

Ionosphere and its Influence on Radio Communications

R S Dabas



R S Dabas is at the Radio Science Division of National Physical Laboratory, New Delhi and is responsible for characterization of ionospheric media for radio communication applications. His main area of research is the study of equatorial and low latitude ionosphere i.e. its properties and dynamics.

1. Discovery of Ionosphere

In 1882, the Scottish physicist Balfour Stewart suggested that the continuous but minor variations in Earth's magnetic field might be caused by the presence of a layer of air capable of conducting electricity in the upper atmosphere. Movement of this layer in the terrestrial field could produce electric currents by dynamo effect; these currents could, in turn, generate magnetic fields, which would be superimposed upon the normal magnetic field observed at the Earth's surface. Stewart's suggestion attracted little interest until after December 1901 when the Italian radio pioneer G Marconi succeeded in transmitting 'wireless' signals over a distance of about 2000 miles across the Atlantic Ocean. Physicists had previously regarded this feat as impossible because radio waves, like light, travel in straight lines; detection on Earth's surface should thus not be possible beyond the line of sight. The explanation, proposed independently in 1902 by Oliver Heaviside in England and Arthur E Kennelly in the United States, for Marconi's unexpected achievement was that there exists an electrically conducting layer in the atmosphere at a height of about 80 km. Such a layer would reflect radio waves, like a passive repeater station, and return them to Earth at a considerable distance beyond the horizon.

The concept of the conducting Kennelly–Heaviside layer, as it was called, was generally accepted, but for more than two decades there was no direct evidence of the manner in which it functioned. In 1924, Edward V Appleton and M A F Barnett in England, using directional antennas were able to relate the angles with the vertical made by transmitted and received continuous radio wave signals. In this way they showed that the signal received at a distance was indeed a reflection of the trans-



mitted wave, high in the atmosphere.

The following year, G Breit and M A Tuve in the United States devised a technique for determining the height of the reflecting region. They transmitted a short pulse of radio waves upward and, by means of an oscilloscope, determined the time taken for the pulse to be reflected back to the receiver a few miles away on Earth's surface. Since the radio waves travel with the speed of light, the height of the reflecting layer in the atmosphere can be calculated from the time delay. The altitude of Kennelly-Heaviside layer was found to be some 80-100 km. Subsequently, in the course of investigations with radio waves of shorter wavelength, (i.e. higher frequency) Appleton discovered another reflecting region, at a height of roughly 200 to 400 km, which was at one time known as the Appleton layer. To simplify the identification of the conducting layers in the atmosphere, Appleton proposed the nomenclature now universally employed; he called the lower of the two layers as the E layer and the upper one as the F layer. In the daytime the latter is split into two parts designated by the symbols F1 and F2, respectively. A layer below the E layer discovered later, became known as the D layer.

The ability of the conducting layers to reflect radio waves is dependent upon the presence of electrically charged particles, especially electrons; in other words, the air in these layers is ionized. R Watson Watt, a British engineer who contributed greatly to the development of radar, therefore proposed the name ionosphere for the whole ionized region in the atmosphere. The ionization is the result of the absorption of radiations of various kinds from the Sun, and it is a combination of air density and composition and radiation intensity that leads to the formation of the ionosphere region of the atmosphere extending upward from an altitude of about 60 km.

Because of its great importance in long distance radio communication, as well as on account of its intrinsic scientific interest, the ionosphere has been the subject of extensive study during the past 50 years. Even though F2 layer electron density is of the

The ionization is the result of the absorption of radiations of various kinds from the Sun, and it is a combination of air density and composition and radiation intensity that leads to the formation of the ionosphere region of the atmosphere extending upward from an altitude of about 60 km.



Under suitable conditions the ionosphere reflects long, medium and short radio waves and is used for radio communication and navigation.

order of one thousandth of the neutral density and the ionosphere is regarded as a weakly ionized plasma embedded in the earth's magnetic field, it is of great scientific interest and has many applications. Under suitable conditions the ionosphere reflects long, medium and short radio waves and is used for radio communication and navigation.

All regions of our atmosphere are defined in terms of ionization, temperature and composition and play an important role in the major areas of human endeavor like communication, navigation, and environmental control. In terms of ionization, the region which extends from 60 to several thousand kilometers above the earth is called the ionosphere. Different constituent gases are partially ionized in it. A typical daytime electron density distribution as well as the neutral gaseous temperature distribution in the upper atmosphere, is shown in *Figure 1*.

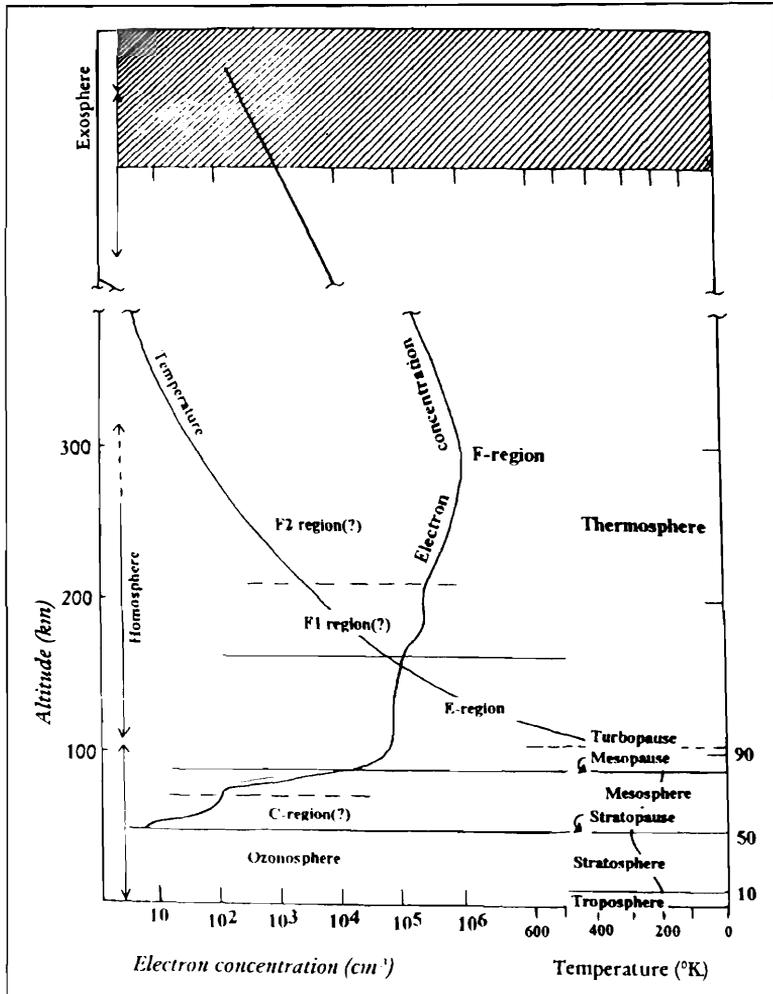
2. Indian Contribution

Around 1930 ionospheric research began with a bang in S K Mitra's laboratory in the University of Calcutta. A few years after the discovery of the ionosphere by Appleton and Barnett using a medium wave transmitter (made available by the Calcutta station of the Indian State Broadcasting Service) and installing a receiving system some 75 km away, the first experimental evidence of the E-region of the ionosphere was obtained in India by Mitra's group in Calcutta University in 1930. The results were published in *Nature*. After that a series of papers came in quick succession relating to the behavior of the ionospheric layers over Calcutta. It was remarkable that the quality of ionospheric research achieved at that time with simple equipment was so excellent. E and F layer heights (then usually called Kennelly-Heaviside layer and Appleton layer, respectively) were measured to be, on an average, 90 and 230 km over Calcutta.

The first experimental evidence of the E-region of the ionosphere was obtained in India by Mitra's group in Calcutta University in 1930.

At that time scientists were looking for additional layers in the ionosphere and there was indirect evidence of a layer below the E layer at levels where collisions between electrons and ions and

Figure 1. A schematic representation of atmosphere regions.



gaseous molecules are frequent and so absorption of medium radio waves is large. In 1935 Mitra and Shyam announced the reception of regular echoes from low heights (55 km) and called that D layer. S K Mitra's research activities ranged over many topics: considerations of the 'fringe' of the atmosphere, distribution of atmospheric constituents, night airglow; suggestion of the mechanism behind the geomagnetic control of the ionosphere discovered by Appleton, thunderstorms and sporadic E ionization.

One of the most important works of Mitra was his explanation for the Appleton ionization anomaly. He pointed out that the

One of the most important works of Mitra was his explanation for the Appleton ionization anomaly.

The total population of ions and electrons depends on the competition between ionization process, ion and electron diffusion, geomagnetic effects and mass motions. These processes superimposed upon the chemical and physical processes determine the structure of the ionosphere.

peaks in electron density at F layer heights observed by Appleton on both sides of the geomagnetic equator occurred around magnetic dip values around 28° and is due to guided transfer of ionization from dip equator along magnetic field lines to low latitudes on both sides of the equator. Mitra set up the Institute of Radiophysics and Electronics, which has grown into a National Centre of Excellence. The Institute has produced some of India's most distinguished radio scientists.

3. Formation of Ionospheric Layers and their Characteristics

With the possible exception of a contribution from cosmic radiation in the formation of the lowest layer, the ionized regions in the upper atmosphere owe their existence, directly or indirectly, to the ionizing power of the solar radiation. As this radiation penetrates the earth's atmosphere, it produces ions and electrons. The nature of the ions that are produced depends upon the energy of the radiation and composition of the atmosphere at the altitude at which ionization occurs. Immediately following the production of an ion species, a number of processes occur that changes the species of ions and decrease the number of electrons and ions that exist. The total population of ions and electrons depends on the competition between ionization process, ion and electron diffusion, geomagnetic effects and mass motions. These processes superimposed upon the chemical and physical processes determine the structure of the ionosphere.

D region is the lowest lying region and hence is produced by the most penetrating of the ionizing radiation that impinge upon the earth. A secondary electron producing source is the visible and ultraviolet solar emissions, which can provide the energy necessary to release weakly held electrons from negative ions.

In the E region, the ionization of molecular species is dominant and electron-ion losses are solely due to dissociative recombination processes. The electron density in this region is determined by the photo-equilibrium equation.



The unique characteristic of the F1 layer is that the principal ion produced is atomic, whereas the principal electron loss process is dissociative recombination with a molecular ion, and that the ion interchange is also a process in this region.

Under quiet daytime conditions, below an altitude of about 200 km, all charged constituents are in or nearly in a condition of photochemical equilibrium. In fact, photochemical processes determine the structure of the ionosphere at these heights at all times of the day and night. At great heights (above 250 km), this is not the case. At these heights the chemical loss rate of the ions is comparable to their diffusion rate though the neutral gas and transport mechanism begin to influence the ion and electron number densities. With ascending altitude, diffusion becomes more and more important, finally dominating the picture at greater heights. The competition of chemical loss (recombination) and diffusion leads to the formation of a peak in the number density profile of the ions and electrons, which is the well-known F2 peak. The geomagnetic field plays an important role in determining the ionization distribution since the transport of ionization with respect to the field is an important process in the F2 region.

The ionization density in each of these layers has a peak value at a certain altitude above and below which it decreases. The D region present only during the daytime normally has a peak density around 90 km, which may drop to around 60 km due to enhanced solar X ray flux. The E region containing both the normal E layers and at times patches of sporadic E (E_s), has a peak density around 120 km. Following sunset, the electron density in this region decreases by a factor of 10 or more within a very short time before attaining a nighttime equilibrium density. The F region is a combination of two different regions, namely F1 and F2. The F1 region, which disappears after sunset, has a peak density around 200 km and is more pronounced in summer than in winter and at high sunspot numbers. In the F2 region, the altitude of the peak density occurs at about 300 km in the daytime and at higher altitudes in the night.

The geomagnetic field plays an important role in determining the ionization distribution since the transport of ionization with respect to the field is an important process in the F2 region.



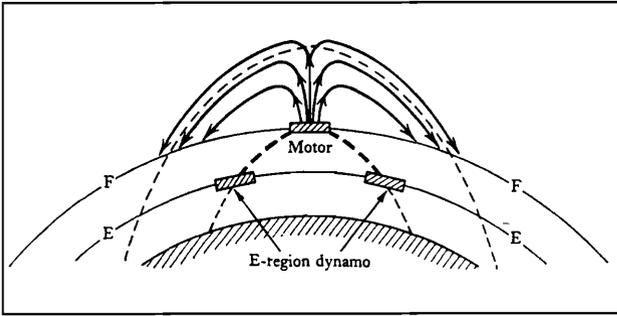


Figure 2. *The F region geomagnetic anomaly. Near the equator the electric fields of the atmospheric dynamo in the E layer are conveyed upwards along geomagnetic field lines to 'motor' in the F layer where they produce an upward movement of plasma during the day. The raised plasma then diffuses down along magnetic field lines to produce enhanced concentration at places on each side of the equator, and decreased concentration at the equator.*

Shortly after sunset, the absolute density near the peak of the F region often increases due to plasma transport processes before decreasing to a nighttime value and drifting downwards. The peak densities of the ionospheric layers vary with time of the day, season, latitude, and solar activity and with a variety of irregular and random disturbances. The 'topside ionosphere' is the name given to the rest of the ionosphere above the F layer peak. In a simple model, the density of the topside ionosphere decreases exponentially with height until the ionization density is below detectable levels.

Near the geomagnetic equator, the earth's magnetic field, B (in North-South direction) is horizontal and there is east-west electric field due to the dynamo effect by atmospheric motions at E region heights (see *Figure 2*). The electric field is eastward during the day and westward during the night. As a consequence, ionospheric plasma from equatorial F region moves upward and then diffuses downward along sloping magnetic field lines to low latitudes on both sides of equator. The electron concentration is thus depleted on the magnetic equator and enhanced in two regions, one on each side as shown in *Figure 3*. The phenomenon is known as equatorial ionization anomaly. The distribution of ionospheric plasma is also affected by solar and magnetic disturbances like the occurrences of solar flares and magnetic storms. Varying solar activity essentially controls ionospheric parameters (like peak electron densities in different layers as well as their altitudes). There is a short (27 days) and long-term (11 years) periodicity in the solar activity. In general, it has been noticed that there is a close relationship between long term variation in sunspot numbers (black spots visible on the Sun's surface – photosphere) and critical frequency (electron density – see section 4.2.2) of the F2 region as seen from *Figure 4*. A highly significant aspect of the sunspots is their association with strong magnetic fields and they mostly occur in pairs on a line roughly parallel to the solar equator. It is believed that

strong local magnetic fields are the cause of sunspots. Sunspot number is a quantitative measure of sunspot activity and the total annual sunspot number exhibits definite variations from year to year. The average time, as seen from *Figure 4*, from minimum to maximum is about 5 years and from maximum to minimum 6 years with an 11-year solar cycle. Sunspot numbers are being generally used for predictions of ionospheric parameters for radio propagation and other applications. Thus, the richness in the variety of phenomena and characteristics applies to the study of radio propagation in the ionosphere as well as to the investigation of the physics of the ionosphere itself.

4. Radio Communication and the Ionosphere

The ionosphere extending from 60 km to about 1000 km altitude, plays a major role in aiding long distance HF (3 to 30 MHz) communications as well as in the deteriorating performance of satellite radio systems in the VHF (30 to 300 MHz), UHF (300 to 3000 MHz) and even higher frequency bands. Before discussing how ionosphere influences radio propagation, it is necessary to describe briefly some of the points like radio refractive index of ionosphere, interaction of radio waves with ionosphere, reflection of radio waves by the ionosphere, etc.

4.1 Radio Refractive Index of the Ionosphere

If the collisions between electrons and neutral molecules are neglected, we may, in a first approximation, treat ionized gas as

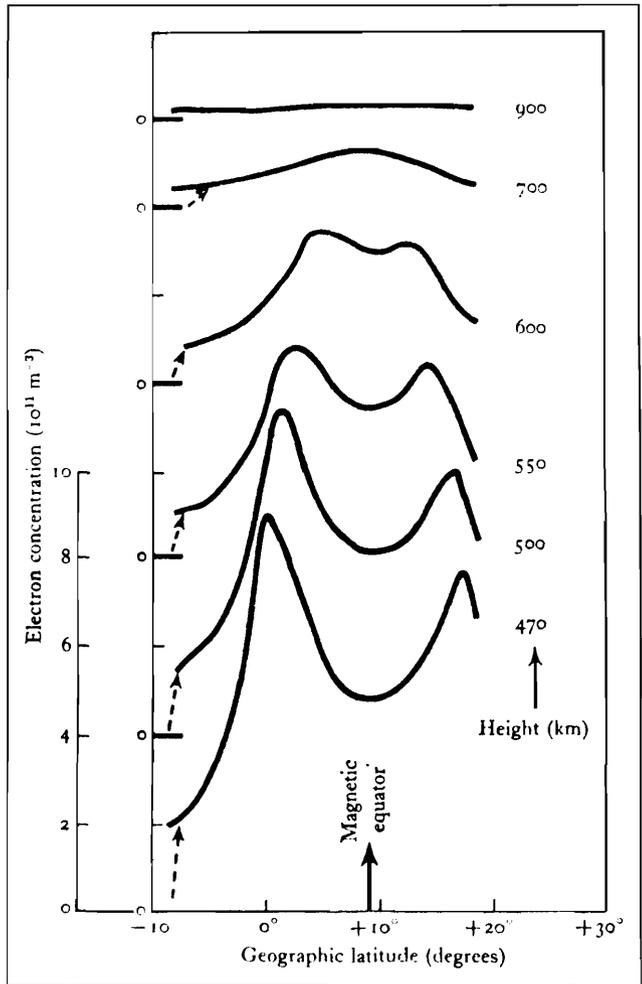


Figure 3. In the daytime topside ionosphere, the electron concentration is increased along geomagnetic lines of force that reaches up to a height of about 700 km. When the concentration is plotted, at lower heights, against latitude, there is thus a minimum on the magnetic equator and maximum on each side. The maxima are closer together at greater heights.

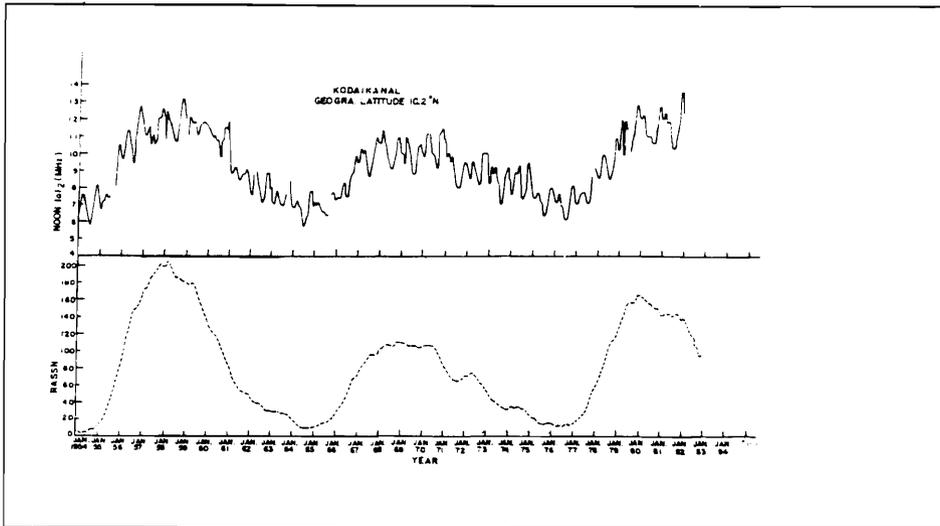


Figure 4. Sunspot numbers (RASSN) for the three cycles since 1954 (lower part) and noon foF2 (critical frequency of F2 layer) values for Kodaikanal. This shows the solar activity variations of the F region critical frequency/peak electron density.

a perfect dielectric whose radio refractive index is given by

$$n = (1 - 80.8 * 10^{-5} N/f^2)^{1/2}, \tag{1}$$

where n is the refractive index of the ionized gas, N is the number of electrons per cubic centimeter and f is the frequency of radio waves.

This expression for n is unusual in more than one respect. One, the absolute value of the refractive index is less than unity, whereas all solids, liquids and gases have a refractive index more than unity. Second, the refractive index of the ionospheric media is seen to be frequency-dependent. At a certain value of f it reduces to zero, while at the lower frequencies it becomes negative.

4.1.1 Phase and Group Velocity: From the field theory it is known that in a pure dielectric, radio waves are propagated at the velocity given by

$$U_p = c/n \text{ m/s}, \tag{2}$$

which is defined as the velocity of travel of an equiphase surface in the direction of the wave normal and called the *phase velocity* of the radio wave. In ionospheric media, n is less than unity, and so the phase velocity exceeds that of light in free space. Besides,

The media in which the velocity of waves varies with frequency are called *dispersive* because this results in dispersion of the composite signal, that is, the separation of its component frequencies.



the phase velocity varies with frequency. The media in which the velocity of waves varies with frequency are called *dispersive* because this results in dispersion of the composite signal, that is, the separation of its component frequencies.

The velocity of the composite wave is defined in one more way so as to take care of this distortion. This is done by introducing the concept of *group velocity* which is defined as the velocity of the crest of a group of interfering waves where the component wave trains have slightly different individual frequencies and phase frequencies. The *group velocity* is defined by the relation

$$U_g = c/nm/s. \quad (3)$$

Since in a homogeneous ionized gas n is always less than 1, the group velocity is also always lower than the velocity of light in free space. In ionospheric propagation, knowledge of phase velocity is important in dealing with refraction and reflection, because wave trajectory is decided by the phase velocity, whereas group velocity is important in measuring the group delay of radio waves reflected or propagated through ionosphere.

4.2 Interaction of Radio Waves with Ionosphere

An ionized region can affect the transmission of radio waves in at least two ways. Firstly, under suitable conditions, charged particles (electrons) can remove energy from an electromagnetic wave and thus attenuate the signal; in the limiting case, the energy of the wave can be absorbed completely. Second, a wave travelling from one place to another in which the electron density is different will undergo a change in its direction of propagation, due to the effect of free electrons on the velocity of a radio wave. These two effects of an ionized medium on electromagnetic radio waves are explained here in more detail.

4.2.1 Absorption of Radio Wave Energy passing through the Ionosphere: When a radio wave enters the ionosphere and encounters a significant concentration of free electrons, some energy of the radio wave is transferred to the electrons, which

In ionospheric propagation, knowledge of phase velocity is important in dealing with refraction and reflection, because wave trajectory is decided by the phase velocity, whereas group velocity is important in measuring the group delay of radio waves reflected or propagated through ionosphere.



Radio waves suffer little loss of energy or signal strength in passing through the higher levels of the ionosphere and most of the attenuation occurs at the lower part of the lower altitudes.

are thus set into oscillation at the same frequency as that of the radio waves. The electrons can lose some of this energy as a result of collisions with neutral particles (atoms or molecules) in the upper atmosphere and it results in attenuation of radio waves when passing through the ionosphere. But, if there are no such collisions, the oscillating electrons will reradiate electromagnetic waves at the same frequency and restore the original radio waves without loss. Appreciable attenuation of radio signals can occur in the D region (lowest region of the ionosphere) where collision of electrons with neutral particles is more due to the higher density of neutral molecules as compared to upper regions of the ionosphere like E and F regions. In addition, collision frequency of electrons with neutral particles is inversely proportional to the square of the frequency of the radio waves and hence attenuation is greater, the lower the frequency, i.e., the longer the wavelength. As a general rule, therefore, radio waves suffer little loss of energy or signal strength in passing through the higher levels of the ionosphere and most of the attenuation occurs at the lower part of the lower altitudes, i.e. in the D region of the ionosphere.

4.2.2 Refraction and Reflection of Radio Waves in the Ionosphere: It is evident from (1) that for a given frequency ' f ' the refractive index decreases in passing from a medium of lower to higher electron density. Hence a beam of radio waves going upward will be refracted downward, i.e. the angle with the vertical increases when the electron density increases and hence a radio wave will refract downward. The same conclusion can be drawn from the fact that the phase velocity of the wave increases with the increase in electron density whereas group velocity decreases.

Now consider the situation in which a radio wave of frequency f , enters the ionosphere in which the electron density increases with altitude. At first, N in (1) is small, and the velocity with which the wave travels is nearly same as velocity of light in free space. But as N increases with altitude, with constant f , it is seen that the phase velocity U_p must increase while the refractive



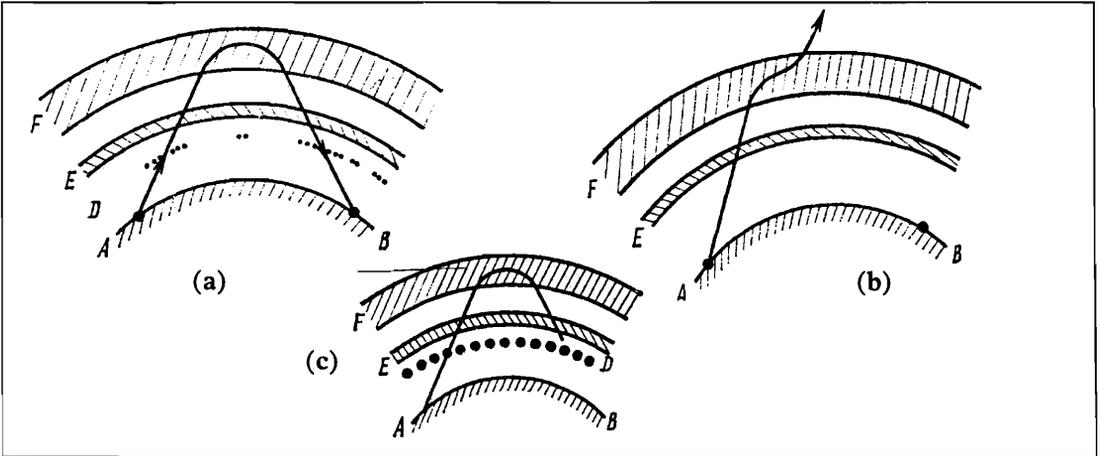


Figure 5. Ionospheric propagation: (a) normal propagation, (b) propagation with transmitter frequency exceeding the critical frequency, and (c) propagation under conditions of increased absorption in D and E regions.

index decreases. Eventually, when the quantity under the square root sign in (1) becomes zero, the refractive index is zero, i.e., when

$$1 - 80.8 N_c / f^2 = 0.$$

or

$$N_c = f^2 / (8.1 * 10^{-5}) = 1.2 * 10^4 f^2. \tag{4}$$

At this value of electron density, represented by N_c and called the *critical electron density*, the radio wave can no longer be propagated in the forward (upward) direction and so it is reflected back to Earth (see Figure 5a).

In (4), the frequency has been taken as fixed and the corresponding critical electron density calculated. The procedure can be reversed and the *critical (or penetrating) frequency* derived; this is the highest frequency that can be reflected by a given electron density or the lowest frequency that can pass through. The *critical frequency 'f_c'* is obtained from (4) as

$$f_c = 9 * 10^{-3} (N)^{1/2} \tag{5}$$

for an electron density of N . As we know, the electron density of different ionospheric regions is different and maximum electron density is that of the F2 region. The maximum frequency when transmitted vertically and reflected back from a particular region of the ionosphere is called the critical frequency of that



Electron density of the ionosphere varies with local time, season, latitude, and longitude and with solar and magnetic activity conditions.

region and is normally denoted by f_oE , f_oF1 and f_oF2 , etc. for E, F1 and F2 regions, respectively.

Based on the above reasoning, in long-distance radio communication via ionosphere only high frequency (HF) band (3 to 30 MHz frequencies) is normally used (see *Figure 5*). This is because of the fact that the radio waves of lower frequency are too greatly attenuated in passing through the D region (*Figure 5c*), whereas those of higher frequency will not be reflected because the maximum critical frequency, i.e. for the F2 region is exceeded (*Figure 5b*).

As described earlier, electron density of the ionosphere varies with local time, season, latitude, and longitude and with solar and magnetic activity conditions. So, it is important to study these changes in ionospheric electron density because they affect the long distance radio communications applications via ionosphere. Different techniques are being used to study the characteristics of the ionosphere.

4.3 Ionospheric effects on Satellite Communications.

In view of the preceding discussions, it will be appreciated that the ionosphere is a useful tool for long-distance radio communication in the HF band. A completely opposite situation arises in connection with the communication from a satellite or a space probe to Earth where waves of very high frequency (VHF) i.e. well above the critical frequency of the F2 region are used to avoid their reflection back to space. For satellite and space based systems, the ionosphere acts as a deteriorating medium and affects the performance of these systems. There are a myriad of radio propagation effects depending upon the nature of the media through which rays must propagate between earth and space. These are refraction, absorption, polarization, propagation time delay, Doppler frequency shift, ionospheric scintillation, etc. In tracking and navigation applications, range error results from both a time delay due to an increase in path length (caused by bending of radio waves due to refraction) and to a



reduction in the radiowave signal velocity. Inhomogeneities in the refractivity due to the presence of ionospheric irregularities (between 200-1000 km altitude) cause fluctuations in signal power and phase of a radio wave propagating through ionospheric media and the phenomena is known as scintillation. Ionospheric scintillation causes both enhancement and fading about the medium level as the radio signal transits the disturbed ionospheric region. *Figure 6* shows an example of amplitude scintillations, observed at 4 GHz in INSAT satellite signals at the Satellite Earth Station, Sikandarabad, near Delhi during the last solar maximum period (October 1989). Such types of scintillation can severely affect the performance of any VSAT terminal.

With the increasing demands placed upon space based systems in terms of accuracy and data transmission rates, scintillation is probably the single most important deleterious factor in utilising the earth-space propagation path. Ionospheric scintillation disturbs satellite communications, global positioning navigation systems, radar system performance, and radio astronomical observations, all resulting in a serious degradation of data quality. Two areas of the globe particularly troubled by scintillation are sub-auroral to polar latitudes and a belt surrounding the geomagnetic equator. Hence, the degradation of earth-space communication links is most severe for propagation paths which transit the auroral and equatorial ionosphere, and the effect is most intense in the latter. *Figure 7* gives a synopsis of the global distribution of scintillation at L band (1.5 GHz) during the solar maximum and minimum periods. As seen from the figure, scintillation intensity increases with increase in solar activity and it is more severe near magnetic equator as compared to sub-auroral to polar regions. Scintillation is basically a night- time phenomenon.

Some of the effects due to ionospheric scintillation on earth-space propagation are listed below:

- When depth of fading due to scintillation exceeds the

Ionospheric scintillation disturbs satellite communications, global positioning navigation systems, radar system performance, and radio astronomical observations, all resulting in a serious degradation of data quality.



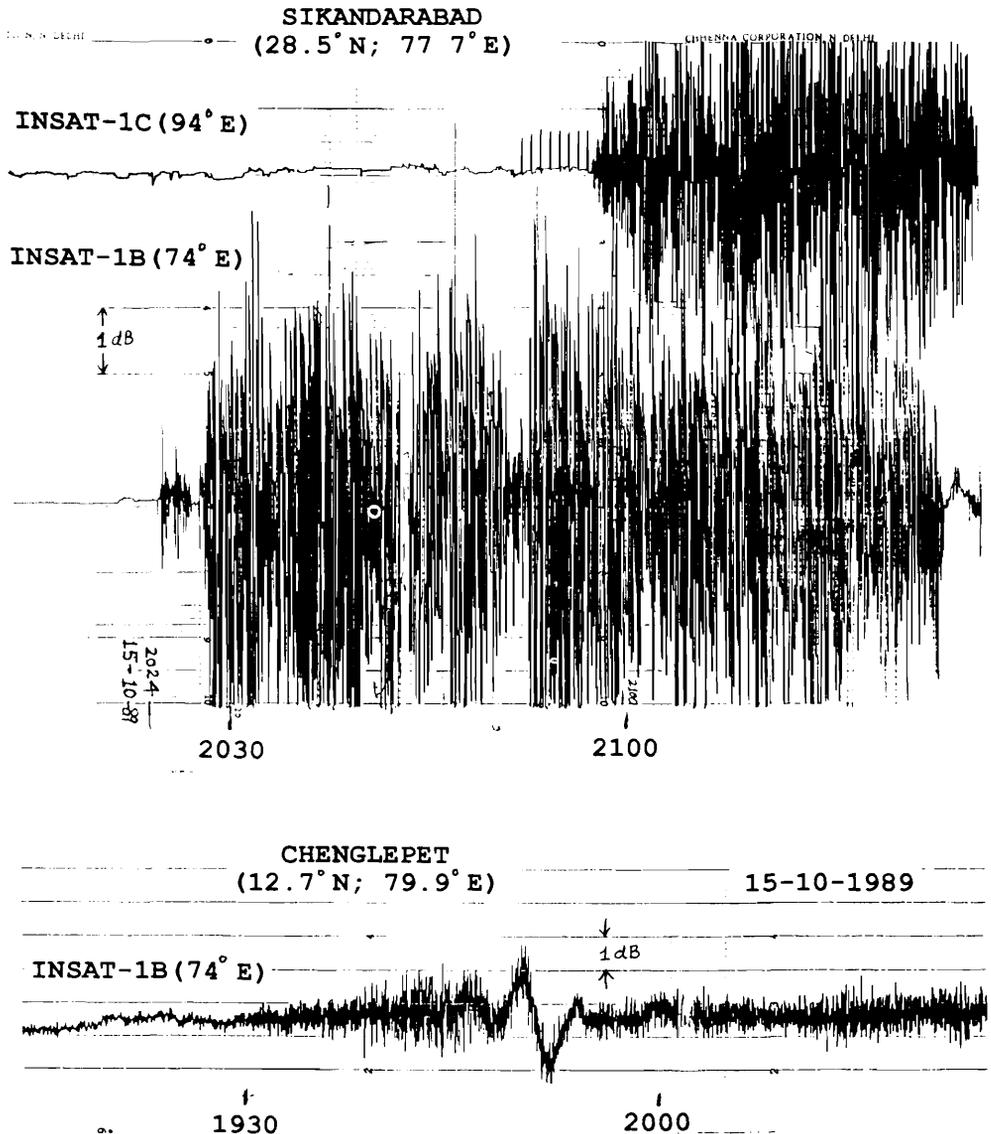


Figure 6. An example of amplitude scintillations observed at 4 GHz INSAT satellite signals at the Satellite Earth Station, Sikandarabad, near Delhi during last solar maximum period (October 1989). Such types of scintillation can severely affect the performance of any VSAT terminal.

fading margin of a receiving system, message error is encountered in satellite communication system.

- In GPS navigation system, amplitude scintillation may cause data loss and cycle slips and phase scintillation can cause

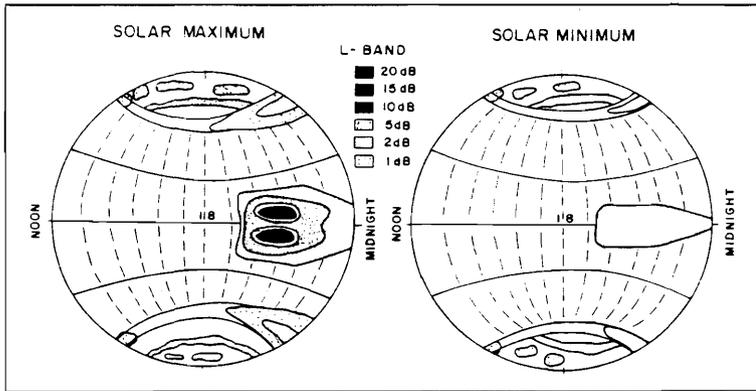


Figure 7. Global morphology of L-band scintillations

loss of phase lock, etc.

- In radar systems, scintillation may disturb radar target signatures or conversely produce fictitious additional targets.
- Power limitation on satellites and low antenna gain on aircraft cause the down link to the aircraft to be the most path sensitive to amplitude scintillation.
- In high speed digital communication systems, if the duration of fading is longer than information pulse length and the fading depth reaches a value below system sensitivity, information bit could be lost and result in increases in bit error rates.

Multipath time delay spread due to the presence of refractive index inhomogeneities in the ionosphere may also cause system degradation. If the multipath causes significant information pulse lengthening, then this may cause significant inter symbol interference.

Even after more than two decades of intensive investigations by theoretical and experimental geophysicists, the space-time variability of night time F region irregularities remains an outstanding problem. Most of the current attention is directed towards the scintillation cause-effect relationship. In the equatorial region irregularities are generated after local sunset initially over the magnetic equator due to R-T instability processes and then rise up and align along magnetic field lines and propagate to higher latitudes.

Suggested Reading

- [1] M Dolukhanov, *Propagation of radio waves*, Mir Publishers, Moscow, 1971 (Russia).
- [2] Siegfried J Bauer, *Physics of Planetary Ionosphere*, Springer-Verlag, Berlin Heidelberg, New York, 1973 (USA).
- [3] J A Ratcliffe, *An Introduction to the Ionosphere and Magnetosphere*, Cambridge University Press, 1972, (UK).

Address for Correspondence
R S Dabas
Radio and Atmospheric Sciences
Division,
National Physical Laboratory,
New Delhi 110012, India.

