CONTEMP. PHYS., 1975, VOL. 16, NO. 5, 469–488

# Ionospheric Irregularities and Radio Scintillations

B. H. BRIGGS
Department of Physics, University of Adelaide, South Australia

## 1. Introduction

In 1948 the first radio star was discovered. It was noticed soon afterwards that this source, Cygnus A, appeared to fluctuate in intensity in a random way. At first it was thought that the fluctuations were inherent in the source itself, and perhaps rather similar to the fluctuating radio emission known to be emitted by the Sun. However, simultaneous observations showed that there was no correlation between fluctuations recorded at two places 210 km apart, while fairly good correlation was obtained at two places 4 km apart (Smith 1950, Little and Lovell 1950). This suggested that the effects were produced locally, probably in the Earth's atmosphere.

In considering which regions of the Earth's atmosphere are most likely to affect radio waves, it is natural to consider first the region in which the refractive index departs from unity by the greatest amount—the ionosphere. In the ionosphere, the presence of free electrons has the effect of reducing the refractive index below unity. In fact, the refractive index can have any value between unity and zero depending on the wave frequency and the number of free electrons per unit volume. (This is a much larger variation than anything which the neutral atmosphere could produce; its refractive index is only very slightly greater than unity even in the densest region close to the Earth's surface.) It was only necessary to suppose that the electron density in the ionosphere was non-uniform or ' patchy ' in order to arrive at a simple explanation of the observed ' radio scintillation '. This would be quite analogous to optical scintillation of stars except that at optical frequencies the situation is reversed; free electrons cannot vibrate at the very high frequencies of light and so the refractive index of the ionosphere is effectively unity. For light, the regions of the atmosphere where the refractive index departs most from unity are the lower electrically neutral regions of greatest density. Although the departures from unity are small, they are nevertheless sufficient to produce optical scintillation through the effects of winds and turbulence, as was already well known.

Confirmation that the ionosphere was the cause of the radio scintillation phenomenon was soon forthcoming by a comparison with radio soundings of the ionosphere by the echo technique. In these experiments, a radio transmitter on the ground sends a short pulse of radio waves upwards, and the reflected pulse is received on a nearby receiver. It was observed that echoes from the higher region of the ionosphere (the so-called ' F region ' at a height of about 350 km) did not return at a definite time after the transmitted pulse, but were spread over a considerable range of times, indicating that the radio waves were being reflected from irregularities, not from a well defined height. Echoes of this type are called ' spread-F ' echoes. There was a close correlation between the times at which spread-F echoes were observed and the times when radio scintillations of Cygnus A were observed. Both phenomena occurred mainly at night, except occasionally during magnetic storms, when they could occur in the daytime.

When artificial satellites were launched in the period 1957 onwards it was possible to observe scintillations of radio signals received from transmitters which they carried. This was done at first with orbiting satellites and later with geostationary satellites. This work confirmed deductions already made from observations on radio stars, and added new information about the heights and global distribution of the ionospheric irregularities. In the last few years it has been possible to measure the electron density fluctuations in the ionosphere by means of probes carried on satellites, thus providing direct *in situ* information about the scales and intensities of the irregularities.

To a radioastronomer ionospheric scintillation is a nuisance; it troubles him just as optical scintillation troubles the optical astronomer. Ionospheric physicists, however, have been able to make use of the phenomenon to study the irregularities of electron density in the ionosphere, and this is what the present article is mainly about. The study of ionospheric irregularities is of scientific interest because the irregularities must be produced by turbulent processes, plasma instabilities, incoming particle streams, or other physical processes which need to be studied if the upper atmosphere is to be understood completely. The subject is also of practical importance because of the use of the ionosphere for radio communication, and possible scintillations of signals from the recently introduced communication satellites.

## 2. The ionosphere

We shall be concerned in this article with radio waves transmitted through the ionosphere from radio stars and satellites. It is first necessary to understand the factors which determine whether a radio wave can penetrate the ionosphere or will be reflected back from it. We also need to know how fluctuations of electron density will affect waves which traverse the ionosphere.

The refractive index of an ionized gas depends on the number of electrons per unit volume (the electron density $N$) and the wave frequency $f=\omega/2\pi$. If collisions of electrons with other particles are neglected, and the Earth's magnetic field is ignored, the refractive index $n$ is given by the formula

$$n=\left\{1-\frac{Ne^2}{\epsilon_0 m_e \omega^2}\right\}^{1/2} \tag{1}$$

where $e$ and $m_e$ are the charge and mass of the electron. Suppose a radio wave is travelling vertically upwards from the ground, as in the echo sounding experiments. Initially $N=0$ and $n=1$. When the wave first enters the ionized region, $N$ will be an increasing function of height, and $n$ will decrease. The wave may (depending on its frequency) reach a level where $n=0$. It can be shown that the wave will be reflected from this level (the group velocity tends to zero, and the wave velocity to infinity). Low frequency waves may reach this point in the lower part of the ionosphere, because the value of $N$ required to make $n=0$ is quite small. If the wave frequency is increased, reflection will take place at a greater height because a larger value of $N$ is required. Now there will be some level at which $N$ is a maximum (it cannot increase without limit because the ionosphere must ultimately merge with interplanetary space where $N$ is very small) and this maximum $N=N_{\max}$ actually occurs in the F region of the ionosphere at a height in the region of 300–400 km. The highest frequency which will be reflected is therefore found by putting $N=N_{\max}$, $\omega=\omega_c$, and $n=0$ in eqn. (1). This gives

$$\omega_c{}^2 = \frac{e^2 N_{\max}}{\epsilon_0 m_e} \tag{2}$$

The frequency $f_c=\omega_c/2\pi$ is called the critical frequency of the F region. It can easily be measured using ground based transmitters and receivers. It is clear that $f_c$ is also the highest frequency which can penetrate the ionosphere and reach the ground if the source is a radio star or a satellite beyond the atmosphere. In practice $f_c$ is of the order of 2–15 MHz depending on time of day, season, etc.

Most experiments on radio scintillations use frequencies considerably higher than the critical frequency so that the waves not only can penetrate the ionosphere but are not too much disturbed by it. If $\omega \gg \omega_c$ it is clear from eqn. (1) that $n$ is only slightly less than unity and to the first order is given by the expression

$$n \simeq 1 - \tfrac{1}{2}\,\frac{Ne^2}{\epsilon_0 m_e \omega^2} \tag{3}$$

If there is a fluctuation of electron density of magnitude $\Delta N$ around the mean, the associated fluctuation of refractive index will be given by

$$\Delta n = -\tfrac{1}{2}\,\frac{e^2}{\epsilon_0 m_e \omega^2}\,\Delta N = \frac{-\lambda^2 r_e}{2\pi}\,\Delta N \tag{4}$$

where $\lambda$ is the wavelength of the radio wave, and $r_e=e^2/(4\pi\epsilon_0 m_e c^2)$ is introduced as a useful constant to simplify the equation (it is actually the ‘ classical radius ’ of the electron). Eqn. (4) is basic to the understanding of ionospheric scintillation.

For a general review of the physics and chemistry of the ionosphere, reference may be made to Rishbeth 1973.

## 3. Radio scintillations: theory

A radio star emits radiation over a wide spectrum of radio frequencies, but the receiver will accept and amplify only frequencies which lie within a relatively narrow band. The centre frequency of the receiver pass-band determines the wave frequency $\omega$ with which we are concerned, and we will neglect effects arising from the finite width of the receiver response. The waves from the radio star will be plane because the source is effectively at infinity. Sometimes the finite angular diameter of the radio star needs to be taken into account, but we will not consider this factor here.

If the waves are from a satellite the frequency is determined by the transmitter on the satellite. The wavefronts will be spherical because the satellite is at a finite distance from the Earth.

To outline the essential features of the theory of diffraction by the irregular ionosphere we will consider the case of a radio star (fig. 1). The modifications needed to take account of the curvature of the wavefront in the case of a satellite are relatively straightforward, and do not introduce any new matters of principle. To further simplify the problem initially, we will also assume that the electron density varies only in the plane of the diagram, and does not vary in the direction perpendicular to the diagram.

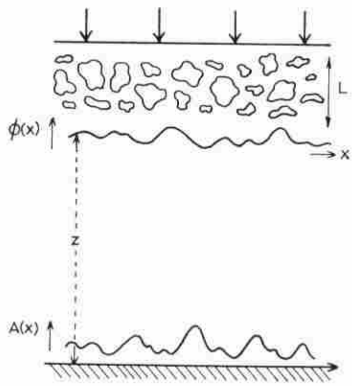


Fig. 1. A plane wavefront from a radio star is distorted in passing through the ionosphere, because of the random variations of refractive index. The phase $\phi(x)$ on emergence from the ionosphere varies randomly across the wavefront. At the ground, both amplitude and phase variations occur, as explained in the text; only the amplitude $A(x)$ is sketched in the figure.

It is clear that the initially plane wavefront will be distorted in passing through the irregular ionosphere, and will emerge from the bottom of the ionosphere with random perturbations impressed on it. Assuming that the wave frequency $\omega$ is high enough for equations of Section 2 to be valid, it is clear that these perturbations will be of phase only, and can be calculated by integration through the medium along straight line paths. This is because refraction will be negligible if the refractive index departs only slightly from unity.

Therefore, the phase at $x$, relative to what it would be if the ionosphere were absent, is given by

$$\phi(x) = \frac{2\pi}{\lambda} \int \Delta n \, \mathrm{d}z \tag{5}$$

the integral to be evaluated over all parts of the path for which $\Delta n$ is significantly different from zero.

Combining eqns. (4) and (5) we obtain

$$\phi(x) = -\lambda r_{\mathrm{e}} \int \Delta N \, \mathrm{d}z \tag{6}$$

This equation brings out two important features of ionospheric scintillation. Firstly, the phase variation impressed on the wave is proportional to the integrated variation of electron density along the ray path, and secondly the phase variation is proportional to the wavelength. Therefore, we expect scintillation effects to become less important as the radio wavelength becomes smaller.

We now need to consider how the random fluctuations of electron density $\Delta N$ can be more precisely specified. Useful quantities for describing any type of random variable are the r.m.s. departure from the mean $[\langle\Delta N^2\rangle]^{1/2}$, which describes the magnitude of the fluctuations, and the auto-correlation function $\rho_{\Delta N}(r)$ which describes the scale of the fluctuations. The auto-correlation function measures the degree of correlation between the electron density at two points a distance $r$ apart. In general $\rho_{\Delta N}(r)$ will fall off from unity at $r=0$ to a small value at $r=r_1$, say, in which case we can use $r_1$ as a measure of the 'scale' of the fluctuations. Alternatively, the fluctuations can be Fourier analysed into sinusoidal components, and specified by the way the amplitude varies as a function of frequency. By 'frequency' we mean in the present case 'spatial frequency', because the Fourier components are sinusoidal in space, and each is specified by its spatial period $d$ (distance from crest to crest) or alternatively by the number of cycles per unit length $k=1/d$ which is the spatial frequency†. The square of the modulus of the Fourier spectrum, which tells us how the power varies as a function of spatial frequency $k$ is often called the power spectrum. It can be shown that the power spectrum is the Fourier transform of the auto-correlation function, and so the two descriptions are closely related. For the moment we will use the description in terms of the auto-correlation function and will return to the spectral description later on.

A useful functional form for $\rho_{\Delta N}(r)$, much used in the literature, is the Gaussian:

$$\rho_{\Delta N}(r)=\exp(-r^2/r_0^2) \tag{7}$$

In this case, $\rho_{N\Delta}(r)$ falls to $e^{-1}\simeq 0{\cdot}37$ when $r=r_0$, and so $r_0$ is a convenient measure of the scale. The Gaussian form is chosen mainly for its mathematical convenience. It has some physical justification because diffusion tends to produce irregularities in which the density varies in this way. (Recent evidence suggests, however, that it may not always be a good choice for the irregularities in the $F$ region, as will be shown later.) Its convenience lies in the fact that it includes only one length parameter, and usually leads to expressions which are integrable.

Suppose a wave travels a distance $L$ through an ionized medium in which there are fluctuations of electron density of r.m.s. value $[\langle\Delta N^2\rangle]^{1/2}$ and which have an auto-correlation function given by eqn. (7). What is the r.m.s. value $\phi_0=[\langle\Delta\phi^2\rangle]^{1/2}$ of the phase departures from the mean phase across the emerging wavefront (fig. 1)? This problem can be solved by applying eqn. (6) and carrying out the required integration through the medium. We will quote the result:

$$\phi_0=\pi^{1/4}\lambda r_e(r_0L)^{1/2}[\langle\Delta N^2\rangle]^{1/2} \tag{8}$$

The significant factors in this equation have a simple physical interpretation. The medium may be thought of as an assembly of electron 'clouds' of size $r_0$. *One* cloud produces a phase change of $2\pi r_0\Delta n/\lambda$ as the wave passes through it (i.e. $2\pi/\lambda$ times the 'optical path'). From eqn. (4), we see that the phase change due to one electron cloud of excess density $\Delta N$ is proportional to $r_0\lambda\Delta N$. But in a distance $L$ the wave will traverse $L/r_0$ such clouds. We

† Note that $k$ is often used in the literature for the *wave number* $2\pi/d$, rather than $1/d$. Unfortunately there is no standard symbol for $1/d$.

must not simply multiply by this factor, because the clouds may have either increased or decreased density and their effects add randomly rather than coherently. The resultant fluctuation in such a situation is proportional to the square root of the number which are effective. Multiplying the factor $r_0\lambda\Delta N$ by the factor $(L/r_0)^{1/2}$ we obtain all the important terms in eqn. (8), but not of course the numerical factor $\pi^{1/4}$.

We know that in practice, radio star scintillations take the form of fluctuations of both amplitude (or intensity) and phase. In practice the intensity fluctuations are easier to observe, and have therefore received most study. Since the ionosphere can impress only phase fluctuation across the wavefront, it is at first sight rather surprising that intensity fluctuations are observed. We now consider this problem, which will involve a consideration of some aspects of diffraction theory (Ratcliffe 1956).

The key to the paradox is that we do not observe the fluctuations as soon as the wave emerges from the lower side of the ionosphere. We observe at the ground, and the wave has therefore propagated a considerable distance through free space before it reaches our receiver. Diffraction effects taking place during this propagation can generate intensity fluctuations when the wave arrives at the ground. Of course, it is well known that phase fluctuations can be converted to amplitude fluctuations in diffraction situations; this is the principle of the phase contrast microscope. The phase contrast microscope, however, is a rather complex instrument. In the present case, free space propagation only is involved.

To understand how this effect comes about let us consider, for simplicity, the case where the phase variation across the emerging wavefront is sinusoidal:

$$\phi(x) = \phi_1 \cos\left(\frac{2\pi x}{d}\right) \tag{9}$$

Here $\phi_1$ is the magnitude of the phase deviation and $d$ the distance between maxima (the spatial period) (fig. 2, left). This assumption is not as restrictive as may appear at first sight, because provided the problem is linear, we can always analyse the actual phase variation $\phi(x)$ into sinusoidal components by

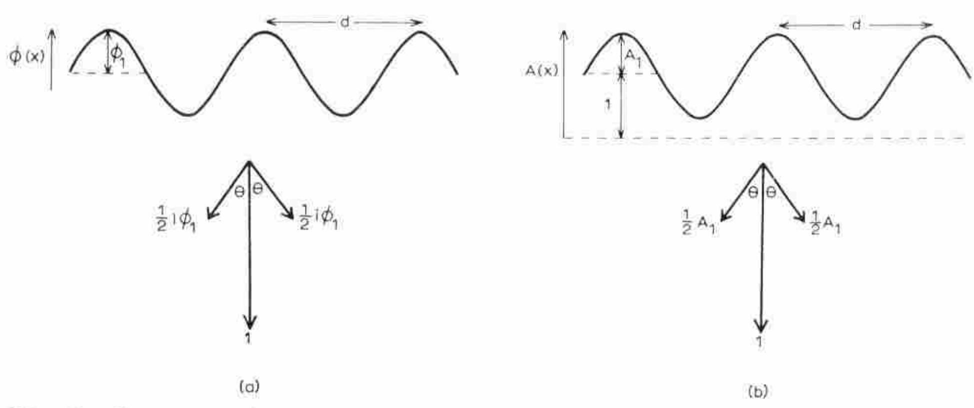


Fig. 2. Spectra produced by sinusoidal phase and amplitude gratings. *Left*: A phase grating of period $d$ and amplitude $\phi_1$ produces first order spectra $\frac{1}{2}i\phi_1$ at angles $\pm\theta$ given by the formula $\sin\theta = \pm\lambda/d$. *Right*: An amplitude grating produces spectra at the same angles, but *in phase* with zero order spectrum.

Fourier analysis, and consider one Fourier component at a time. That the problem *is* linear, for small phase deviations, can be demonstrated as follows. The actual variation of wave field $E(x)$ across the wave front is given in complex form by

$$E(x) = \exp\ \{i\phi(x)\} \tag{10}$$

where we assume, for simplicity, that the field has unit amplitude. This equation says nothing more than that the wave has unit amplitude and a phase $\phi(x)$ at the point $x$. Eqn. (10) can be expanded as a series

$$E(x) = 1 + i\phi(x) + \frac{i^2\phi^2(x)}{2} + \ldots \tag{11}$$

and if $\phi(x)$ is small (i.e. less than one radian) we need retain only the first two terms:

$$E(x) = 1 + i\phi(x) \tag{12}$$

Therefore, each Fourier component of $\phi(x)$ corresponds to a single Fourier component of the wave field $E(x)$, and components may simply be added linearly to obtain the total wave field. This is not, of course, true if the higher order terms such as $\phi^2(x)$ are included. Then two Fourier components of $\phi(x)$, when added and squared, produce *new* Fourier components (sum and difference terms) in $E(x)$, not present in $\phi(x)$; the problem is non-linear, and much more difficult to handle.

Considering the linear case, we can substitute for $\phi(x)$ from eqn. (9) into (12) to obtain an expression for the wave field:

$$E(x) = 1 + i\phi_1 \cos\left(\frac{2\pi x}{d}\right) \tag{13}$$

This is the starting point for any diffraction problem; we need to know the wave field across the aperture or diffracting 'screen' before we can discuss how it becomes modified by diffraction as it propagates beyond the screen.

Now eqn. (13) and fig. 2 remind us very much of an ordinary diffraction grating, for which the theory is familiar from elementary optics. The only difference is that an ordinary grating has alternate transparent and opaque strips, so that the variations are of amplitude not phase, and they are certainly not sinusoidal. However sinusoidal amplitude gratings are easily discussed theoretically and have been constructed photographically (e.g. by photographing Young's interference fringes). It is known that they differ from ruled gratings only in that a zero order and two first order spectra only are produced; there are no higher orders. The angles at which the first order spectra occur are given by the usual grating formula

$$\sin\theta = \pm\frac{\lambda}{d} \tag{14}$$

where $\theta$ is the angle from the normal to the grating.

Let us now compare the amplitude and phase gratings. A sinusoidal amplitude grating would be described mathematically by the equation

$$E(x) = 1 + A_1 \cos\left(\frac{2\pi x}{d}\right) \tag{15}$$

The mean amplitude is assumed to be unity, and we must have $A_1 < 1$ so that the amplitude does not go negative. It will be seen that this equation is exactly the same as eqn. (13) for the 'phase grating' except that $A_1$ replaces $i\phi_1$. It therefore is reasonable to suppose (and it can easily be proved rigorously) that the phase grating will produce first order spectra at the same angles as the amplitude grating (as given by eqn. (14)) and that the only difference will be in the *phases* of these spectra relative to the zero order beam. These spectra are all plane waves; their relative amplitudes, directions and phases as they leave the gratings are shown schematically by the arrows (wave normals) in figs. 2 (*a*) and (*b*).

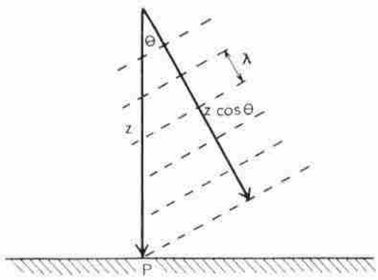


Fig. 3. A wavefront travelling at an angle $\theta$ to the vertical arrives at $P$ after travelling a distance $z \cos\theta$, which is less than the distance $z$ travelled by the wave at $\theta = 0$.

We now come to the crucial point. In propagating down to the ground, the phases of the spectra at $\pm\theta$ change relative to the zero order or undeviated wave because they travel different distances, as shown in fig. 3. Suppose the undeviated beam at $\theta = 0$ progresses a distance $z$, and arrives at the ground at $P$. A wavefront at an angle $\theta$ arrives at $P$ after travelling a distance $z \cos\theta$. The phases of the first order spectra have therefore *gained* in phase, relative to the undeviated wave, by an amount $2\pi(z - z\cos\theta)/\lambda$. In practice, $d$ is much larger than $\lambda$, and $\theta$ is therefore small (see eqn. (14)). Therefore, we can expand $\cos\theta$ as $(1-\theta^2/2)$, and the phase lead can be written $\pi z\,\theta^2/\lambda$. When this phase lead is $\pi/2$ the first order spectra from the phase grating will be *in phase* rather than in quadrature with the undeviated wave, as in fig. 2 (*b*), and will therefore add to it to produce only amplitude variations over the ground, and no phase variations. The distance $z_0$ at which this will occur is given by

$$z_0 = \frac{\lambda}{2\theta^2} = \frac{d^2}{2\lambda} \tag{16}$$

since $\theta \simeq \lambda/d$ from eqn. (14). Thus at a distance of $d^2/2\lambda$ beyond a sinusoidal phase grating, we will observe sinusoidal variations of amplitude, with the same spatial period $d$ as the phase grating. If the screen is in horizontal motion with speed $V$ these will be observed as amplitude fluctuations of period $d/V$. The fractional variation of amplitude is just $\phi_1$, as is clear from a comparison of figs. 2 (*a*) and 2 (*b*) (assuming that $A_1$ and $\theta_1$ are both small compared with unity).

At twice this distance from the screen the spectra will be in quadrature again with the undeviated wave, and we return to a pure phase variation with no amplitude variation, and so on, the fluctuations changing from phase to amplitude and back again periodically in every increment of distance $\Delta z = d^2/2\lambda$. In between these planes, variations of both phase and amplitude will occur. If we now consider the more realistic case of a random phase variation or 'random phase screen', we see that every Fourier component will change periodically from phase to amplitude but the changes will occur at different distances from the screen, because the critical distance $z_0$ depends on the value of $d$. If we examine the Fourier spectrum of the amplitude $A(x)$ across the ground at a distance $z$ we would expect certain Fourier components to be missing, because they will be represented only by phase fluctuations. The periods $d_{\mathrm{m}}$ of these missing Fourier components at a distance $z$ can easily be found by setting $z$ equal to an even multiple of the critical distance $d^2/2\lambda$, which gives at once

$$d_{\mathrm{m}} = \left(\frac{\lambda z}{m}\right)^{1/2} \tag{17}$$

where $m$ is an integer. The 'missing' spatial frequencies will, of course, be $k_{\mathrm{m}}$, where $k_{\mathrm{m}} = 1/d_{\mathrm{m}}$.

Close to the phase screen, every Fourier component will be in the process of changing from phase to amplitude for the first time, and therefore the percentage amplitude fluctuations will increase with distance, starting from zero close to the screen, and reaching a constant value at a large distance where *some* components are always in the 'amplitude' state.

We do not intend to pursue the mathematical development further than this, but will now quote some results. The discussion given so far will provide the background needed to understand the significance of these results.

To obtain definite results it is, of course, necessary to assume a specific form for the auto-correlation function (and therefore for the Fourier spectrum also) of the ionospheric irregularities. If we take the Gaussian case (eqn. 7) it turns out that the correlation function of the phase of the emerging wavefront is also Gaussian with the same scale factor $r_0$, i.e.

$$\rho_\phi(r) = \exp\left(-r^2/r_0^2\right) \tag{18}$$

The wavefront is two-dimensional, and the irregularities are assumed to be isotropic, i.e. the correlation between the phase at two points depends only on the distance $r$ between these points and not on their orientation. The *spatial power spectrum* of the phase irregularities is the Fourier transform of $\rho_\phi(r)$ and is

$$W(k) = \exp\left(-\pi^2 r_0^2 k^2\right) \tag{19}$$

Eqn. 19 says that if the phase variations are Fourier analysed into periodic components of different spatial frequencies, the power for the frequency $k$ is $W(k)$.

Fig. 4 shows how the spectrum for the amplitude fluctuations can be found at any distance $z$ from the screen. The Gaussian spectrum of the phase fluctuations is shown at the top (eqn. (19)). This has to be multiplied by the quasi-periodic function shown in the middle diagram, giving a resultant spectrum as shown in the lower diagram. The missing frequencies occur at

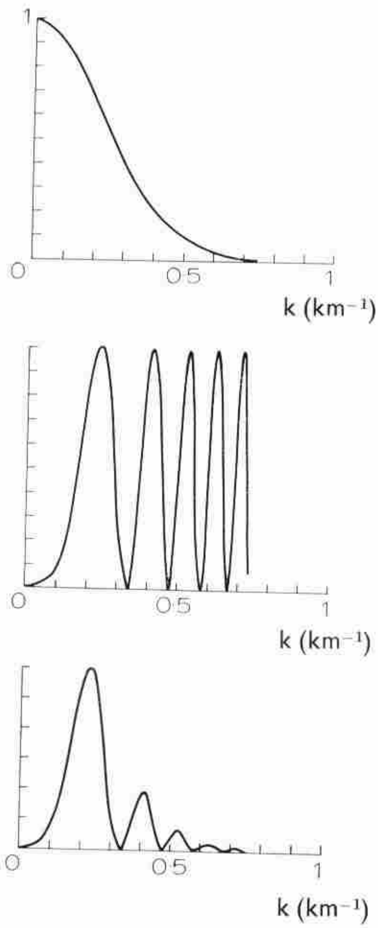


Fig. 4. This diagram illustrates how the Fourier spectrum of the amplitude fluctuations at a given distance from a phase screen is related to the spectrum of the phase fluctuations. Each diagram is a section of a two-dimensional spectrum obtained by rotating the diagrams around the vertical axis. The horizontal axis is spatial frequency $k$ (measured in units of cycles per kilometre). *Top*: spectrum of the phase fluctuations. In the example this is Gaussian with a scale parameter $r_0 = 1$ km, but any form of spectrum could be used. *Centre*: Fresnel filter function for $\lambda = 15$ m and $z = 600$ km. For these values, $z\lambda = 9\ \text{km}^2$ and $1/(\lambda z)^{1/2} = 0{\cdot}33\ \text{km}^{-1}$. The zeros occur at $k = 0{\cdot}33$, $0{\cdot}33 \times \sqrt{(2)} = 0{\cdot}47$, $0{\cdot}33 \times \sqrt{(3)} = 0{\cdot}575$, $0{\cdot}33 \times \sqrt{(4)} = 0{\cdot}66$, as expected from eqn. (17). *Bottom*: The spectrum of the amplitude fluctuations, obtained by multiplying together with two upper curves.

the expected positions as given by eqn. (17). Since the problem is really two-dimensional, the actual spectra are two-dimensional, and are to be obtained by rotating each diagram around the vertical axis through the origin.

The oscillations in the spectrum are sometimes called ' Fresnel oscillations ', and the multiplying function which produces them is sometimes called the ' Fresnel filter function '. This is because of its close connection with the Fresnel zone size. The first Fresnel zone for a distance $z$ has a radius of $\sqrt{\lambda z}$, and the reciprocal of this length gives the position of the first zero. The first main peak of the Fresnel filter function occurs at a scale approximately equal to *the diameter* of the first Fresnel zone. Fig. 4 is really saying that irregularities of this size are most effective in producing amplitude fluctuations at the distance $z$. Irregularities larger than the Fresnel zone are ' filtered out ' (the filter function falls to zero at the origin). This does not mean that large scale irregularities are undetectable at the ground. They can produce phase fluctuations, and associated refraction effects, but not amplitude fluctuations. Of course, the primary spectrum does not have to be Gaussian. The Fresnel filter multiplies the original spectrum whatever it may be.

The Fresnel oscillations are only well-developed if the layer of irregularities is thin and the phase deviation small. They become shallow maxima and minima, and eventually disappear, if the layer is thick and/or the phase deviation is greater than one radian.

As a measure of the intensity of the scintillation phenomenon the quantity $S$ defined below is often used:

$$S=\frac{\{\langle(A^2-\langle A^2\rangle)^2\rangle\}^{1/2}}{\langle A^2\rangle} \tag{20}$$

It will be seen that $S$ is the r.m.s. deviation of the intensity $A^2$ from its mean value $\langle A^2\rangle$ divided by the mean intensity. It is adopted partly because it is easy to calculate, and partly because the intensity $A^2$ is usually measured in practice, especially in studies of radio star scintillations. $S$ is sometimes referred to as the ' scintillation index ', although different workers have used different measures of this index.

The scintillation index depends on the total energy of the spatial spectrum, i.e. on the area under the lower curve in fig. 4. As the wave propagates away from the phase screen, the zeros in the spectrum move to the left, and the area under the spectrum increases. Fig. 5 shows how $S$ increases with distance for various values of $\phi_0$. Provided $\phi_0<1$, the variation is of the form expected from the qualitative argument given earlier. That is, $S$ starts from zero at $z=0$ and increases monotonically, eventually becoming constant for large values of $z$. The larger $\phi_0$ the larger the value of $S$ for any given value of $z$. When $\phi_0>1$ a new feature appears. A peak in $S$ now occurs at a particular distance from the screen, and beyond this point, $S$ decreases as $z$ increases. The theory is more complicated in this case as the problem is non-linear and cannot be treated by the method so far described. We will return to this ' focusing ' phenomenon shortly.

It is not a difficult matter to extend the diffraction type of treatment (assuming $\phi_0<1$) to include the way the scintillation depth would be expected to vary as a function of the zenith angle at which the radio star or satellite is located (Briggs and Parkin 1963). The results are much as might be anticipated on simple arguments. When a source is at a large value of zenith angle, the length $L$ of eqn. (8) is increased because of the oblique path through the irregular ionosphere, and this leads to a larger value of $\phi_0$. Also, since the irregularities are far away from the observer, the wave has ample distance in which to build up the amplitude fluctuations to a large value as given by the curves of fig. 5. Both effects lead to an increase of $S$ with zenith angle, which can be worked out for particular values of the parameters—irregularity scale $r_0$, height, and radio wavelength. Additional factors enter if the irregularities are anisotropic and elongated along the Earth's magnetic field, as is often believed to be the case. The integral in eqn. (6) then leads to the largest value of $\phi_0$ when the irregularities are viewed along their longest axis, because $\Delta N$ then remains coherent over the greatest possible distance. The scintillation depth therefore has a maximum when the radio source (star or satellite) is viewed in the direction of the *magnetic zenith*—the point in the sky at which the observer is looking directly along the magnetic field lines.

Both the zenith angle effects and the magnetic alignment effects are shown in the theoretical curves of fig. 6. The curves are for different values of the axial of the irregularities ($\alpha$). Isotropic irregularities have an axial ratio of

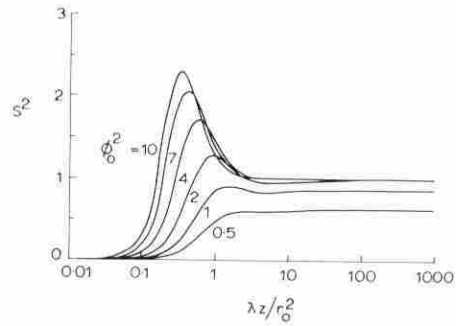


Fig. 5. This diagram shows how the scintillation depth $S$ increases with distance $z$ beyond a phase screen. The numbers on the curves are the value of $\phi_0{}^2$ the mean square phase deviation. A normalized distance variable $\lambda z/r_0{}^2$ is used, where $r_0$ is the scale parameter of the phase fluctuations. When $\phi_0$ is of the order of 1 rad or greater the curves show a peak at a particular value of $z$; this is a focusing phenomenon (see text). (After Bramley and Young 1967.)

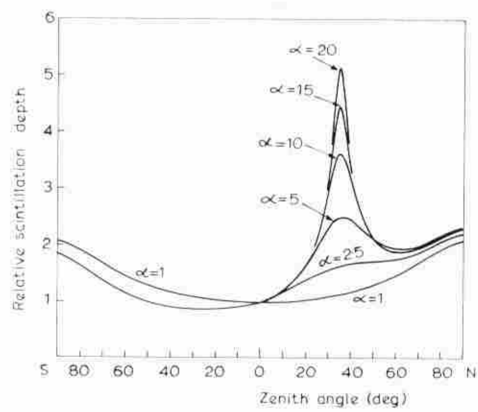


Fig. 6. These curves show how the relative scintillation depth $S$ would be expected to vary in the magnetic meridian plane if scintillations of the satellite Explorer VII were observed at Brisbane, using a frequency of 20 MHz. The curves are for different values of the axial ratio $\alpha$ of the ionospheric irregularities, assumed to be at a height of 300 km. The peak near a zenith angle of about 40°N occurs when the radio waves travel along the direction of the Earth's magnetic field at 300 km, this being the assumed direction of the 'long' axis of the irregularities. (After Briggs and Parkin 1963.)

unity and show the zenith angle effect only. These results are based on the assumption of a Gaussian spectrum for the irregularities and also assume $\phi_0 < 1$.

The linear treatment outlined so far can be expected to be a reasonable approximation for high radio frequencies, for which the refractive index is close to unity, and the r.m.s. phase deviations $\phi_0$ will be small. However, we are often concerned with experimental situations where $\phi_0$ is large, either because the frequency is low, or because the path length through the ionosphere is very large, as when a radio star or satellite is observed at small angles of elevation. What can we say about the diffraction effects when $\phi_0$ is larger than one radian?

As indicated already, this is a difficult non-linear problem. Analytical approaches have been confined to various limiting cases and are highly mathematical; some cases of interest cannot be solved analytically at all. We will not review this work, but outline instead an interesting numerical approach used by Buckley (1975), which has the advantage of giving results which are easy to understand physically.

Buckley's approach is as follows. First a single sample or ' realization ' of a section of a random function which is to represent the phase variation of the wavefront emerging from the ionosphere is constructed in a digital computer. This function can be given any desired statistical properties. For example, in fig. 7 the random curve at the top of the diagram is a short section of a

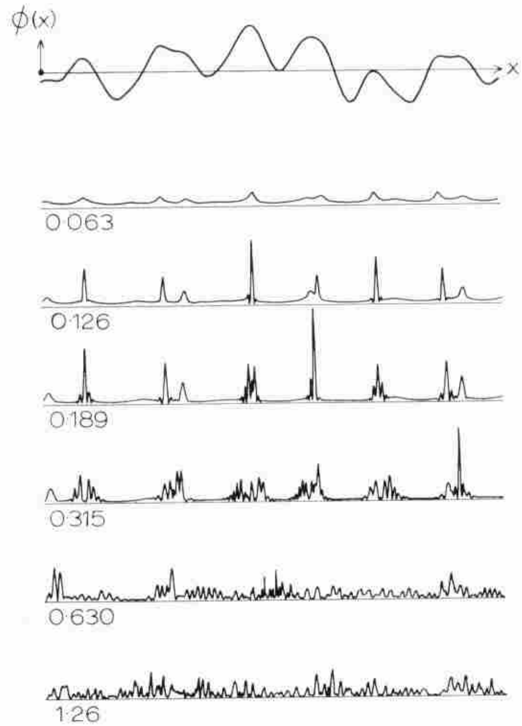


Fig. 7. Diffraction patterns calculated numerically at different distances $z$ from a random phase screen. The screen has a Gaussian auto-correlation function with scale parameter $r_0$. The curves are labelled by the normalized distance parameter $\lambda z/r_0^2$. (After Buckley 1975.)

random function which has a Gaussian auto-correlation function, as given by eqn. (18), and a normal distribution of phase about the mean, with an r.m.s. phase variation $\phi_0$ of 10 radians. The wave field in any plane below this phase screen is then calculated by ordinary diffraction theory. That is, the field at any point is calculated (in principle) as the sum of Huygens wavelets coming from each element of the phase screen, starting with the appropriate initial phase. The calculation of each of the 'diffraction patterns' illustrated is tedious, but, with a computer, not a difficult problem.

If statistical results are required, the whole process is now repeated many times, using a fresh sample or 'realization' of the random screen as a starting point. Then, by combining all the results, it is possible to obtain, for example, the probability distribution of amplitude, and the scale of the pattern and to see how these change as the wave propagates beyond the screen.

We will not discuss these statistical results but it is of interest to examine qualitatively some of the features involved. We can do this by considering the results shown in fig. 7. Starting very close to the phase screen, we see that the amplitude fluctuations are small, and there is one peak for each 'concave' section of the phase function. In this region geometrical optics is valid, and we can easily explain these fluctuations by the divergence and convergence of rays, refracted by the phase irregularities which act as irregular 'lenses' or 'prisms'. Further from the screen we notice large isolated peaks of amplitude, with a complex structure. In this region, the phase 'lenses' tend to focus the wave, though they do not all focus in quite the same plane due to their differing curvatures. The focal distance beyond the screen can be predicted roughly by geometrical optics, but the fine structure inside each focusing peak is a diffraction effect. The focusing effects account for the fact that the scintillation depth $S$ has a maximum at a certain distance from the screen when $\phi_0 > 1$ (see fig. 5). At very large distances, well beyond the focusing plane, the fluctuations of amplitude become irregular and of much smaller spatial scale than the phase fluctuations from which they originate. In fact the scale is found to be $r_0/\phi_0$, where $r_0$ is the scale of the phase screen (cf. eqn. (18)), and $\phi_0$ the r.m.s. phase deviation. This is a general result which can be demonstrated in other ways. It shows that we must be cautious in deriving scales of ionospheric irregularities from the observed amplitude fluctuations, unless we can be certain that the phase variation is less than one radian. If the phase deviation is large, the irregularities in the ionosphere will be larger than the scales we derive from observations of diffraction patterns.

## 4. Observations

Ionospheric scintillation is essentially a sporadic phenomenon; sometimes it is present and sometimes it is not. Nevertheless we would expect any such phenomenon to show characteristic patterns of occurrence if examined statistically over a long enough period. For example, its occurrence may vary diurnally and seasonally, and may be a function of latitude. It may also vary with the sunspot cycle, and perhaps be related to the occurrence of magnetic storms, aurorae and other phenomena. Extensive observations over long periods are needed to establish facts of this kind. Such studies are perhaps the kind of science which Rutherford would have dismissed as 'stamp collecting'. But such work is needed, both in order to provide facts against which theories of origin of the irregularities can be tested, and also because the phenomenon is of practical importance. Navigation and communication satellite signals can be affected by ionospheric scintillation. Radioastronomical observations can be upset, not only by amplitude scintillations, but also by associated phase fluctuations, which cause positional errors. With sufficient knowledge it might be possible to determine optimum sites for radiotelescopes, and to decide the best times at which to make observations.

It has been known from the times of the earliest observations, that ionospheric scintillation is mainly a night-time phenomenon, and the association with 'spread-F' echoes suggested that the irregularities were situated mainly in the F region of the ionosphere. Weaker daytime scintillation is occasionally observed, and could have its origin in the lower 'E region' around 100 km (methods for determining the heights of the irregularities with more precision will be described later). There is no marked seasonal variation of occurrence.

Geographically, two regions of increased scintillation occurrence are generally recognized. One is centred on the geomagnetic equator and extends about $\pm 20°$ in latitude on either side. This scintillation belt expands towards mid-latitudes with declining solar activity. There is a negative correlation with magnetic activity in this zone, i.e. scintillation tends to disappear when magnetic storms occur. The other region of increased occurrence is in polar regions at latitudes greater than about 65°. The boundary of this zone moves to lower latitudes around midnight and during increased magnetic activity.

The mid-latitude region, 20° to 65°, has reduced, but still significant scintillation occurrence, and as most observers work in this zone, has been most extensively studied. Correlation with magnetic activity is positive, and scintillation is more likely to occur at sunspot maximum than at sunspot minimum (Briggs 1964).

Apart from the statistics of occurrence, information is needed about the shapes and sizes of the ionospheric irregularities, and the heights at which they occur. There are several methods of observation, making use of radio stars and satellites, which can give this information, using as a basis the diffraction theory outlined in Section 3.

The most direct method for determining sizes and shapes is to observe the diffraction pattern formed over the ground by the use of spaced receivers. With sufficient receivers spread out over the ground the form of the pattern at one instant of time could be determined. In practice, such an experiment would be prohibitively expensive. Instead, three spaced receivers are usually employed, and scintillations are recorded simultaneously at the three points. A receiver separation of the order of 1 km to 5 km is required. The required statistical information about the diffraction pattern is then determined from the degree of correlation of the scintillations at the three points, in a manner which will now be explained.

The situation is illustrated in fig. 8 (*a*), where a plane wave from a radio star, after passing through the ionospheric irregularities at $I$ forms an amplitude pattern $A(x)$ over the ground. Suppose for the moment that this pattern is sampled by two receivers, $R_1$ and $R_2$ at a separation $d_{12}$, and that the ionosphere is moving with a velocity $V$ in the direction from $R_1$ to $R_2$. Then the pattern $A(x)$ will also move over the ground with velocity $V$, and the variations recorded by $R_1$ will also be recorded by $R_2$ but after a time delay

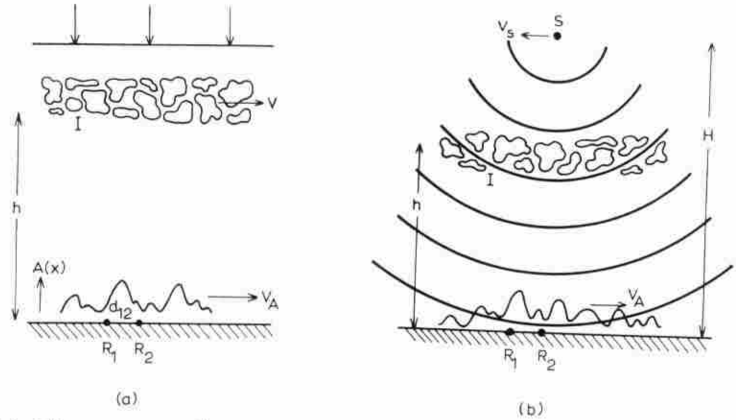


Fig. 8. (*a*) Plane waves from a radio star produce a pattern over the ground which moves with a speed equal to that of the ionospheric irregularities ($V_A = V$). (*b*) Spherical waves from a satellite produce a pattern which has a different scale and a different speed from the irregularities which produce it.

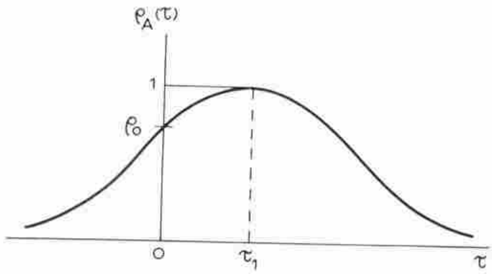


Fig. 9. Cross-correlation function between the scintillations recorded at two points $R_1$ and $R_2$ on the ground. The time delay for maximum correlation is related to the drift velocity of the pattern over the ground.

$\tau_1 = d_{12}/V$. Now suppose the correlation $\rho_A(\tau)$ between the amplitude fluctuations recorded by $R_1$ and $R_2$ is calculated for different values of time shift $\tau$ between the records. We would expect to obtain a curve like that shown in fig. 9. For a time shift $\tau_1$ the records match up exactly, and the correlation has a maximum value of unity. Such a curve is called a 'cross-correlation function'. In the example illustrated, the velocity $V$ could be found from the value of $\tau_1$, $d_{12}$ being known. To determine the scale of the pattern we can use the value of correlation $\rho_0$, say, calculated for $\tau = 0$. This is the actual correlation between the records obtained at points separated by a distance $d_{12}$ and must be related to the scale of the pattern. However, to obtain a definite scale parameter, we must make some assumption about the form of the auto-correlation function of $A(x)$. For example, if we assume that $A(x)$ has a Gaussian auto-correlation function with scale parameter $l$, so that

$$\rho_A(r) = \exp(-r^2/l^2) \tag{21}$$

where $\rho_A(r)$ is the correlation between two receivers separated by $r$, then, since $\rho_A(r) = \rho_0$ when $r = d_{12}$, we can find $l$ by substitution of these values into eqn. (21). It is then usual to assume that $l$ is a useful measure of the scale of the ionospheric irregularities. As pointed out earlier, this is only so if $\phi_0 < 1$, and it also ignores the modifications to the spectrum which occur through the effect of the Fresnel filter (fig. 4) even if this condition is satisfied. There are indications that these assumptions, as well as the assumption of Gaussian spectra and correlation functions, may require re-examination in future work. However, for the present purpose we will accept the sizes obtained essentially in the manner outlined.

The real problem is, of course, two-dimensional, and so three receivers are used, three cross-correlations are calculated between the receivers taken in pairs, and then, by a slightly more elaborate calculation, the size of the irregularities in two planes can be determined. In this way it has been shown that the irregularities tend to be elongated with their long axes along the direction of the Earth's magnetic field in the F region and with typical dimensions 1 km × 5 km. The elongation is especially marked in the equatorial scintillation zone. Also, with three receivers, we have the possibility of determining the drift velocity $V$ in both magnitude and direction.

It is interesting to see what information can be obtained if the same basic experimental technique is applied to records of scintillations of signals from an orbiting satellite. The factors involved here are somewhat different, as illustrated in fig. 8 (*b*). The satellite is moving with velocity $V_S$ at a height $H$ above the ground. The velocity of the satellite is always very much greater than that of the ionospheric irregularities. The motion of the latter can be neglected, and the motion of the amplitude pattern is therefore due entirely to the motion of the satellite. The following facts can easily be established by applying diffraction theory:

(i) If $r_0$ is the scale of the ionospheric irregularities, the scale of the pattern on the ground is $r_0' = r_0 H/(H-h)$.

(ii) If $V_A$ is the velocity of the pattern over the ground, then $V_A = -V_S h/(H-h)$.

It will be seen that these results are just of the form which would apply if the ionospheric irregularities were casting geometrical 'shadows' on to the ground. This reflects the fact that diffraction theory must always go over to geometrical shadow optics as a special case.

By applying correlation analysis to records obtained at three spaced receivers, $r_0'$ and $V_A$ can be found. Then $r_0$ can be found using (i) and $h$ can be found using (ii). (Note that $H$ and $V_S$ are known if the satellite orbit is known). Thus, when compared with the radio star method, it will be seen that the satellite method gives similar information about scales, no information about ionospheric velocities, but additional, and very useful information about the heights at which the irregularities occur.

Observations of this kind have shown that the irregularities occur most frequently in the height range 200 to 800 km. There is some evidence that the height decreases during magnetic storms. Often the irregularities are above the height of maximum electron density $N_m$, and therefore not observable as spread-F echoes using ground based sounders. Ionosphere sounders carried or orbiting satellites have shown that spread-F echoes occur on the 'topside' of the ionosphere and are related to scintillation phenomena. Occasionally, especially when daytime scintillation is observed, heights as low as 100 km are indicated. (A useful collection of papers on the use of satellite transmissions was presented at a Symposium held in Boston in 1969, the proceedings of which have been edited by Aarons 1970.)

Zenith angle effects at a single station have been used to study irregularities sizes and shapes. The basic technique is to compare the observed variations with theoretical predictions, and then to find which parameters give the best fit with the observations. The peak in scintillation depth when a source of radio waves is viewed along the Earth's magnetic field has been observed. Sizes, shapes, and heights generally agree well with those obtained by other methods, though some observations suggest that magnetic field aligned 'sheets' as well as 'columns' are needed to explain all the observations (Singleton 1970). Another check on the theory has been obtained by demonstrating the existence of 'fringes' in the power spectrum of scintillation due to the Fresnel filter effect (Rufenach 1972, Singleton 1974).

When the r.m.s. phase deviation $\phi_0$ becomes large, the theory predicts focusing phenomena. Large values of $\phi_0$ are most likely to be encountered on low radio frequencies, and when a source is viewed at a large zenith angle so that the path length $L$ through the ionosphere is large. Since focusing is basically a phenomenon of geometrical optics, the focusing peaks, if they occur, should be broad-band, i.e. they should occur simultaneously over a wide band of radio frequencies. Consequently, focusing effects are most easily detected and studied if observations are made over a range of frequencies rather than at a single frequency. This is possible using a radio star as a source. A radio star is a wide-band source of 'noise', and it is possible to sweep a radio receiver rapidly backwards and forwards over a band of frequencies, recording the intensity of the source all the time. The observations are usually recorded on a moving film, in such a way that time runs along the film and frequency across the film. For obvious reasons, an instrument of this type is called a 'radio spectrograph'. The first records of this type were obtained by Wild and Roberts (1956), observing the strong radio star Cygnus A. Fig. 10 shows an example of focusing type scintillations observed by these workers. The record runs from 40 MHz to 70 MHz, and the white regions on the record indicate high intensity. Several 'ridges' of high intensity extend over the frequency band, some showing fringe structures within them.

It is instructive to compare these with the theoretical results shown in fig. 7. It is important to note that the occurrence of these apparently 'isolated' focusing effects does not necessarily imply that isolated clouds of electrons occur in the ionosphere to provide the necessary 'lenses'. The theoretical curves of fig. 7 were produced from a random phase screen, and yet reproduce the focusing phenomena very well. Also, in the case of the record shown in fig. 10, irregularities with scales of a few kilometres would explain the observations, if $\phi_0$ is large enough. On the other hand, other workers observing wide-band focusing effects have found it necessary to postulate somewhat larger ionospheric structures, of the order of 20 km in extent, to explain their observations (Warwick 1964).

Communication satellites operate on very high frequencies where ionospheric scintillation would be expected to be small or non-existent. However, recent observations in the GHz band ($\sim 10^9$ Hz) have shown that near the equator, and to a lesser extent in polar regions, scintillations can be significant (Wernick


Fig. 10. Radio spectrogram showing scintillations of Cygnus A recorded simultaneously over the frequency band from 40–70 MHz. Some bright regions (which correspond to strong signals) extend over the whole of this frequency range and show a fine internal fringe structure. (After Wild and Roberts 1956).

and Liu 1974). This is clearly of practical importance. The irregularities which can cause such scintillation must have a scale of the order of 10–100 m; scales of kilometres or greater could not be effective in producing amplitude scintillations, because of the effect of the 'Fresnel filter' (cf. fig. 4). This raises the question as to the true spectrum of the ionospheric irregularities. The assumption of a Gaussian spectrum, used in most theoretical work, has been made mainly for mathematical convenience. Observations of large scale structures by the radio spectrograph technique, and small scale structures by the observations of scintillations on GHz frequencies suggests that the spectrum may really be very broad. Perhaps there is more experimental 'selection' than has commonly been supposed, and the usually quoted typical scales of a few kilometres may arise through the operation of the Fresnel filter.

This brings us to the last topic to be discussed in this article—the question of direct *in situ* observations by satellite probes. If a probe carried on a satellite could measure the fluctuations of electron density through which the satellite was travelling in its orbital motion, the true spectrum of electron density fluctuations could be calculated, free from any experimental bias. This is provided, of course, that the electron density probe has a sufficiently rapid frequency response (which is not always the case). The Ogo 6 satellite carried a retarding potential analyser, capable of detecting changes in ion concentration as small as 0·01 per cent. The frequency response was such that spatial scales of irregularities as small as 35 m could be resolved. The spectrum of the irregularities was calculated for the range of scales between 70 m and 7 km, and was found to be such that the intensity of irregularities of scale $R$ was proportional to $R^n$ (i.e. of power law form) with the exponent $n$ very close to the value 2. Or, in terms of spatial frequency, irregularities are present with an intensity inversely proportional to the square of their spatial frequency (Dyson *et al.* 1974). This is a very interesting result. Recent scintillation studies have also indicated that a power law spectrum may be appropriate (Rufenach 1972). As a final comment on the spectrum, we may just note the very remarkable fact the irregularity scales as small as 70 m have been shown to exist in a region of the atmosphere where the mean free path (of neutral particles) is of the order of 10 km.

## 5. Theories of the origin of the ionospheric irregularities

If we are interested primarily in the physics of the upper atmosphere, the fundamental question is: what causes the irregularities of electron density?

So many theories have been proposed that a separate review would be required to discuss them. We will just list the possibilities; for a detailed though not completely up to date review, the reader may refer to Herman (1966).

In the equatorial region, most mechanisms proposed rely on either plasma instabilities, or transfer of irregularities from below along magnetic field lines. For higher latitudes, in addition to these possibilities, an irregular downward flow of heat from the protonosphere into the night ionosphere has been proposed. At still higher latitudes, near the auroral zones, ionization from the influx of energetic charged particles is an obvious possibility. Transfer of irregularities down the magnetic field lines from convective processes in the magnetosphere has also been proposed.

To decide between these and other theories is one of the outstanding problems of ionospheric physics at the present time.

## References

Aarons, J. (editor), 1970, *Symposium on the application of atmospheric studies to satellite transmissions, Boston, September* 1969, *Radio Science*, **5,** 867.
Bramley, E. N., and Young, M., 1967, *Proc. I.E.E.*, **114,** 553.
Briggs, B. H., 1964, *J. Atmos. Terr. Phys.*, **26,** 1.
Briggs, B. H., and Parkin, I. A., 1963, *J. Atmos. Terr. Phys.*, **25,** 339.
Buckley, R., 1975, *J. Atmos. Terr. Phys.* In the press.
Dyson, P. L., McClure, J. P., and Hanson, W. B., 1974, *J. Geophys. Res.*, **79,**1497.
Herman, J. R., 1966, *Rev. Geophys.*, **4,** 255.
Little, C. G., and Lovell, A. C. B., 1950, *Nature*, **165,** 423.
Ratcliffe, J. A., 1956, *Rep. Prog. Phys.*, **19,** 188.
Rishbeth, H., 1973, *Contemp. Phys.*, **14,** 229.
Rufenach, C. L., 1972, *J. Geophys. Res.*, **77,** 4761.
Singleton, D. G., 1970, *J. Atmos. Terr. Phys.*, **32,** 789.
Singleton, D. G., 1974, *J. Atmos. Terr. Phys.*, **36,** 113.
Smith, F. G., 1950, *Nature*, **165,** 422.
Warwick, J. W., 1964, *Radioscience* D, **68,** 179.
Wernick, A. W., and Liu, C. H., 1974, *J. Atmos. Terr. Phys.*, **36,** 871.
Wild, J. P., and Roberts, J. A., 1956, *J. Atmos. Terr. Phys.*, **8,** 55.

*The Author:*

Dr. B. H. Briggs graduated at the University of Cambridge, and after some years of research on radar during the war, returned to complete Part II of the Tripos in 1947. His Ph.D. work was carried out in the Radio Section of the Cavendish Laboratory under J. A. Ratcliffe, and involved a study of irregularities and drift motions in the ionosphere. He continued with research in the same general fields at the Cavendish Laboratory until 1962, when he went to the Physics Department of the University of Adelaide, where he now has the post of Reader in Physics. His present research interests still include ionospheric irregularities and movements, but have broadened to include also interplanetary scintillations, and the propagation of light waves and sound waves in the lower atmosphere.