectors of $\mathbf{A}$. The initial amplitude of each eigenmode is found by evaluating $\mathbf{b} = \mathbf{X}^{-1}\mathbf{y}$. Taking the magnitude of both sides of this expression, we get the relation $\|\mathbf{b}\| \leq \|\mathbf{X}^{-1}\|\|\mathbf{y}\|$. Now consider the case where the initial conditions are small; if $\|\mathbf{X}^{-1}\|$ is very large, this implies that the coefficients of the eigenmodes can be very large. Since the initial conditions are small, these modes must very nearly cancel each other. As the system evolves, there is no guarantee that these large-amplitude modes will continue to cancel each other. Each of the normal modes of the system either grows or decays at the exponential rate given by the eigenvalue corresponding to that mode. The nonmodal growth is the result of the set of eigenfunctions no longer effectively canceling each other as the individual modes grow or decay at different rates. The response of the system is no longer dominated by a single eigenmode but is formed from the superposition of multiple eigenmodes and will in general change as a function of time. A further discussion of pseudospectra and their relation to the nonorthogonality of the eigenfunctions is given in the appendix. The reader should refer to the references in the appendix for a detailed discussion on pseudospectra.

### 4.2. Numerical Results and Interpretation

The pseudospectra of the matrix $\mathbf{A}$ in the vicinity of the least stable eigenvalues are plotted in Figures 4 and 5. Figures 4 and 5 are organized in manner similar to Figures 2 and 3. In Figures 4a, 4b and 4c, there is no shear. In Figures 4d, 4e and 4f there is mild shear. In Figures 5a, 5b and 5c the magnitude of the shear is consistent with measured values reported by *Kudeki et al.* [1981], and Figures 5d, 5e and 5f have a shear which is 50% larger than that. Note that the axes of the plots are not the same when there is shear and when there is no shear. For these calculations, time has been scaled so that one time unit corresponds to 300 s. The units of frequency shown here correspond to $1/300$ s$^{-1}$. The eigenvalues are denoted by the symbol +.

When the shear is zero and the system is close to normal, the contours of the $\epsilon$−pseudospectra are close to a distance $\epsilon$ from the eigenvalues. Figures 4a, 4b and 4c display the least stable eigenvalues and the contour of the $\epsilon = .01$ pseudospectrum for 8 km, 16 km and 32 km wavelengths.

On the basis of pseudospectra we can conclude that large-amplitude transient growth cannot occur. These results suggest that the product $\|\mathbf{X}\|\|\mathbf{X}^{-1}\|$ is close to unity.

For sufficient shear the $\epsilon$-pseudospectral contours are no longer close to a distance $\epsilon$ from the eigenvalues. Figure 4(d) shows the pseudospectra in the vicinity of the least stable eigenvalues for an 8 km horizontal wavelength with a mild shear. Even a perturbation with a magnitude as small as $10^{-8}$ can move the eigenvalues a significant amount. Given that the eigenvalues are so highly sensitive to small perturbations in the system, it is doubtful that attempts to calculate their precise values will produce physically meaningful results. All of the eigenvalues of the system are in the stable, left

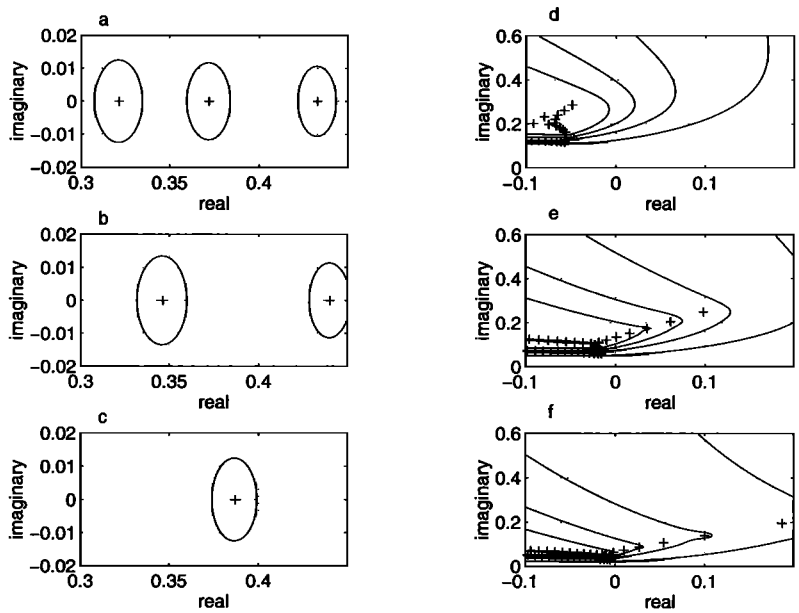


**Figure 4.** Pseudospectra of the matrix $\mathbf{A}$ for (a) an 8 km wavelength without shear, (b) a 16 km wavelength without shear, (c) a 32 km wavelength without shear, (d) an 8 km wavelength with a mild shear, (e) a 16 km wavelength with a mild shear, and (f) a 32 km wavelength with a mild shear. In Figures 4a - 4c a single contour for a value of $\epsilon = 0.01$ is shown. In Figures 4d - 4f the contours are $\epsilon = 10^{-8}$, $10^{-6}$, $10^{-4}$ and $10^{-2}$.

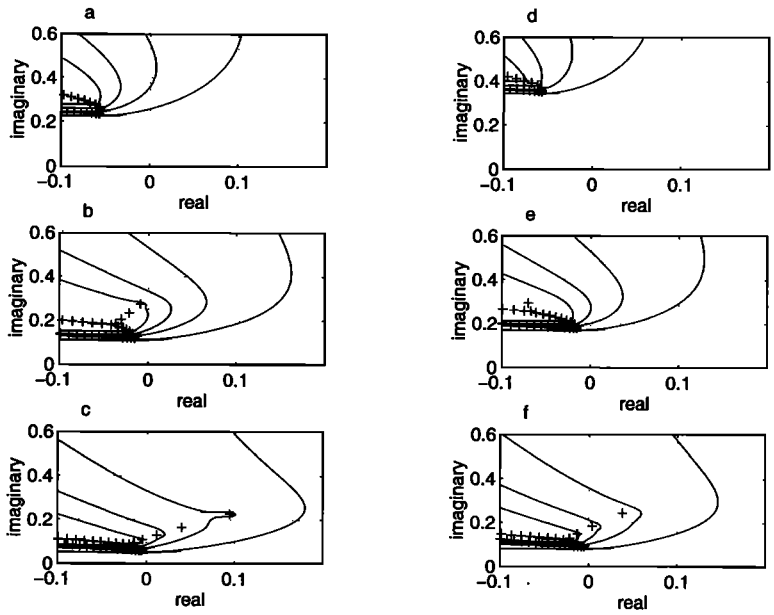


**Figure 5.** Pseudospectral contours of the matrix **A** for (a) an 8 km wavelength with shear, (b) a 16 km wavelength with shear, (c) a 32 km wavelength with shear, (d) an 8 km wavelength with a strong shear, (e) a 16 km wavelength with a strong shear, and (f) a 32 km wavelength with a strong shear. Contours are drawn for $\epsilon = 10^{-8}$, $10^{-6}$, $10^{-4}$ and $10^{-2}$.

half plane. However, a perturbation to the matrix **A** of magnitude $10^{-6}$ can move the eigenvalues into the unstable right half plane. The maximum real part of the $\epsilon = 10^{-6}$ pseudospectral contour is approximately 0.02. From the expression given in (13), this gives a lower bound on the maximum possible transient amplification of a factor of 20,000. This is within a factor of 20 of the maximum transient response seen in Figure 2d, which resulted from integrating 25 trajectories using random initial conditions.

As the shear is increased to the nominal value described in section 3 (Figures 5a and 5d), the eigenvalues move slightly to the left, but the pseudospectral contours move significantly to the left. The pseudospectral contours moving to the left implies that the transient response will be significantly reduced. The estimate of the lower bound on the maximum possible growth is reduced by more than two orders of magnitude. This large reduction in the transient growth is consistent with the results of numerically integrating the system shown in Figures 2d and 3a. From this we can see that it is pseudospectra that describe the behavior of the system, not the eigenvalues.

Similarly, we can compare the transient behavior seen in the numerical integrations shown in section 4.1 with the pseudospectral contours for 16 km and 32 km wavelengths. The large-amplitude transient response at 16 km for nominal and strong shear shown in Figures 3b and 3e again correspond to strong pseudoresonances in the right half-plane (Figures 5b and 5f). At the longest wavelength, 32 km, when the shear is increased (Figures 5c and 5)), the pseudospectral contours move to the left, but not as far as the eigenvalues. It is this increased distance in the complex plane between the eigenvalues and the pseudospectral contours that allows the system to initially grow faster than the least stable eigenmode.

The matrices considered in this paper are nondegenerate; the eigenvectors are linearly independent and hence form a complete basis. The eigenvectors are, however, nonorthogonal. *Fu et al.* [1986] conjectured that the set of eigenfunctions may not form a complete basis and that this may be the source of the transient behavior. The results presented in this section show clearly that it is the nonorthogonality of the eigenfunctions which is the source of the large transient response.

## 5. Optimal Trajectories

In this section we consider the optimal response ot the system in the sense of *Butler and Farrell* [1992] by calculating both the maximum possible initial growth rate and the maximum possible transient amplification as a function of time and wavenumber. These calculations provide an upper limit on the transient response and also give additional insight into the nature of the system. An important question for comparison with experimental observations is, what is the distribution of irregularities as a function of wavenumber? The calculations described in section 5.1 allow us to answer this question.

### 5.1. Optimal Initial Growth Rate

According to (10) the linear system is described by the matrix **A**. The time rate of change of the square

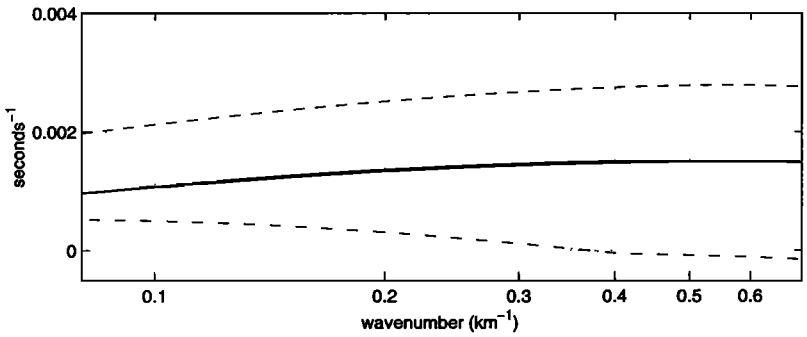


**Figure 6.** The eigenvalues and the maximum possible initial growth rate with and without shear as a function of wavenumber. The eigenvalues are given by the lowest, dotted-dashed curve, the maximum possible initial growth rate without shear is given by the solid curve, and the maximum possible initial growth when shear is included is shown in the upper, dashed curve.

of the magnitude of the density perturbations can be written as

$$\frac{d}{dt}|\mathbf{n}|^2 = (\mathbf{n}, \mathbf{An}) + (\mathbf{An}, \mathbf{n}) \tag{14}$$

where ( , ) is the inner product. Writing this as a matrix product gives

$$\frac{d}{dt}|\mathbf{n}|^2 = \mathbf{n}^{\mathrm{H}}(\mathbf{A} + \mathbf{A}^{\mathrm{H}})\mathbf{n} \tag{15}$$

Dividing by the quantity $|\mathbf{n}|^2$ gives:

$$\frac{d}{dt}\ln(|\mathbf{n}|^2) = \frac{\mathbf{n}^{\mathrm{H}}(\mathbf{A} + \mathbf{A}^{\mathrm{H}})\mathbf{n}}{|\mathbf{n}|^2} \tag{16}$$

The right-hand side of (16) is the Rayleigh quotient of the Hermitian matrix $\mathbf{A} + \mathbf{A}^{\mathrm{H}}$. It is a well-known theorem of linear algebra that the maximum of this expression is equal to the largest eigenvalue of $\mathbf{A} + \mathbf{A}^{\mathrm{H}}$ and that this maximum occurs when the vector n is equal to eigenvector corresponding to the largest eigenvalue. From (16) it is clear that the largest eigenvalue of $\mathbf{A}+\mathbf{A}^{\mathrm{H}}$ is equal to the maximum possible growth rate of the quantity $|\mathbf{n}|^2$. For arbitrary initial conditions this growth rate corresponds to the largest possible initial growth of the system. Note that when the eigenvector corresponding to the largest eigenvalue is not an eigenvector of A, this maximum growth rate cannot be maintained, so that for our system it is only possible for the system to grow at this rate in the limit as $t \to 0$.

In Figure 6 the lower (dotted-dashed) curve is the real part of the least stable eigenvalue for a shear of magnitude consistent with the measured values of *Kudeki et al.* [1981]. As shown by the computation of the pseudospectra near marginal stability the eigenvalues are extremely sensitive to small perturbations to the system. Away from marginal stability, where the least stable eigenmode is either strongly damped or rapidly growing, the eigenvalues are not sensitive to perturbations to the system. The solid curve is the maximum real part of the least stable eigenvalue when there is no shear. The upper (dashed) curve is the largest eigenvalue of the matrix $(\mathbf{A} + \mathbf{A}^{\mathrm{H}})/2$. This gives the maximum possible initial growth rate of the density perturbations. The factor of one half compensates for the square in (16).

We can see from this plot that the shear strongly stabilizes the eigenmodes of the system at the high wavenumbers. If the growth of the least stable eigenmode were representative of the growth of the system over the times of interest, these results would imply that the dominant horizontal scale would be at wavelengths of the order of 70 km. This plot clearly demonstrates the large difference that exists between the initial growth of the system and its long-time behavior described by the eigenvalues. The lower (dotted-dashed) curve shows that all of the normal modes of the system are decaying with time for wavenumbers greater than approximately 0.4 km$^{-1}$. The upper curve shows that there can be initial growth that is considerably faster than even the least stable normal mode of the no-shear system.

At first glance it is surprising that shear can cause the density perturbations to grow more rapidly than when there is no shear. If the shear represents a stabilizing effect, how can shear lead to initial growth at a rate that is faster than if the shear were not present? The answer can be seen by considering the role of the vertical electric field in the potential equation (5). The system of equations describes the motion of the plasma as a result of the perturbation electric fields. The equation for the perturbation electric field (5) has a form similar to Poisson's equation where the terms on the right-hand side play the role of charge density. The increased initial growth rate shows that the vertical field has resulted in a larger effective charge density. When the vertical electric field is removed from the potential equation, the curve describing the maximum initial growth rate and the growth of the least stable normal mode are the same.

Although this result shows that it is theoretically possible for the system with shear to grow initially faster

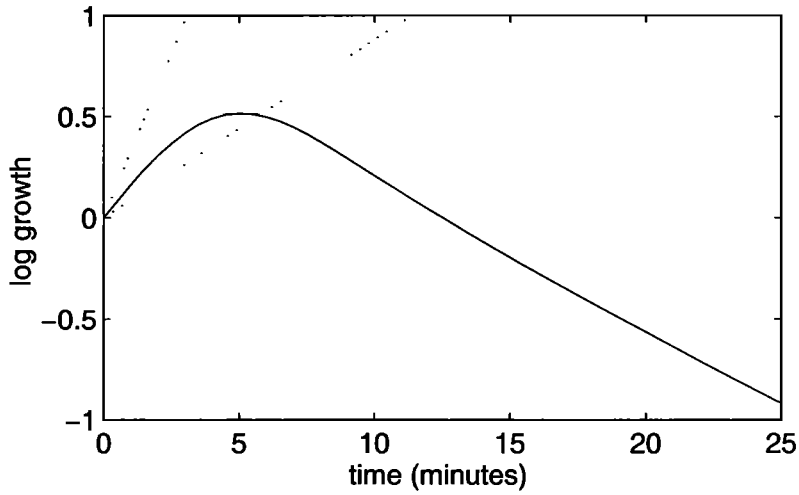


**Figure 7.** Trajectory corresponding to maximum possible initial growth for a 16 km wavelength including shear. The slope of the upper, dotted line corresponds to the maximum possible growth rate. The slope of the lower, dotted line corresponds to the growth rate of the fastest growing eigenmode when there is no shear.

than the no-shear system, numerical integration using random initial conditions shows that when shear is included, the initial growth is typically comparable to the growth without shear. The initial conditions corresponding to the maximum possible growth in the limit as $t \to 0$ do not produce significant transients. The transient growth corresponding to the maximum possible initial growth is shown as a solid line in Figure 7. The slope of the lower dotted line corresponds to the maximum possible growth rate when the shear is zero. The upper dotted line corresponds to the maximum possible growth rate with nominal shear. In the limit as $t \to 0$, the perturbations are growing at the rate given by the upper dotted line. This figure shows the growth rate decreases very rapidly, and by 10 minutes, the perturbation is already beginning to decay.

### 5.2. Optimal Growth as a Function of Wavenumber

With no shear the system quickly approaches the growth rate associated with the least stable eigenvalue. Since the growth can be characterized by a single number, comparing the growth at different wavenumbers is straightforward. When shear is nonzero, this is no longer the case. The growth rate of the transients changes as a function of time, so that the response of the system can no longer be characterized by a single number. We would like to make a calculation that quantifies the transient amplification as a function of wavenumber and to carry out this calculation for several instances of time as the system evolves. The details of how the system evolves depend on the initial conditions. Figure 2 shows that random initial conditions can produce a variety of transient behavior. One possibility for quantifying the transient response is to calculate the maximum possible growth. (This approach was used by *Butler and Farrell* [1992] to study transient behavior in fluid flows in pipes and channels.) We do this to estimate how the transient amplification varies as a function of wavenumber at a given time. By performing this calculation at several times, insight can be gained into how the transient response evolves. While it is unlikely that

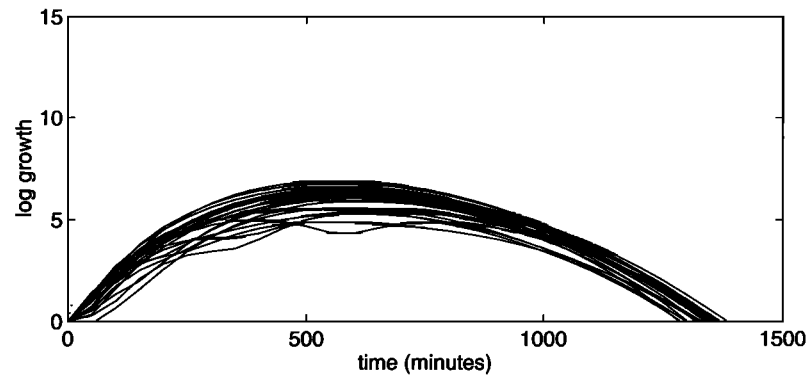


**Figure 8.** The same data as shown in Figure 2d with the maximum possible growth at each time overlayed as a dotted curve.

the initial conditions will be such that the maximum amplification is achieved, we show by example that the maximum possible growth is a reasonable estimate of the behavior of the system for random initial conditions.

The perturbation density at a given time $t$ is given by

$$\mathbf{n}(t) = e^{\mathbf{A}t}\mathbf{n}(t=0) \tag{17}$$

where $e^{\mathbf{A}t}$ is the matrix exponential. The maximum possible transient amplification at time $t$ for the system described by the matrix $\mathbf{A}$ is $\|e^{\mathbf{A}t}\|$. In the examples shown below, the norm was calculated by finding the largest singular value of the matrix $e^{\mathbf{A}t}$.

Figure 8 shows the maximum possible growth of the system as a function of time overlayed on a plot displaying the results of numerically integrating the system starting with random initial conditions (as was done in Figures 2 and 3). For times less than two hours the optimal growth curve (shown as a dotted line) is within three e-foldings of the envelope of the response resulting from the random initial conditions. From Figure 8 we see that as well as being an upper bound on the response of the system, the maximum possible growth is reasonably representative of the behavior of the system resulting from random initial conditions.

The maximum possible growth as a function of horizontal wavenumber was calculated at several times separated by 30 min and the results are shown in Figure 9a - 9e. At the earliest time, 30 min, the response of the system has a broad maximum. By 60 min a dominant wavenumber near 0.3 $km^{-1}$ has emerged. After another 30 min as shown in Figure 9c, the peak has moved to a wavenumber of approximately 0.25 $km^{-1}$. This trend of the peak response moving to lower wavenumbers as time increases continues in Figures 9d and 9e. At the last time shown, 2.5 hours, the maximum response has moved to a wavenumber less than 0.2 $km^{-1}$.

The duration of the enhanced electric field which drives the instability varies night to night, but is typically on the order of 2 hours. Even at 2.5 hours , the dominant wavenumber is considerably higher than that predicted by eigenanalysis. The distribution of the irregularities as a function of wavenumber is continually evolving over the times shown in Figure 9. For times less than 2.5 hours, the response of the system is clearly dominated by transient effects.

These results show an interesting progression of the wavenumber spectrum of the irregularities. Initially, a broad range of wavenumbers will grow. As time increases, the growth rate at the higher wavenumbers decreases. As this happens, the response of the system becomes dominated by lower wavenumbers. This process will continue, and the dominant wavenumber will asymptotically approach the wavenumber of the fastest growing normal mode of the system. The least stable eigenmode and the growth rate associated with it only describe the response of the system in the limit as $t \to 0$.

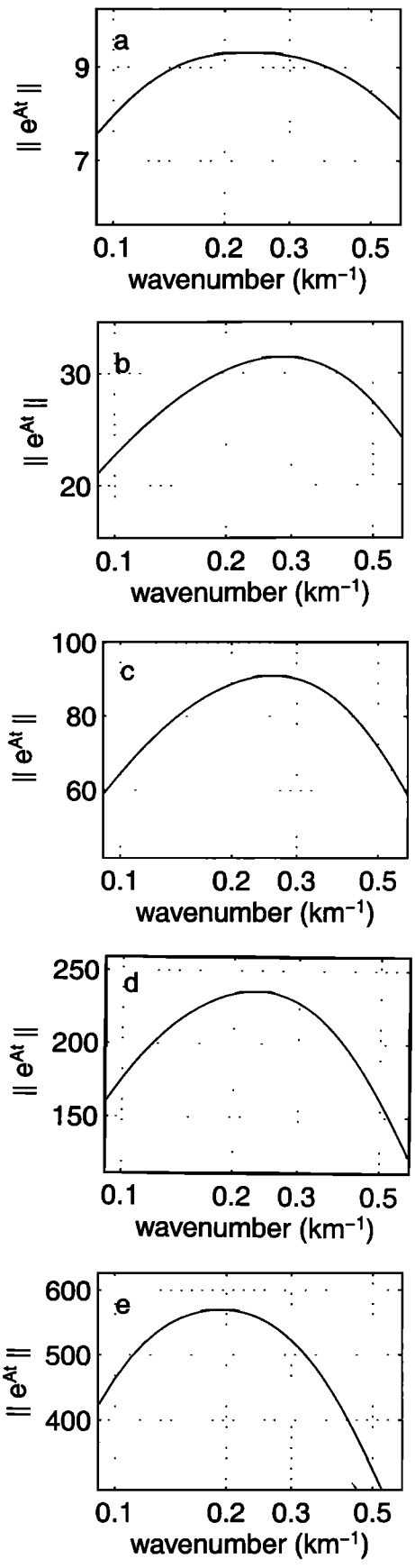


**Figure 9.** The maximum possible transient amplification as a function of wavenumber at (a) 30 min, (b) 60 min, (c) 90 min, (d) 120 min, and (e) 150 min.

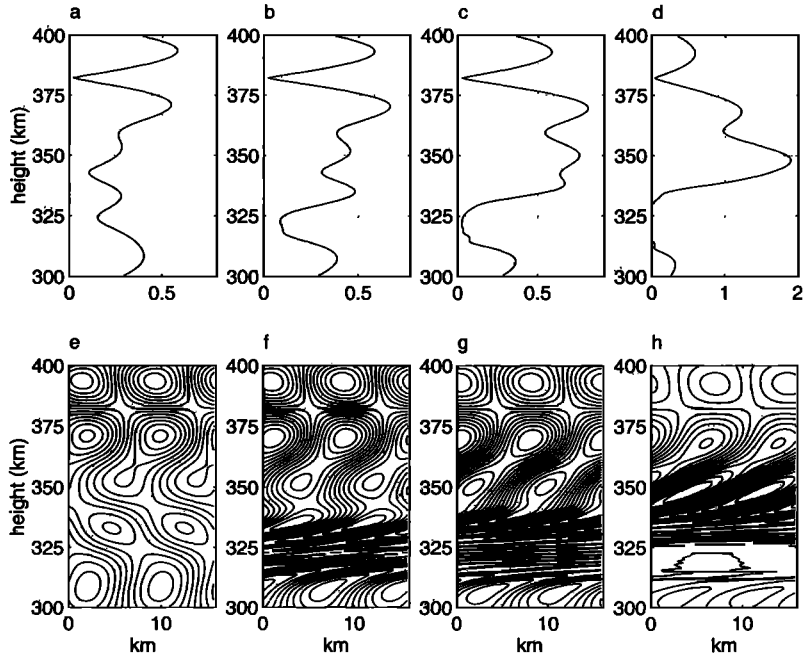


**Figure 10.** (a) - (d) The magnitude of the perturbation density as a function of altitude. (e) - (h) The structure of the perturbation in two dimensions. Perturbation density as a function of altitude at $t = 0$ is shown in Figures 10a and 10e $t = 10$ min as shown in Figures 10b and 10f, $t = 20$ min as shown in figures 10c and 10g, $t = 40$ min as shown in Figures 10d and 10h

## 6. Qualitative Description of the Instability

The instability results from a process analogous to the hydrodynamical instability which occurs when a heavy fluid is above a lighter fluid in the presence of gravity. It is useful to consider the expression for the growth rate that can be derived from local analysis when the shear and the vertical variations in the irregularities are neglected. As is shown by *Kelley* [1989], the instability grows as $e^{\gamma t}$, where $\gamma$ is equal to $(E_0+g/\nu_{in})\nabla \log[n_0(z)]$ plus diffusion. As this expression shows, the existence of the ambient zonal field and gravity (which are combined in the term $E_0$) in the presence of a density gradient drives the instability, and diffusion has a stabilizing influence. Recall from section 2 that $E_0$ is modeled as a constant so that the driving term is directly proportional to $\nabla \log[n_0(z)]$. When a nonuniform vertical electric field is added, the source of the instability is still the ambient field, gravity, and the gradient in density. The nonuniform field causes the irregularities to $\mathbf{E} \times \mathbf{B}$ drift in the zonal direction at a rate which changes with altitude. Qualitative insight into the behavior of this system can be gained by looking at the time evolution of the structure of the unstable region at a single horizontal wavenumber. In Figures 10 and 11, density perturbations are shown at several times during the evolution of the system. The results shown here are for a 16 km horizontal wavelength. The density gradient and shear are given in Figure 1. Figures 10a-10d show the magnitude of the perturbations as a function of altitude. Figures 10e -10h show the same data as contours in two dimensions. Figure 10 shows the initial conditions and the early stages of the evolution at 10, 20 and 40 min. Figure 11 shows the later stages as the transient response reaches a maximum and then begins to decay.

At the earliest times the growth rate of the system is determined primarily by the term $(E_0+g/\nu_{in})\nabla \log[n_0(z)]$. Note for example, the pronounced maximum in the irregularities in Figure 10d in the vicinity of 340 km where $\nabla \log[n_0(z)]$ is a maximum. However, the system does not continue to evolve in this manner. The plots of the density perturbations in Figures 10 and 11 show that the density perturbations have become stretched by the shear flow into horizontally elongated structures. Associated with this stretching in the horizontal direction, sharp gradients in the vertical direction have been formed.

Consider, for the moment, the region in the vicinity of 320 km altitude. In Figure 10f we see the initial stretching, in Figure 10g the vertical gradients have steepened and in 10h, after another 20 min, the structure is no longer present, and the amplitude is close to zero. This sequence illustrates the process of stretching and diffusion by which the shear stabilizes the system. As the perturbations become stretched by the shear, the growth rate is reduced because of the presence of diffusion. In this region the diffusion ultimately removed the perturbations. As can be seen in Figure 1b, the region

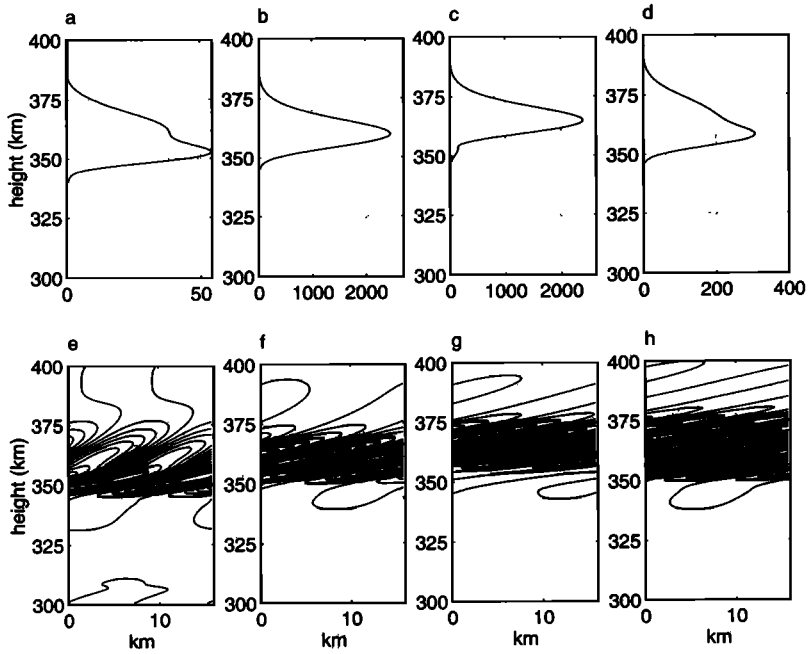


**Figure 11.** (a) - (d) The magnitude of the perturbation density as a function of altitude. (e) - (h) The structure of the perturbation in two dimensions. Perturbation density as a function of altitude at $t = 200$ is shown in Figures 11a and 11e $t = 500$ min as shown in Figures 11b and 11f, $t = 1000$ min as shown in figures 11c and 11g, $t = 1500$ min as shown in Figures 11d and 11h

slightly above 320 km is where the shear is the largest.

Ultimately, as can be seen in Figure 11, the irregularities at all altitudes become stretched into vertically steepened structures. This steepening increases the dissipation due to diffusion and the growth rate at all altitudes is reduced. At altitudes less than approximately 350 km the dissipation due to the stretching and diffusion became greater than the growth due to the driving term described above and the amplitude of the perturbations eventually went to zero.

The process by which the shear generates the high vertical wavenumbers is fundamental to the dynamics of the system. One way to think about how this structure is created is to consider just the portion of the evolution equation (6) which causes the stretching, *i.e.*

$$\frac{\partial n(z,t)}{\partial t} = i\omega(z)n(z,t) \tag{18}$$

where $\omega(z)$ is the oscillation frequency at the altitude $z$ which is given by $k_x E_z(z)$. The solution of this equation is

$$n(z,t) = e^{i\omega(z)t}n(z,t=0) \tag{19}$$

At each altitude the perturbation density is oscillating at a different frequency. As time increases, high $k$ structure will be created as a result of the eventual rapid change in the phase with altitude.

The driving term for the instability is proportional to $\nabla \ln[n_0(z)]$. From Figure 1(c) we can see that this term is greater than zero everywhere on the bottomside of the layer. As the density perturbations at the lower altitudes are being removed by the shear, the density perturbations at the highest altitudes are continuing to grow. As the instability grows the combination of these two effects causes the location of the density perturbations to increase in altitude with time. At the last time shown, the amplitude of the transient response has decreased significantly. We note that through this combined process of growth and diffusion, the shape of the perturbations has broadened slightly and the altitude corresponding to the peak amplitude of the perturbations has decreased slightly.

## 7. Discussion

Our work began as an attempt to understand the long-lived irregularities on the bottomside of the $F$ layer that are often observed in VHF radar measurements. We constructed a numerical simulation of a two-dimensional model and analyzed the results. Now, we ask the following question: Are the results of the modeling consistent with what is observed in the VHF radar measurements made at the Jicamarca Observatory?

As can be seen in many published examples of the phenomena [*Woodman and LaHoz*, 1976; *Kelley*, 1989], bottomside irregularities develop soon after sunset and can often persist for several hours. When bottomside irregularities grow to sufficient amplitude, nonlinear terms begin to dominate the dynamics and bubble-type features will develop. The absence of these bubble-type features in the radar measurements suggests that the irregularities have not grown to the point that the non-

linear terms are significant. The transient behavior described in sections 3 and 6 is qualitatively consistent with this phenomena. The rapid initial growth may explain how the irregularities are visible soon after sunset; and the subsequent slower growth may explain how these features may be present for several hours without nonlinear features developing.

The radar at the Jicamarca Radio Observatory operates at 50 MHz. The strength of the 50 MHz radar returns is proportional to the amplitude of the three $m$ component of the irregularities in the vertical direction. The steepened vertical structures which are produced by the shear may act as a source of the small-scale irregularities which are measured by the radar.

Our calculations predict a gradual progression of the dominant horizontal wavelength of the irregularities. This is a new view of how this instability will evolve. *Valladares et al* [1983] identified a class of long-lived irregularities on the bottomside of the $F$ layer known as bottomside sinusoidal structure (BSS). One characteristic of BSS is that there is a pronounced peak in the wavenumber spectrum corresponding to a horizontal wavelength of approximately 1 km. This aspect of the phenomena is not consistent with the results of our modeling, and we do not believe that the production of BSS can be explained by the effect of a shear on the gravitational interchange instability.

The development of bubble-type features which rise out of the regions of bottomside irregularities is a fundamentally nonlinear phenomenon. On the basis of only a linear model we cannot directly address the question, will the irregularities remain confined to the bottomside of the layer? Solution of the full nonlinear model is presented by *Flaherty* [1997]. An interesting issue that has not been explored is, do the structures which develop as a result of the shear have an effect on the structure of the nonlinearities which develop when the irregularities become sufficiently large?

The irregularities which exist on the bottomside of the $F$ region can be categorized as bottom-type or bottomside spread $F$ [*Woodman and La Hoz*, 1976]. Bottom-type irregularities drift westward (opposite to the direction of the thermospheric wind) and typically extend less than a few kiolmeters in altitude. The westward drift is evidence that the vertical electric field in these regions is controlled by the $E$ region dynamo through the coupling between the $E$ and $F$ regions [*Kelley*,1989]. Bottomside irregularities drift eastward and extend tens of kilometers in altitude. The coupling to the $F$ layer is not as strong and the vertical electric field is controlled by the $F$-region dynamo. An impressive collection of measurements of these phenomena is presented by *Hysell and Burcham* [1998].

The two-dimensional model we have investigated does not account for the coupling that exists between the $E$ region and the $F$ region. We have limited the scope of our efforts to the role of the shear (which is present in both bottomside and bottom-type irregularities) on the development of the instability. A numerical model which attempts to account for the effects of the coupling between the $E$ and $F$ regions is presented by *Hysell and Burcham* [1998]. The coupling to the $E$ region effectively shorts out long-wavelength irregularities One conclusion from their work is that their models produce bottomside irregularities which do not evolve into plume-type structures.

## 8. Conclusion

Equatorial spread $F$ can severely degrade the performance of communications and radar systems. As part of the National Science Foundation space weather initiative, there is considerable interest in developing the ability to predict the occurence of strong equatorial spread $F$ [*Basu et al.*, 1996]. An understanding of the bottomside irregularities which exist prior to the development of the large-scale plume events is a necessary component in developing such a predictive capability. This paper has addressed one component of this problem by investigating the role of shear on the evolution of the gravitational interchange instability.

## Appendix: Nonorthogonality of the Eigenvectors and the Pseudospectra

It is instructive to rewrite the matrix $\mathbf{A}$ in the definition of the pseudospectra as $\mathbf{X}\Sigma\mathbf{X}^{-1}$, where $\mathbf{X}$ is a matrix whose columns are the eigenvectors of $\mathbf{A}$ and $\Sigma$ is a diagonal matrix with the eigenvalues of $\mathbf{A}$ on the diagonal. Rewriting the quantity $\|(z\mathbf{I}-\mathbf{A})^{-1}\|$ using this substitution, we get:

$$\|(z\mathbf{I}-\mathbf{A})^{-1}\| \leq \|\mathbf{X}\|\|\mathbf{X}^{-1}\|\|(z\mathbf{I}-\Sigma)^{-1}\| \qquad \text{(A1)}$$

If the matrix $\mathbf{A}$ is normal, that is, there exists a complete set of orthogonal eigenvectors, then the columns of X can be normalized so that $\|\mathbf{X}\| = \|\mathbf{X}^{-1}\| = 1$. In this case, $\|(z\mathbf{I}-\mathbf{A})^{-1}\| \leq \|(z\mathbf{I}-\Sigma)^{-1}\|$, and in fact, the inequality can be replaced by an equality [*Kato*, 1976]. The quantity $\|(z\mathbf{I}-\Sigma)^{-1}\|$ is equal to the reciprocal of the distance in the complex plane from the point $z$ to the closest eigenvalue. This shows that a point in the $\epsilon$ pseudospectrum must be located within a distance $\epsilon$ from an eigenvalue of the system. When the matrix is nonnormal, it is no longer true that $\|\mathbf{X}^{-1}\| = 1$; in fact, this quantity can become arbitrarily large. When $\|\mathbf{X}^{-1}\|$ is large, we can see from expression (A1) that the $\epsilon$ pseudospectra can extend much farther away from the spectrum.

According to (A1) the $\epsilon$—pseudospectral contours for small values of $\epsilon$ can extend far from the eigenvalues only when the system is nonnormal and the product $\|\mathbf{X}\|\|\mathbf{X}^{-1}\|$ (which is the condition number of the ma-

trix $\mathbf{X}$) is much greater than 1. This implies that the large-amplitude transient amplification is the result of the nonorthogonality of the eigenvectors. An important inequality which exhibits this relationship more directly is to the norm of the matrix exponential $e^{\mathbf{A}t}$ :

$$\|e^{\mathbf{A}t}\| \leq \|\mathbf{X}\|\|\mathbf{X}^{-1}\|e^{\alpha t} \qquad \text{(A2)}$$

where $\alpha$, the spectral abscissa, is the maximum real part of the set of eigenvalues. When the product $\|\mathbf{X}\|\|\mathbf{X}^{-1}\|$ is close to unity, the response is bounded by the growth of the least stable eigenvalue.

**Acknowledgments.** This work was supported by National Science Foundation ATM 9613745. We gratefully acknowledge discussions about spread $F$ and the role of shear with Mike Kelley, Don Farley, Erhan Kudeki, Dave Hysell, Bela Fejer, and Pete Schuck.

Janet G. Luhmann thanks Santimay Basu and P. T. Jayachandran for their assistance in evaluating this paper.

## References

Basu, S., et al., Scintillations, plasma drifts, and neutral winds in the equatorial ionosphere after sunset, *J. Geophys Res . 101*, 26.795, 1996.

Benney, D., and L. Gustavsson, A new mechanism for linear and nonlinear hydrodynamic instability, *Stud. in Appl. Math.*, *4*, 185, 1981.

Boberg, L., and U. Brosa, Onset of turbulence in a pipe, *Z. Naturforsch*, *A43*, 697, 1988.

Budden, K. G., *The Propagation of Radio Waves*, Cambridge Univ. Press, New York, 1988.

Butler, K., and B. Farrell, Three-dimensional optimal perturbations in viscous shear flows, *Phys. Fluids A*, *4*, 1637, 1992.

Ellingsen, T., and E. Palm, Stability of linear flow, *Phys. Fluids*, *18*, 487, 1975.

Flaherty, J. P., Radar observations and numerical modeling of bottomside equatorial spread $F$, Ph.D. thesis, Cornell University, 1997.

Flaherty, J. P., M. C. Kelley, C. E. Seyler, and T. Fitzgerald, Simultaneous VHF and transequatorial HF observations in the presence of bottomside equatorial spread $F$, *J. Geophys. Res.*, *101*, 26811, 1996.

Fornberg, B., *A Practical Guide to Pseudospectral Methods*, Cambridge University Press, Cambridge, 1996.

Fu, Z. F., L. C. Lee, and J. D. Huba, A quasi-local theory of the **E** x **B** instability in the ionosphere, *J. Geophys. Res.*, *91*, 3263, 1986.

Gustavsson. L., Excitation of direct resonances in plane Poiseuille flow, *Stud. in Appl. Math.*, *75*, 227, 1986.

Gustavsson, L., Energy growth of three-dimensional disturbances in plane poiseuille flow, *J. Fluid Mech.*, *224*, 241, 1991.

Guzdar, P. N., P. Satyanarayana, J. D. Huba, and S. L. Ossakow, Influence of velocity shear on the Rayleigh-Taylor instability, *Geophys. Res. Lett.*, *9*, 547, 1982.

Henningson, D. S., A. Lundbladh, and A. Johansson, Energy growth in viscous channel flows, *J. Fluid Mech.*, *250*, 169, 1993.

Huang, C.-S., and M. Kelley, Nonlinear evolution of equatorial spread $F$, 3, plasma bubbles generated by structured electric fields, *J. Geophys. Res.*, *101*, 303, 1996.

Huba, J. D., S. L. Ossakow, P. Satyanarayana, and P. N. Guzdar, Linear theory of the e x b instability with an inhomogeneous electric field, *J. Geophys. Res.*, *88*, 425, 1983.

Hysell, D. L., and J. D. Burcham. JULIA radar studies of equatorial spead $F$, *J. Geophys. Res.*, *103*, 29155, 1998.

Kato, T.. *Perturbation Theory for Linear Operators*, Springer-Verlag, New York, 1976.

Kelley, M. C., *The Earth's Ionosphere*, Academic, San Diego, Calif., 1989.

Kelley, M. C., et al., The Condor equatorial spread $F$ campaign: Overview and results of the large-scale measurements, *J. Geophys. Res.*, *91*, 5487, 1986.

Kudeki. E.. B. G. Fejer, D. T. Farley, and H. M. Ierkic, Interferometer studies of equatorial $F$ region irregularities and drifts, *Geophys. Res. Lett.*, *8*. 377, 1981.

Landahl, M., Wave breakdown and turbulence, *SIAM J. Appl. Math.*, *28*, 733. 1975.

Perkins, F. W., and J. H. Doles III. Velocity shear and the **E x B** instability, *J. Geophys. Res.*, *80*, 211, 1975.

Reddy, S. C., and D. S. Henningson, Energy growth in viscous channel flows, *J. Fluid Mech.*, *252*, 209, 1993.

Reddy, S. C., and L. N. Trefethen, Pseudo spectra of the convection-diffusion operator, *SIAM J. Appl. Math.*, *54*, 1634, 1994.

Reddy, S. C., P. J. Schmid, and D. S. Henningson, Pseudo spectra of the Orr-Sommerfeld operator, *SIAM J. Appl. Math.*, *53*, 15, 1993.

Ronchi, C., Large scale turbulence in the equatorial electrojet, Ph.D. thesis, Cornell Univ., Ithaca, N. Y., 1990.

Satyanarayana, P., P. N. Guzdar, J. D. Huba, and S. L. Ossakow, Rayleigh-Taylor instability in the presence of a stratified shear layer, *J. Geophys. Res.*, *89*, 2945, 1984.

Trefethen, L. N., Pseudospectra of matrices, in *Numerical Analysis edited by D. F. Griffiths and G. A. Watson*, pp. 234–266, Longman, White Plains, N. Y., 1992.

Trefethen, L. N., Pseudospectra of linear operators, *SIAM Rev.*, *39*, 383, 1997.

Trefethen, L. N., A. E. Trefethen, S. C. Reddy, and T. A. Driscoll, Hydrodynamic stability without eigenvalues, *Science*, *261*, 578, 1993.

Tsunoda, R. T., R. C. Livingston, and C. L. Rino, Evidence of a velocity shear in bulk plasma motion associated with the post-sunset rise of the equatorial $F$ layer, *Geophys. Res. Lett.*, *8*, 807, 1981.

Valladares, C. E., W. B. Hanson, J. P. McClure, and B. L. Cragin, Bottomside sinusoidal irregularities in the equatorial $F$ region, *J. Geophys. Res.*, *88*, 8025, 1983.

Woodman, R. F., and C. La Hoz, Radar observations of $F$ region equatorial irregularities, *J. Geophys. Res.*, *81*, 5447, 1976.

Zargham, S., and C. E. Seyler, Collisional interchange instability, 1, Numerical simulations of intermediate-scale irregularities, *J. Geophys. Res.*, *92*, 10073, 1987.

J. P. Flaherty and C. E. Seyler, School of Electrcal Engineering, Cornell, University, Ithaca NY 14853-7501. (e-mail: seyler@ee.cornell.edu)

L. N. Trefethen, Oxford University Computing Laboratory, Wolfson Building, Parks Road, Oxford OX1 3QD, England. (e-mail: Nick_Trefethen@comlab.ox.ac.uk)

(Received May 14, 1998; revised November 30, 1998; accepted December 15, 1998.)