Aust. J. Phys., 1974, 27, 391-400

On the Dynamics of Equatorial Spread F

Tom Beer

Department of Physics, University of Ghana, P.O. Box 63, Legon, Ghana.

Abstract

A theory is proposed to explain the occurrence of spread F in the equatorial ionosphere. It is envisaged as a multistage process involving: movement of ionization drift with atmospheric gravity waves that results in spatial resonance, consequent enhancement of ionization irregularities and eventual strong vertical layering ('pousse-café' effect), which is characterized by the appearance of satellite traces on ionograms; subsequent development of instability in the discrete layers and mixing of the ionization by means of the Rayleigh-Taylor criterion, which leads to blurring of the satellite traces and the observation of spread F. The arguments for the theory are closely related to experimental observations.

Introduction

Spread F is characterized by a blurring of the F region echo on ionosondes. It is generally believed to be caused by partial reflections from small-scale ionospheric irregularities, and observations show that the characteristics, and presumably the method of formation, of spread F at the dip equator are markedly different from the characteristics at non-equatorial locations. Although equatorial spread F has been studied for over 30 years it has remained to a large extent unexplained. The vast amount of data gathered has been reviewed by Herman (1966). Basically, it is well understood how the various radio propagation effects that characterize equatorial spread F can arise if irregularities of electron density aligned with the Earth's magnetic field are present in the ionosphere. The more fundamental task, however, is to account for the generation of the irregularities themselves, and no previous theory has been able to adequately explain equatorial spread F (Farley et al. 1970). Recently Balsley et al. (1972) suggested that the collision-dominated Rayleigh-Taylor instability is capable of generating irregularities that have most of the observed properties. Even though I agree on the importance of the Rayleigh-Taylor instability (it was independently suggested by Koster and Beer 1972), the mechanism proposed by Balsley et al. is still unable to explain the common night-time occurrence of spread F up to 1000 km that is observed on topside ionograms.

Equatorial spread F occurs only at night and in recent years much new information on the behaviour of the night-time equatorial F region has become available. It has been observed that:

(1) The vertical motion of the F layer is directly controlled by the electric field in the E region. This electric field propagates up the highly conducting magnetic field lines which act as equipotentials (Farley *et al.* 1970). During the day the electric field points eastward and the F region ionization moves upwards due to the resulting $E \times B$ drift. The electric field reverses just after sunset and the F region then starts to drift downwards with a speed of the order of 20 m s⁻¹.

(2) The night-time equatorial F region ionization can drift horizontally and eastward with a speed that can exceed 200 m s⁻¹. This drift has frequently been measured in Ghana and elsewhere. A theoretical explanation of the drift has been given by Rishbeth (1971), who envisages the ionization moving with approximately 90% of the neutral wind velocity by means of the presence of polarization electric fields. Faraday rotation experiments in Ghana have confirmed these high horizontal drift motions (Koster and Beer 1972). Fig. 1 depicts the calculated eastward drift velocities.

(3) Transequatorial h.f. radio signals have variations in propagation time and azimuth angle during periods of spread F that are consistent with the presence of atmospheric wave disturbances in the ionization moving from west to east with horizontal wavelengths of approximately 400 km (Röttger 1972, 1973). Furthermore, Raitt and Clark (1973) have reported fluctuations in the equatorial thermospheric electron temperature that appear to indicate the presence of atmospheric waves.



Fig. 1. Calculated eastward drift velocity of F region ionization (after Rishbeth 1971).

A New Theory of Equatorial Spread F

I would suggest that equatorial spread F is a multistage process. The first stage in the production of spread F occurs when the phase velocity of an atmospheric gravity wave matches the component of the ionization velocity in the direction of the wave's phase velocity. Atmospheric gravity waves are manifested by a wavelike perturbation in the neutral gas density and temperature. Wavelike perturbations of the ionization then occur through collisions between the charged particles and the neutral particles and also from changes induced by atmospheric waves in the rate at which the temperature-dependent loss processes operate. If the velocity of the resultant ionization irregularity has a component which is the same as the phase velocity of the gravity wave and in the same direction, then the peaks of the ionization perturbation will continue to stay in the same position relative to the wave. The ionization perturbations will maintain the same phase relation to the wave and as the process continues the ionization irregularity will continue to increase. This process has been termed the spatial resonance effect (Whitehead 1971; Beer 1973*a*). The presence of a number of enhanced ionization layers, one above the other, gives rise to the name 'pousse-café effect' for this theory (Beer 1974). The spatial resonance cannot continue indefinitely and it will cease when the charged particle density of the ionization irregularity has become so large that the system satisfies the collision-dominated Rayleigh-Taylor instability criterion. This is the second stage of the process and it is here that the instability is observed as spread F on ionosondes.

Spatial Resonance

In order to determine whether the spatial resonance can be set up in the F region, we need to know the velocity of the ionization and the parameters of any gravity waves liable to exist in the equatorial F region. The parameters of the possible gravity waves are in turn determined by the height at which the waves are generated so that it is necessary to postulate some source mechanism.

As noted in the Introduction, there is convincing evidence for the existence of atmospheric waves in the equatorial thermosphere (Röttger 1972, 1973; Raitt and Clark 1973), though their source is uncertain. It has been suggested (Beer 1973b) that the supersonic motion of the terminator will generate atmospheric waves, but the expected westward movement of these waves would make them unlikely contenders for a role in the production of spread F, which is always observed to drift eastward. I would suggest here that the gravity waves responsible for the spatial resonance and the subsequent spread F are launched by the equatorial electrojet.

The parameters of atmospheric waves at ionospheric heights have been analysed by Hines (1960, 1964). Short wavelengths are dissipated by molecular viscosity, whose effects become more pronounced as the altitude increases. If the atmospheric waves are generated at the ground then the longer wavelengths will be reflected back to the ground by the stratosphere. However, if the waves are generated by the equatorial electrojet, which lies at an altitude of 100 km, then the long wavelengths which Hines lists as being excluded from the ionosphere will in fact exist there. Thus we can definitely expect to find waves with horizontal wavenumbers less than 10^{-4} m⁻¹ and with vertical wavenumbers less than 10^{-4} m⁻¹ at all periods above the Vaisala Brunt period, which is about 12 min in the thermosphere. As the waves ascend the shorter modes will be dissipated. In all the subsequent analysis it will be assumed that the ionization has a downward velocity of 20 m s⁻¹ and an eastward velocity of 150 m s⁻¹. These appear to be quite typical values for the ionization once it has started its downward drift.

The atmospheric waves capable of generating spatial resonance with the above ionization drift can be found most easily by a geometric construction such as that shown in Fig. 2. Firstly, the ionization drift velocity vector is drawn on cartesian axes representing horizontal and vertical speeds. Then all the lines drawn so as to pass through the tip of the ionization velocity vector represent the phase fronts of waves which will set up a spatial resonance with the ionization. Two illustrative phase fronts are shown in Fig. 2. To construct the phase velocity vector corresponding to each phase front one merely draws the line perpendicular to the phase front that passes through the origin. The intersections of the phase front line with the coordinate axes give the horizontal and vertical trace speeds ω/k_x and ω/k_z , where ω is the wave's angular frequency and k_x and k_z are the horizontal and vertical wavenumbers. It can be seen immediately that the components of the ionization drift velocity provide the minimum values for the trace speeds.

The two illustrative phase fronts in Fig. 2 both represent atmospheric waves that one would expect to find in the F region. The wave with trace speeds 250 and 50 m s⁻¹ has a period of 60 min and wavenumbers $k_x = 6 \times 10^{-6}$ and $k_z = 3 \times 10^{-5}$ m⁻¹. The wave with trace speeds 187 and 100 m s⁻¹ has a period of 22 min with $k_x = 2.4 \times 10^{-5}$ and $k_z = 4.5 \times 10^{-5}$ m⁻¹. It therefore appears that the spatial resonance mechanism will indeed be operative in the equatorial thermosphere.



Fig. 2. Graphical representation of spatial resonance.

Theory

The irregularities responsible for spread F have been observed to be strongly field aligned. This indicates that if spread F is generated by gravity waves, the gravity waves will also be field aligned and have no components whose wave normals are parallel to the magnetic field. Both likely gravity wave source mechanisms would generate gravity waves of this type since the motions of the terminator and of the equatorial electrojet are east-west. Field-aligned irregularities move with the electron drift velocity (Whitehead 1971) so that a mathematical description of this phenomenon only needs to include the parameters of the electron's motion. In fact, as we shall see below, it is irrelevant whether the electron motion or ion motion is considered since the spatial resonance is not controlled by the ionization drift induced by the collisions with the gravity wave but by the motion of the fluctuations in the loss rate.

We will assume that the geographic and dip equators coincide and use a cartesian coordinate system with x eastward and z upward. The electron velocity will be denoted by V and the velocity of the neutral particles by U. During the passage of an atmospheric wave the neutral wind is perturbed and a perturbation velocity is also transmitted to the electrons through their collisions with the neutral particles. At F region heights the Hall motion predominates and the perturbed electron velocity V_1 has components

$$V_{1x} = v^- U_{1z}/\Omega, \qquad V_{1z} = -v^- U_{1x}/\Omega$$
 (1)

in terms of the perturbed neutral wind velocity U_1 , the electron gyrofrequency Ω and the electron-neutral collision frequency v^- .

A quantitative examination of the spatial resonance mechanism is facilitated by transformation into a reference frame moving with the drift velocity of the ionization. The choice of this reference frame means that $V_0 = 0$, where the zero subscript refers to the zero-order background quantities, and also that the Doppler-shifted wave frequency is zero in this reference frame.

The first-order perturbed continuity equation for the ionization n in the F region is then

$$\frac{\partial n_1}{\partial t} + \beta_0 n_1 = -n_0 \left(\frac{\partial V_{1x}}{\partial x} + \frac{\partial V_{1z}}{\partial z} + \frac{V_{1z}}{n_0} \frac{\partial n_0}{\partial z} + \beta_1 \right), \tag{2}$$

assuming that horizontal variations in the background quantities can be ignored. The passage of a gravity wave constitutes a local region of heating and cooling. The attachment-like coefficient β is itself temperature dependent and its perturbations during the passage of a wave are given by (Hooke 1970)

$$\beta_1/\beta_0 = \rho_1/\rho_0 - T_1/T_0 = \{(2R - P)/X\}U_{1x},\tag{3}$$

where P, R, X and Z are the polarization relations for the gravity wave as defined by Hines (1960). The perturbations of the neutral wind speed are sinusoidal and are given by

$$U_{1x}/X = U_{1z}/Z = A_0 \exp\{-i(K_x x + K_z z)\},$$
(4)

where A_0 is the amplitude of the gravity wave and the complex vertical wavenumber K_z with real component k_z is given by

$$K_z = k_z + i/2H,$$

H being the scale height of the neutral atmosphere. Thus equation (2) becomes

$$\frac{\partial n_1}{\partial t} + \beta_0 n_1 = -U_{1x} n_0 \left\{ \frac{v}{\Omega} \left(-\frac{\mathrm{i}K_x Z}{X} + \mathrm{i}K_z - \frac{1}{n_0} \frac{\mathrm{d}n_0}{\mathrm{d}z} \right) + \frac{\beta_0 (2R - P)}{X} \right\}.$$
(5)

Since v^{-}/Ω is very small an evaluation of typical values of equation (5) shows that the variations in the loss coefficient are far more effective in producing equatorial F region spatial resonance than the electron-neutral collisions. Thus

$$\partial n_1 / \partial t + \beta_0 n_1 \approx -U_{1x} n_0 \beta_0 (2R - P) / X.$$
 (6)

For a low frequency gravity wave with $k_z^2 \ge 1/4H^2$

$$P/X = \gamma \omega_{\rm B}/C^2 k_z$$
 and $R/X = i(\gamma - 1)^{\frac{1}{2}}/C$, (7)

where $\omega_{\rm B}$ is the Vaisala-Brunt frequency which is given by

$$\omega_{\rm B} = \left(\frac{(\gamma - 1)g^2}{C^2} + \frac{g}{C^2}\frac{\partial C^2}{\partial z}\right)^{\frac{1}{2}},\tag{8}$$

 $\gamma \approx 1.6$ is the ratio of specific heats and $C = (\gamma g H)^{\frac{1}{2}}$ is the speed of sound.

Equation (6) may be solved to give the rate of growth of spatially resonant irregularities n_1 as

$$|n_1| = (U_{1x}n_0/C^2k_z)\{4(\gamma-1)C^2k_z^2 + \gamma^2\omega_B^2\}^{\frac{1}{2}}\{1 - \exp(-\beta_0 t)\}, \qquad (9)$$

where the value of U_{1x} , which varies sinusoidally in the vertical direction as given by equation (4), determines whether one is examining a peak or trough of n_1 . Applying equation (8) to (9) reveals that

$$|n_1| \approx (U_{1x}n_0/C) \{1 - \exp(-\beta_0 t)\}, \qquad (10)$$

and provided $\beta_0 t \ll 1$

$$|n_1| \approx U_{1x} n_0 \beta_0 t / C.$$
 (11)

Instability

At some time τ after the onset of spatial resonance there will be a number of dense layers of ionization overlaying less dense ones. In general this situation is unstable and the instability that is generated is known as the Rayleigh-Taylor instability. The presence of the un-ionized neutral atmosphere tends to stabilize the situation. Instability occurs when (Liu and Yeh 1966; Balsley *et al.* 1972)

$$n^{-1} dn/dz > \beta_0 v^+/g$$
, (12)

where v^+ is the ion-neutral collision frequency.

The ionization density gradient consists of two parts, due to both the profile of the background ionization n_0 and the spatially resonant irregularities. Thus

$$\frac{1}{n}\frac{dn}{dz} \approx \frac{1}{n_0}\frac{dn_0}{dz} + \frac{1}{n_0}\frac{dn_1}{dz} - \frac{n_1}{n_0^2}\frac{dn_0}{dz}$$
$$\approx \frac{1}{n_0}\frac{dn_0}{dz} + \frac{iU_{1x}k_z}{C}\{1 - \exp(-\beta_0 t)\}$$
(13)

from equation (10), since $U_{1x}/C \ll 1$. The time τ for the onset of instability after the start of spatial resonance is given by

$$\tau = \beta_0^{-1} \ln \left\{ 1 - \left(\frac{\beta_0 v^+}{g} - \frac{1}{n_0} \frac{dn_0}{dz} \right) \frac{C}{k_z U_{1x}} \right\}^{-1},$$
(14)

provided

$$0 < \left(\frac{\beta_0 v^+}{g} - \frac{1}{n_0} \frac{\mathrm{d}n_0}{\mathrm{d}z}\right) \frac{C}{k_z U_{1x}} < 1.$$
 (15)

The right-hand inequality of (15) provides the lower limit for the occurrence of spread F. If the minimum height at which this inequality is satisfied lies above the F region peak then spread F will not be observed on ground-based ionosondes.

Evidence

The pousse-café effect appears capable of explaining some of the hitherto puzzling features of the characteristics of equatorial spread F. In particular, Lyon *et al.* (1961) pointed out that there is a bifurcation of the F region trace on ionosondes which can extend into three or more satellite traces before the onset of spread F. This existence of satellite traces preceding equatorial spread F has also been noted at Huancayo (Pitteway and Cohen 1961), and it appears to be a common feature of both equatorial

and polar spread F (Herman 1966). A number of spatially resonant ionization irregularities formed by the dominant atmospheric wave would produce this characteristic splitting. Lyon *et al.* (1961) pointed out that the observed splitting corresponded to layers separated by 25 km, but that the average separation was 80 km. This would correspond to the vertical gravity wavelength and is consistent with the current ideas on atmospheric gravity wave theory. They also mentioned that the time of first appearance of continuous spreading of echoes varied between zero and 15 min.

The results of Lyon *et al.* (1961) will give a measure of the time τ . Accordingly, equation (14) has been evaluated between 240 and 700 km, assuming no background gradient in the ionization density and using

$$k_z = 2\pi/80 \text{ km}^{-1}, \quad \beta_0 = 10^{-4}(50/H)\exp(300 - z/H) \text{ s}^{-1},$$

 $v^+ = 2.6 \times 10^{-15} M^{-\frac{1}{2}} N \text{ s}^{-1},$

where N is the neutral particle concentration (in m⁻³), z the altitude (km), H the scale height (km) and M the molecular weight. Values of N, M and H were obtained from CIRA (1965) for the case of an exceptionally low level of solar activity (CIRA model 1) and the case of a very high level of solar activity (CIRA model 8), the latter of which corresponds to the conditions encountered by Lyon *et al.* (1961), who took their readings during a period when the sunspot number was 250 and the ionospheric F region peak was at about 550 km or higher.

In order to obtain a value for the amplitude of U_{1x} , the *a priori* assumption was made that U_{1x} was due to fluctuations of 1 m s^{-1} in the equatorial electrojet at 100 km altitude. It was assumed that U_{1x} grows as $\exp(z/2H)$ until viscous damping or some other dissipative mechanism predominates, and that above 400 km U_{1x} has a constant amplitude of 250 m s^{-1} as illustrated in Fig. 3. This assumption agrees with the experimental observation of Balsley *et al.* (1972) that the *F* region convection velocities can be spread over as much as 600 m s⁻¹.

The results for τ , the time between the start of spatial resonance and the onset of instability, are depicted in Fig. 4. It may be seen that under the conditions encountered by Lyon *et al.* (1961), where the satellite traces occurred between altitudes of 450 and 500 km, τ varies between 10 and 20 min. Since at these heights the background electron density is greater than or equal to zero, the observed τ would be somewhat less than the values depicted in Fig. 4. This is in agreement with their results.

The lower limit of instability depends on the upward ionization gradient at the base of the ionosphere. However, since $\beta_0 v^+/g$ attains large values below 250 km, the lower limit will always be very close to this altitude. In the region where $n_0^{-1} dn_0/dz$ is negative there is also an upper limit to the onset of instability. This occurs when $\beta_0 v^+/g$ is so small as to be negligible and

$$C/k_z U_{1x} > |n_0^{-1} dn_0/dz|^{-1}$$
 for $n_0^{-1} dn_0/dz < 0$.

Hanson and Sanatani (1971) have found that the occurrence of equatorial spread F is related to the presence of minute concentrations of long lived Fe⁺ ions. Though Hanson *et al.* (1973) view the Fe⁺ ions as a major instigator of equatorial spread F,



Fig. 3. Assumed amplitude variation of the horizontal perturbation wind velocity U_{1x} (in ms⁻¹) induced by atmospheric waves.



Fig. 4. Calculated results for τ , the time between the start of spatial resonance and the onset of instability in the absence of any background ionization gradients.

the normal vertical rise of the ionization in the pre-sunset period will carry a considerable quantity of the meteoric ions deposited in the E region to great heights. Thus the occurrence of the Fe⁺ ions could well be merely related to the normal upward $E \times B$ drift of the ionization. Hanson and Sanatani note that the presence of at least a trace amount of Fe⁺ is a necessary but not sufficient condition for spread F. I believe that the presence of Fe⁺ is merely an indicator as to whether the daytime rise of the ionosphere has been sufficient to bring the F region peak above the lower limit of instability, in which case Fe⁺ ions will have also been transported upwards. If the $E \times B$ drift during the day is weak, Fe⁺ ions will not be transported into the F region; most of the F region will lie below the lower limit of instability and spread F will not occur.

Discussion

In order to observe spread F on ionograms the results of Fig. 4 indicate that the F region peak must be above 250 km during periods of low solar activity and above 300 km during periods of high solar activity. This then provides an explanation for the observed correlations between solar activity, magnetic activity and spread F. During periods of solar activity, the E region electric fields, which are responsible for the daytime rise of the F region, are greater than during periods of solar minimum (Matsushita 1968) and it has been observed that the F region peak is much higher during solar maximum (when it can exceed 500 km) than during solar minimum (when it may always remain below 300 km). On the other hand, it would appear that the reason for the negative correlation of equatorial spread F with magnetic activity (Lyon *et al.* 1960) arises because during magnetic storms the E region daytime electric field is suppressed and the F region rise is inhibited (Maeda and Sato 1959; Rastogi *et al.* 1971). Furthermore, the thermospheric temperature rises and the lower limit of instability is then higher than it is during non-stormy periods.

I have suggested that the gravity waves responsible for the spatial resonance and the subsequent spread F are launched by the equatorial electrojet. This would also explain why this mechanism of spread F is limited to regions near the dip equator. Gravity waves launched by the electrojet travelling to mid latitudes will be inclined to the vertical sufficiently to increase their phase velocity beyond the maximum capable of generating spatial resonance. Further, gravity waves generated by the electrojet travelling to the F region will have an upward flow of energy corresponding to an upward group velocity. Since the vertical phase and group velocities are oppositely directed (Hines 1960), the phase velocity will be downwards so that spatial resonance in this case will only occur when the F region is moving downwards. Observations of the F-region vertical velocity reveal, however, that the start of spread F can occur for both upward and downward motions of the F region (Farley et al. 1970; Skinner and Kelleher 1971), There are two possible reasons for this. Firstly it is possible that during the upward movement the ionization gradient at the base of the F region increases sufficiently so that $n_0^{-1} dn_0/dz$ is greater than $\beta_0 v^+/g$ and instability takes place directly, without the prior intervention of spatial resonance. This seems to be the case envisaged by Balsley et al. (1972). Secondly, spatial resonance would occur for upward moving ionization if there were a reflection of gravity waves in the upper part of the thermosphere. This is liable to occur just after sunset in the lower F region when the illuminated upper F region is at a higher temperature. If we denote the lower scale height by H_1 and the upper scale height by H_2 and neglect temperature gradients, the reflected waves will have a frequency ω that is given by

since the Vaisala-Brunt frequency $\{(\gamma-1)g/\gamma H\}^{\frac{1}{2}}$ acts as a high frequency limit on

The fact that a spectrum of gravity waves is capable of setting up a spatial resonance with the ionization indicates that the whole of the F region will be covered by turbulent layers, though the quantitative description of this requires a knowledge of the amplitude distribution of the gravity wave spectrum.

References

Balsley, B. B., Haerendel, G., and Greenwald, R. A. (1972). J. geophys. Res. 77, 5625.

Beer, T. (1973a). Planet. Space Sci. 21, 297.

Beer, T. (1973b). Nature 242, 34.

Beer, T. (1974). 'Atmospheric Waves' (Adam Hilger: London).

CIRA (1965). 'Cospar International Reference Atmosphere' (North-Holland: Amsterdam).

Farley, D. T., Balsley, B. B., Woodman, R. F., and McClure, J. P. (1970). J. geophys. Res. 75, 7199.

Hanson, W. B., McClure, J. P., and Sterling, D. L. (1973). J. geophys. Res. 78, 2353.

Hanson, W. B., and Sanatani, S. (1971). J. geophys. Res. 76, 7761.

Herman, J. R. (1966). Rev. Geophys. 4, 255.

Hines, C. O. (1960). Can. J. Phys. 38, 1441.

Hines, C. O. (1964). Can. J. Phys. 42, 1424.

Hooke, W. H. (1970). J. geophys. Res. 75, 7239.

Koster, J. R., and Beer, T. (1972). Physics Dept. Univ. Ghana Report.

Liu, C. H., and Yeh, K. C. (1966). Radio Sci. 1, 1283.

Lyon, A. J., Skinner, N. J., and Wright, R. W. (1960). J. atmos. terr. Phys. 19, 145.

Lyon, A. J., Skinner, N. J., and Wright, R. W. (1961). J. atmos. terr. Phys. 21, 100.

Maeda, K., and Sato, T. (1959). Proc. Instn Radio Engrs 47, 32.

Matsushita, S. (1968). Geophys. J. R. astr. Soc. 15, 109.

Pitteway, M. L. V., and Cohen, R. (1961). J. geophys. Res. 66, 3141.

Raitt, W. J., and Clark, D. H. (1973). Nature 243, 508.

Rastogi, R. G., Chandra, H., and Misra, R. K. (1971). Nature phys. Sci. 233, 13.

Rishbeth, H. (1971). Planet. Space Sci. 19, 357.

Röttger, J. (1972). J. atmos. terr. Phys. 35, 1195.

Röttger, J. (1973). Z. Geophys. 39, 799.

Skinner, N. J., and Kelleher, R. F. (1971). Ann. Geophys. 27, 282.

Whitehead, J. D. (1971). J. geophys. Res. 76, 238.

Manuscript received 16 October 1973