The Role of the Ionosphere in Radio Wave Propagation

J. H. DELLINGER
NONMEMBER AIEE

Contents

I. Introduction

When Marconi electrified the world in 1901 by sending radio signals across the Atlantic Ocean, he incidentally proved that the upper atmosphere is electrified. Diffraction was insufficient to explain the bending of the electromagnetic waves around the 30 degrees of the earth's curved surface. The waves could not penetrate the earth. So there was only one way they could go the incredible distance, and that was by reflection from one or more conducting strata in the atmosphere. So reasoned Professor A. E. Kennelly, and he published[1] the idea a few months after Marconi's demonstration. Oliver Heaviside reached the same conclusion and published[2] it at nearly the same time; his paper however mentioned only a single layer. This postulated layer was known for a number of years as the Kennelly-Heaviside layer, and the entire ionized region in the upper atmosphere which affects the transmission of radio waves is now called the ionosphere. Radio waves transmitted by means of it are called sky waves, in contradistinction to the ground waves, those which are propagated along the earth's surface.

While it might at first be astonishing that the extremely rare upper atmosphere could contain enough material to affect the passage of radio waves, yet the very thinness of the air is what permits this to happen. At a height of 100 kilometers (62 miles) above the earth's surface, for example, the air pressure is probably much less than a millionth of what it is at the earth's surface, or the equivalent of a vacuum in an electric-discharge vacuum tube. The air particles are separated so far from one another that collisions are far less frequent than in the lower atmosphere, and when an atom is ionized by solar radiation it remains ionized for a considerable time. Therefore at any given time a large proportion of the air particles are in an ionized condition. When a radio wave passes through such ionized air the electric force of the wave accelerates the ions, which take on a certain amount of motion at the wave frequency, this accelerated motion in turn giving rise to radiation. There is thus an interchange of energy between the ionized air particles and the radio waves, with a net effect of reflection or refraction of the waves.

It has been found from radio experiments that this ionized condition does not increase uniformly as the air pressure decreases with altitude. Because of varying distribution of chemical composition of the air with height, and because of the different gases' differing capability of absorption of solar radiation of different frequencies, there are certain strata or layers in the air in which a maximum of ionization exists, that is, the ionization is greater than it is either above or below the layer.

The ionization of the ionosphere layers is principally due to ultraviolet radiation from the sun. The detailed processes by which the ionization is produced and maintained at any given level, are obscured by the almost complete lack of precise knowledge of the composition, state of dissociation, and temperature at those levels. The ionization produced in the daytime is carried over into the night by chemical, electrical, or other reactions which are at present unknown. Some speculation has appeared in print on some of the details, but that is outside the scope of this paper, which is essentially a presentation of the known facts. These facts are fairly complicated, and it is noteworthy that while some of them have been explained by some of the speculative theory, such theory has in each case been developed to fit the facts after the facts had been discovered.

Ionosphere Structure

Since the distribution of energy in the spectrum of radiation from the sun, and likewise possibly the chemical composition and temperature of the air at different heights, vary at different times of the day and year, the ionized layers in the atmosphere do not remain always at the same height but vary diurnally, seasonally, and otherwise, in both height and ionization. There may be a considerable number of such layers at a given time. Of these, two are permanent, and two others are semi-permanent. The two permanent ones are called the *E* and *F* layers. The *E* layer is at a height of 90 to 140 kilometers at different times, usually about 110 kilometers. The term *F*-layer is ordinarily reserved for the other layer as it exists at night; in the daytime during most of the year it divides into two layers which are called the F_1 and F_2. The night *F* layer is at a height of 180 to 350 or more kilometers. The F_1 layer exists in the daytime (except in winter), at a height of 130 to 250 kilometers. The F_2 layer exists every day, at a height of 250 to 350 or more kilometers in the summer, dropping to about 150 kilometers in the winter day. (The "virtual" heights, defined later, are higher than these values.) The fourth layer, which is semipermanent, is the *D* layer; it exists only in the daytime and its height is of the order of 50 to 90 kilometers. Little has been done on the determination of the quantitative characteristics of the *D* layer, its effects being largely inferred rather than directly

Paper number 39-47, recommended by the AIEE committee on communication, and presented at the AIEE winter convention, New York, N. Y., January 23–27, 1939. Manuscript submitted November 23, 1938; made available for preprinting December 29, 1938.

J. H. Dellinger is chief of the radio section of the National Bureau of Standards, Washington, D. C.

1. "On the Elevation of the Electrically Conducting Strata of the Earth's Atmosphere," A. E. Kennelly, *Electrical World and Engineer*, volume 39, March 15, 1902, page 473.

2. "Telegraphy," Oliver Heaviside, Encyclopedia Britannica, tenth edition, volume 33, December 19, 1902, page 215.

observed. Existing knowledge covers mainly the E, F, F_1, and F_2 layers.[2a]

The structure of the ionosphere may be visualized in an elementary way from figure 1, which is for a typical summer daytime condition, the F_1 and F_2 layers both being present as well as the E. This is drawn to scale, so the angles of reflection of radio waves from the layers may be estimated correctly. The three layers are shown as mere thin lines, for simplicity. The layers have in fact a certain thickness, and the density of ionization varies somewhat in this thickness. At the right of the diagram is a rough illustration of a possible distribution of ionization density with height.

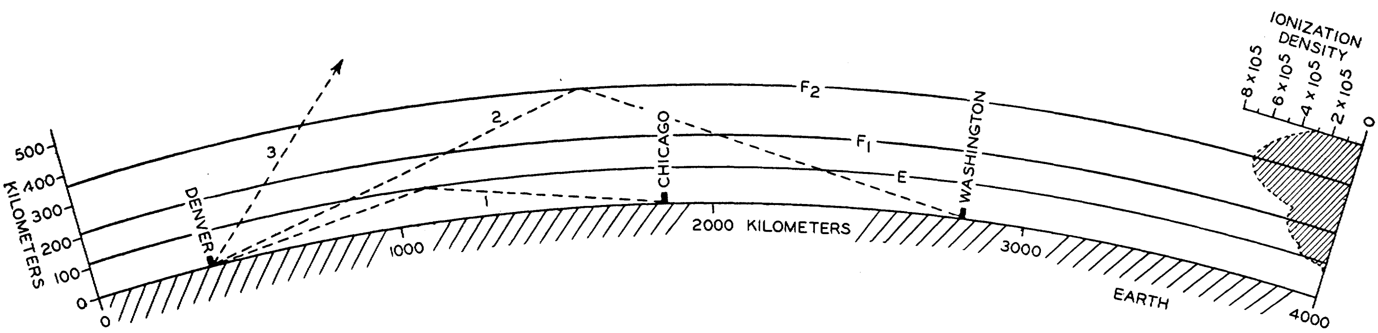


Figure 1. The principal layers of the ionosphere, drawn to scale for a particular time of day and year

Dotted lines indicate two of many possible paths of radio waves from a transmitter to a receiver as transmitted by reflection from the ionosphere layers. This picture, simple as it is, does in fact represent the basic mechanism of radio-wave transmission over long distances. When we consider the variations of ionization and height of the layers with time, and the effects of the ionization upon the received intensity and the limits of transmissible frequency at any particular time, the picture loses its simplicity. The purpose of the present paper is to outline the principal known facts and implications of this complicated situation. We shall see that the phenomena of long-distance radio transmission may be completely explained by the ionosphere.

Ionosphere Characteristics

There are three principal ionosphere characteristics which determine radio transmission. These are the height, the ionization density, and the energy absorption capability of each of the ionosphere layers. Since each layer has a certain thickness it is necessary to define the sense in which the term height is used. A ray or wave train starts to bend toward the ground immediately upon entering the layer, as illustrated in figure 2, and follows a curved path in the layer until it emerges at the same vertical angle at which it entered. It has been shown[3] that the time of transmission along the actual path BCD in the ionized layer is the same as would be required for transmission along the path BED if there were no ionized particles present. The height h, from the ground to E, the intersection of the two projected straight parts of the path, is called the virtual height of the layer. This is the important quantity in all measurements and applications.

The virtual height of a layer is measured by transmitting a radio signal from A, and receiving at F both the signal transmitted along the ground and the echo, or signal reflected by the ionosphere, and measuring the difference in time of arrival of the two. The signal is a special, very short pulse in order that the two may be separated in an oscillograph, as the time differences are mere thousandths of a second. The difference between the distance $(AE + EF)$ and AF is found by multiplying the measured time difference by the velocity of light. From this and the known distance AF, the virtual height h is calculated. It is usually convenient to make AF zero, that is, to transmit the signal vertically upward and receive it at the same place (and the term "virtual height," rigorously defined, is for this case).

The effectiveness of the ions in reflecting the waves back to earth depends on the number of ions present in a unit of volume, that is, the ionization density. It requires a greater density of ionization to reflect the waves back to earth, the higher the frequency. It has been shown that, for electron ionization, the relation[4] (for the ordinary ray) is

$$N = 0.0124 f^2 \tag{1}$$

where N is the number of electrons per cubic centimeter and f is the highest frequency, in kilocycles per second, at which waves sent vertically upward are reflected back to earth. Waves of all frequencies higher than this pass on through the ionized layer. This frequency is called the critical frequency, and measurement of it is, with the equation just given, a means of measuring the maximum ionization density in an ionized layer.

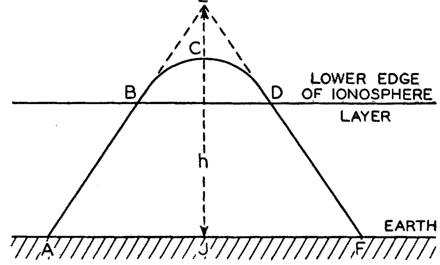


Figure 2. Diagram for explanation of virtual height of ionized layer

The critical frequency is thus the practical measure of the ionization density, the second of the principal quantities determining the role of the ionosphere in radio transmission. The way the critical frequency is measured is to

2a. The D, E, and F layers were first identified by E. V. Appleton (Papers of the URSI General Assembly, Washington, 1927, volume 1, part 1, page 2). The F_1 and F_2 layers were first identified by J. P. Schafer and W. M. Goodall (*Nature*, volume 131, 1933, page 804) and S. S. Kirby, L. V. Berkner, and D. M. Stuart (IRE *Proceedings*, volume 21, 1933, page 757). The D layer was first shown to serve as a reflecting layer by N. Smith and S. S. Kirby (*Physical Review*, volume 51, 1937, page 890).

3. Page 571 of article, "A Test of the Existence of the Conducting Layer," G. Breit and M. A. Tuve, *Physical Review*, volume 28, 1926, page 554.

4. Relation evolved through the work of Eccles (1912); Larmor (1924); Breit and Tuve (1926); Appleton (1931); Gilliland, Kenrick, and Norton (1931); and given as equation 5*a* of "Studies of the Ionosphere and Their Application to Radio Transmission," S. S. Kirby, L. V. Berkner, D. M. Stuart, Bureau of Standards *Journal of Research*, volume 12, 1934, page 15 (research paper 632). A summary of the scientific aspects of the ionosphere is given in "The Physics of the Ionosphere," H. R. Mimno, *Reviews of Modern Physics*, volume 9, 1937, page 1; this article also presents a valuable bibliography, comprising 309 references (the literature is so extensive that even this list is far from complete).

determine the virtual height of the layer by the method mentioned above, using vertical or nearly vertical transmission (that is, with the transmitter and receiver not far apart), and keep increasing the frequency until signals are no longer received back from the layer. The highest frequency at which signals are received back from the layer is the critical frequency of that layer. (For some purposes this definition is not exact; see discussion of "sporadic *E*" in section IV.) The results of such a measurement are illustrated in figure 3. Starting at a frequency below 2,000 kilocycles per second, the virtual height is found to be about 110 kilometers, and remains at about this height until about 3,000 kilocycles per second. The critical frequency of the *E* layer at the time of this measurement is thus 3,000 kilocycles per second. At this frequency the waves penetrate the *E* layer and go on up to a higher layer, the F_2. The F_2 layer has a greater ionization density and so it reflects back waves of frequency greater than 3,000 kilocycles per second. It is not until frequencies greater than 12,000 kilocycles per second are used that the F_2 layer fails to reflect them, in the case illustrated.

Near the critical frequency the waves are excessively retarded in the ionized layer, which accounts for the abnormal rise of the curve at the critical frequency. At the right of the curve appear two critical frequencies for the F_2 layer. This is an indication of double refraction of the waves due to the earth's magnetic field, two components of different polarization being produced. One is called the ordinary ray and the other the extraordinary ray. The symbols *o* and *x* are used for these, respectively. The critical frequency of a layer *n* is represented by the symbol f_n, and to such symbol the *o* or *x* is added as a superscript. Thus the critical frequencies of the F_2 layer for the ordinary and extraordinary rays are indicated by the respective symbols, $f_{F_2}{}^o$ and $f_{F_2}{}^x$. In the case of the *E* layer, usually the ordinary ray predominates and the extraordinary ray is so weak it does not affect radio reception.

Automatic recording equipment is used by the National Bureau of Standards and a few other laboratories to record the virtual heights and critical frequencies of the ionosphere layers. A sample record is shown in figure 4. In this case (afternoon in May 1933) the *E*-layer critical frequency was 2,900 kilocycles per second and the F_1-layer critical frequency was 3,800 kilocycles per second. The dark curved line in the upper right corner of the figure of the same shape as the line below it, is a multiple reflection, that is, the wave was not only reflected back from the F_2 layer, causing the lower line, but was then reflected back up from the earth's surface and down again a second time from the F_2 layer.

Measurements of the type just described are usually made at vertical incidence, that is, the waves are transmitted straight up to the ionosphere and received at a place not far from the transmitter. Strange as it may seem, such measurements tell us a great deal about long-distance radio transmission. Thus, from the results of such measurements can be calculated the upper limit of frequency usable over any given distance. A simple relation for reflection from a sharply defined layer follows from consideration of its dielectric constant and index of refraction. To a first approximation,[5]

$$f_m = f_c \sec \phi = f_c/\cos \phi \qquad (2)$$

where f_m is the maximum usable frequency over any distance, f_c is the critical frequency measured as above described at vertical incidence, and ϕ is the angle of incidence of the waves at the ionosphere, that is, the angle between the wave's line of travel *a* and the vertical line giving the virtual height *h* in figure 5. This is only a first approximation, as we shall see later, but it is close enough to permit estimating certain important effects and results. From figure 5, $\sec \phi = a/h$, *h* being the virtual height of the ionized layer. Also,

$$a = \sqrt{\frac{d^2}{4} + h^2}$$

d being the horizontal distance of transmission. Therefore,

$$f_m = f_c\sqrt{\frac{d^2}{4h^2} + 1} \qquad (3)$$

This is the approximate relation between the maximum usable frequency at a distance *d* between transmitter and receiver, and the two important characteristics of the ionosphere layer, the virtual height *h* and critical frequency f_c.

Skip Distance

An interesting consequence of this relation is: the distance at which a given frequency is the maximum usable is also the minimum distance over which that frequency is receivable, that is, there is a zone around the transmitter in which the frequency is not receivable

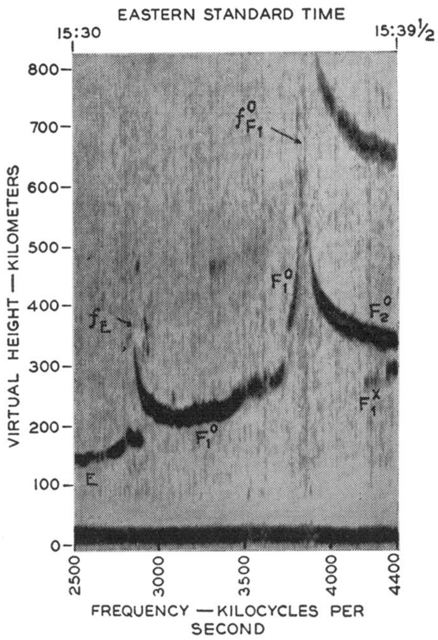
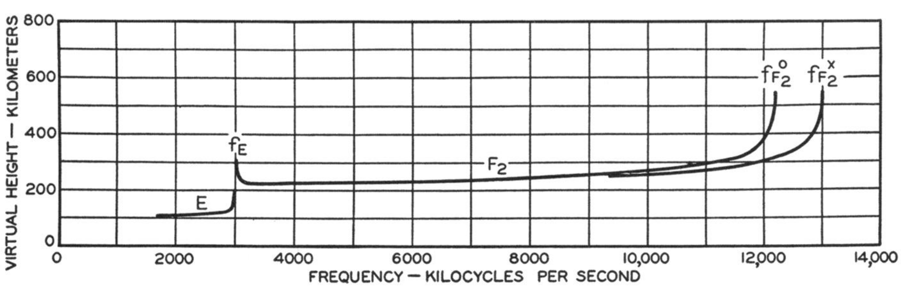


Figure 4. Example of frequency-height record as recorded by automatic multifrequency recorder of ionosphere echoes

May 8, 1933

Figure 3. Relation between observed virtual height and frequency of radio waves



5. Relation following directly from Snell's law of refraction and Eccles-Larmor equations for index of refraction in an ionized medium. Derived by my colleagues in National Bureau of Standards in 1934; published in "Multifrequency Ionosphere Recording and Its Significance," by T. R. Gilliland, National Bureau of Standards *Journal of Research*, volume 14, 1935, page 283 (research paper 768); in document "Study of Question 7 Proposed for the Fourth Meeting of the C.C.I.R., Report Submitted by U.S.A.," 1936; and elsewhere, for example, in "Maximum Usable Frequencies for Radio Sky-Wave Transmission, 1933 to 1937," by T. R. Gilliland, S. S. Kirby, N. Smith, S. E. Reymer, National Bureau of Standards *Journal of Research*, volume 20, 1938, page 627 (research paper 1,096).

but outside of which it is receivable. Thus figure 6 shows the calculated upper limit of frequency of waves reflected back by the ionosphere for various distances, for a particular time (midnight, December 1937, *F* layer). This was calculated by the simple formula (3) with the virtual height h = 310 kilometers and f_c = 4,180 kilocycles per second. The maximum frequency receivable at a distance of, say, 1,200 kilometers is 8,300 kilocycles per second, and this is higher than that receivable at any shorter distance. This means that radio waves of a frequency of 8,300 kilocycles per second are reflected by the *F* layer of the ionosphere and receivable at a distance of 1,200 kilometers or more, but at any shorter distance such waves pass on through the *F* layer to outer space and are lost. Consequently the zone around the transmitter of a radius of 1,200 kilometers is a skipped zone for the frequency of 8,300 kilocycles per second, and the distance of 1,200 kilometers is called the skip distance for that frequency at that time. There is good reception beyond the skip distance and none within it. This is subject to the qualification that there is a short distance close to the station in which reception by the ground wave is possible; this is of no interest in long-distance reception and is outside the scope of this paper. There is also some reception, usually weak and unsatisfactory, within the skipped zone due to scattered reflection of waves from irregular ionized patches in the ionosphere (see section IV hereinafter).

The reality of the skip distance is quite striking. It frequently happens that, in

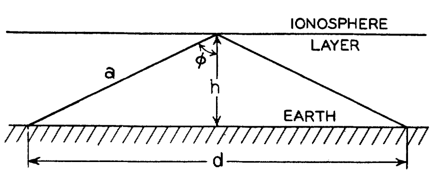


Figure 5. Simplified geometry of angle of incidence of radio waves on ionosphere layer

order to get radio messages through on a certain high frequency over a short distance, two stations not far apart communicate with one another by relaying their messages through a third station a great distance from both. For example, two stations in Pennsylvania 100 kilometers apart may be unable to send messages directly to each other on 15,000 kilocycles per second but both may easily communicate with a station in South America and relay messages through it to each other.

II. Regular Characteristics of the Ionosphere

Since the upper limits of frequency usable for radio transmission over a given distance depend directly upon the virtual heights and critical frequencies of the layers of the ionosphere, and since these are constantly varying with time of day, season, and other factors, it is particularly important to have definite data on the variations of these two ionosphere characteristics. Radio transmission conditions in general follow the critical frequencies; for example, when the critical frequencies are high the maximum usable frequencies and the best frequencies for radio communication are also high. In view of the known relations between radio transmission and ionosphere data, the facts of radio transmission can indeed be summarized in far more compact form in terms of ionosphere data than in terms of direct radio data.

A survey of the factual data of the ionosphere is in some respects like a survey of weather conditions. Like the predominant roles played in weather phenomena by atmospheric pressure, temperature, and humidity, we have the predominant roles played in ionosphere phenomena by the ionization, the virtual height, and the absorption. Our task is to learn the characteristic variations of these factors. While the two realms, weather and ionosphere, have this similarity, they are, so far as we know, independent of each other. Weather is local; the ionosphere phenomena are world-wide. Weather is due to happenings in the troposphere, extending about ten kilometers above the earth's surface, while the ionosphere phenomena occur at heights many times this. The seasonal effects in the ionosphere are synchronous with the sun's position, not lagging a month or two as do the seasons of weather. Ionosphere phenomena exhibit a greater regularity than weather phenomena and are in some respects better understood.

A large body of facts regarding the virtual heights and critical frequencies of the ionosphere layers is now available. Much less is known about the absorption. All three vary from hour to hour, from season to season, and from year to year. Each of these variations would be expected because of their dependence upon radiation from the sun. The ultraviolet radiations from the sun vary in an 11-year cycle. The last minimum was in 1933, and there were maxima in 1927 and 1938. The ionosphere was under study throughout this cycle, and we have consistent data particularly for the last six years. The present is a good time to summarize the facts of the ionosphere, for it is only now that we have come into comprehensive possession of such facts throughout the entire significant period of half a solar cycle.

The only place where the ionosphere is known to have been continuously under observation for this period is the National Bureau of Standards at Washington. The measuring technique[6] involves oscillographic observation of radio echoes of a sharp pulse. Similar work has been done from time to time by numerous observers and laboratories, and continuous observations are now in progress in many countries. Knowledge of the behavior of the ionosphere is coming to be so important in the operation of radio services that the National Bureau of Standards broadcasts information on the ionosphere characteristics and vagaries from its radio station WWV one day each week. These regular broadcasts

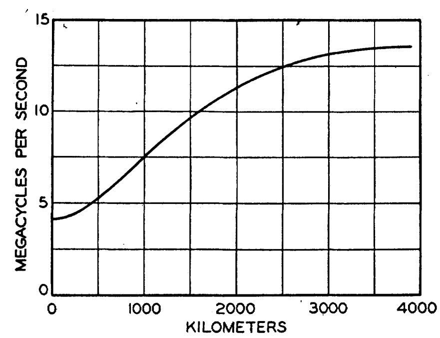


Figure 6. Relation between limit of frequency receivable and distance of transmission, calculated by simplified formula for a particular time (midnight, December 1937, F layer)

were begun in 1937. It may be necessary in the future to have such service daily from a number of places, like the weather-forecasting service.

Virtual Heights and Critical Frequencies

A satisfactory way in which to present and use the data on ionosphere layer

6. "A Test of the Existence of the Conducting Layer," G. Breit and M. A. Tuve, *Physical Review*, volume 28, 1926, page 554.
"Preliminary Note on an Automatic Recorder Giving a Continuous Height Record of the Kennelly-Heaviside Layer," T. R. Gilliland and G. W. Kenrick, Bureau of Standards *Journal of Research*, volume 7, page 783, 1931 (research paper 373); IRE *Proceedings*, volume 20, 1932, page 540.
"Note on a Multifrequency Automatic Recorder of Ionosphere Heights," T. R. Gilliland, Bureau of Standards *Journal of Research*, volume 11, page 561, 1933 (research paper 608); IRE *Proceedings*, volume 22, 1934, page 236.
"Studies of the Ionosphere and Their Application to Radio Transmission," S. S. Kirby, L. V. Berkner, and D. M. Stuart, Bureau of Standards *Journal of Research*, volume 12, 1934, page 15 (research paper 632); IRE *Proceedings*, volume 22, 1934, page 481.

heights and critical frequencies is by means of graphs of monthly averages of these quantities as a function of time of day. The monthly average is significant because the variations from day to day are small, usually less than 15 per cent, except for disturbed days, which constitute a separate subject, treated in section IV. Typical examples of the monthly average graph are given in figure 7, showing summer and winter conditions. Similar curves have been published[7] by the National Bureau of Standards for every month from May 1933 to the present, and continue to be published in the *Proceedings* of the Institute of Radio Engineers each month. Such data are also published[8] in tabular form every quarter.

The virtual heights shown in these graphs are the heights for the lowest frequencies used in the determinations. There is a slight variation of height with frequency, as suggested in figure 3, and as discussed in section III hereinafter, but the variation is small and can be neglected for many purposes. The vertical dashed lines on the graphs are for the times of sunrise and sunset at the ground, not the earlier sunrise time and later sunset time at those heights in the atmosphere where the ionosphere layers are located. While this is surprising at first sight, the ground times of sunrise and sunset are the logical ones to consider, because the sunlight arriving tangentially to the earth's surface cannot reach any atmospheric level above the dark hemisphere without having first traveled down through that and lower levels above the lighted hemisphere, and the ionizing radiation would be largely absorbed in those lower levels.

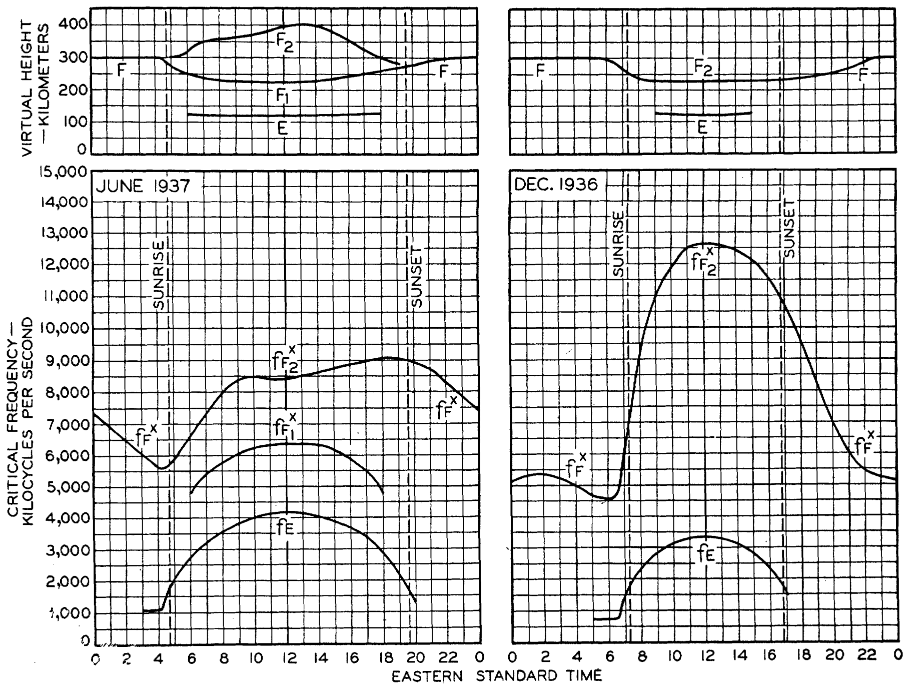


Figure 7. Typical summer and winter monthly averages of the diurnal variation of virtual heights and critical frequencies

Doubtless some of the ionizing radiation does reach the ionosphere somewhat before ground sunrise and after ground sunset, but it is by scattering, diffusion, or other special process.

The abscissas in figure 7 are time of day (Eastern standard time), on a 24-hour basis, for example, 0 = midnight, 14 = 2 p.m., etc. The following symbols are used:

f_E = critical frequency of the E layer. This is the critical frequency of the ordinary ray unless otherwise specified. When it is desired to specifically indicate it as that of the ordinary ray, the symbol is $f_E{}^o$.

$f_{F_1}{}^x$ = critical frequency of the F_1 layer. extraordinary ray.

$f_{F_2}{}^x$ = critical frequency of the F_2 layer, extraordinary ray. The night portions of the $f_{F_2}{}^x$ curve are for the F layer. The symbols for F and F_2 at night are sometimes used interchangeably.

Variations With Time of Day and Season

No regular diurnal or seasonal variation is observed in the virtual height of the E layer, which appears to be almost always between 110 and 120 kilometers. The critical frequency of the E layer has however a regular diurnal and seasonal variation. It varies synchronously with the altitude of the sun, reaching the diurnal maximum at noon, and having a higher diurnal maximum in the summer than in the winter. This behavior accords with the simple theory of ionization by ultraviolet radiation from the sun, according to which, assuming recombination of ions to be sufficiently rapid,[9]

$$f_E = K\sqrt[4]{\cos\psi} \qquad (4)$$

where ψ is the zenith angle of the sun, and K is a factor depending on the intensity of the solar radiation. This holds within fairly close limits for values of ψ not too close to 90 degrees, during the daylight hours of any one day.

No such simple relation characterizes the variation of the critical frequencies of the F, F_1, and F_2 layers. As may be seen from figure 7, the daytime critical frequencies of the F_2 layer are much higher in the winter than in the summer (although reports from the southern hemisphere indicate that the reverse is true there), and the night critical frequencies are somewhat lower in winter than in summer. The diurnal maximum of the F_2 critical frequency almost always occurs after noon, and is much later in the day in summer than in winter. The post-sunset drop in the night F critical frequency is much more rapid in the winter than in the summer; indeed in summer the midnight F critical frequency is sometimes as high as the noon F_2 critical frequency. The outstanding characteristics of the F_2 critical frequency are (1) a seasonal variation inverse to that of the E critical frequency, and (2) a diurnal lagging behind the altitude of the sun. The F and F_2 maximum critical frequency occurs in the winter day, and the minimum in the winter night. The day and night summer values are between these extremes.

7. "Multifrequency Ionosphere Recording and Its Significance," T. R. Gilliland, Bureau of Standards *Journal of Research*, volume 14, 1935, page 283 (research paper 769); IRE *Proceedings*, volume 23, 1935, page 1076.
"Characteristics of the Ionosphere and Their Application to Radio Transmission," T. R. Gilliland, S. S. Kirby, S. E. Reymer, and N. Smith, Bureau of Standards *Journal of Research*, volume 18, 1937, page 645 (research paper 1001); IRE *Proceedings*, volume 25, 1937, page 823.
"Characteristics of the Ionosphere at Washington, D. C., January to May, 1937," T. R. Gilliland, S. S. Kirby, N. Smith, and S. E. Reymer, IRE *Proceedings*, volume 25, 1937, page 1174.
"Characteristics of the Ionosphere at Washington, D. C., ...," published each month in IRE *Proceedings* for the second month before, starting with the September 1937 issue.

8. "Averages of Critical Frequencies and Virtual Heights of the Ionosphere Observed by the National Bureau of Standards, Washington, D. C., 1934–36," T. R. Gilliland, S. S. Kirby, N. Smith and S. E. Reymer, *Terrestrial Magnetism and Atmospheric Electricity*, volume 41, 1936, page 379. Similar summaries in each succeeding quarterly issue.

9. Relation given on page 251 of "Radio Observations of the Bureau of Standards During the Solar Eclipse of August 31, 1932," S. S. Kirby, L. V. Berkner, T. R. Gilliland, K. A. Norton, IRE *Proceedings*, volume 22, 1934, page 247.

The F_2 layer also differs markedly from the other layers in having much greater height in the summer than in the winter. The F_2 layer acts as though it were expanded by the heat in summer, resulting in higher virtual height and also greater volume for a given number of ions and thus lower ionization density and lower critical frequency. This is not a complete explanation, as the F_2 layer has many other anomalies, for example, a slight drop of critical frequency in midwinter from the seasonal maximum.

Year-to-Year Variations

Superposed on the variations of the ionosphere characteristics with season are the long-time variations with the solar cycle, caused by the rising and falling in an 11-year cycle of the sun's ultraviolet radiations. Practically the entire period of the observations under discussion, 1933 to the present, has been one of increasing solar radiation, as we are now at about the top of the cycle. There has consequently been an increase from year to year in the ionization and the critical frequency of all the layers. The trend reversed in 1938 and the critical frequencies will in general decrease until about 1944.

The heights of the various layers do not change appreciably from year to year. The chief change in respect to them is the disappearance of daytime F-layer stratification in the winter as the solar cycle advances. Thus, figure 7 shows no F_1 layer in December 1936, whereas in 1933 to 1935 there was a stratification into F_1 and F_2 layers in winter as well as in the summer although the stratification was much less marked in the winter than the summer.

Both seasonal and year-to-year changes of the E-layer critical frequency are shown in figure 8, curve b. This gives the average f_E at noon for each month. The full-line curve b shows the rise each summer and drop each winter, the climb to higher values each year, and also some minor fluctuations. To examine these three things separately, it is possible to redraw the curve with the seasonal variations eliminated, and this has been done. By means of equation 4, it is possible to calculate the value of the E-layer critical frequency at that point on the earth's surface where the sun's radiation is perpendicular; this is K, or $f_E/\sqrt[4]{\cos\psi}$, and it is plotted as the dashed curve.

For comparison, curve a gives average sunspot numbers for each month. The spots on the sun increase in number and activity as the solar cycle progresses, and it is believed that the intensity of the ultraviolet radiations causing the ionization of the layers of the ionosphere increases along with the sunspots. Neither the sunspots nor the ultraviolet radiations are the cause of the other, but both are manifestations of those profound activities within the sun which undergo the 11-year variation. It is not possible to compare directly against measured ultraviolet radiation from the sun because all the radiations which produce the ionosphere phenomena are absorbed in the ionosphere and do not reach the surface of the earth. The comparison with sunspot numbers is merely a comparison with a crude index of solar activity, but it is of some value.

The detailed correspondence of the dashed curve b with sunspot numbers is in fact impressive. It shows not only that the E-layer ionization follows the general trend of the sunspot numbers, but that the detailed changes from month to month show a fair amount of agreement.

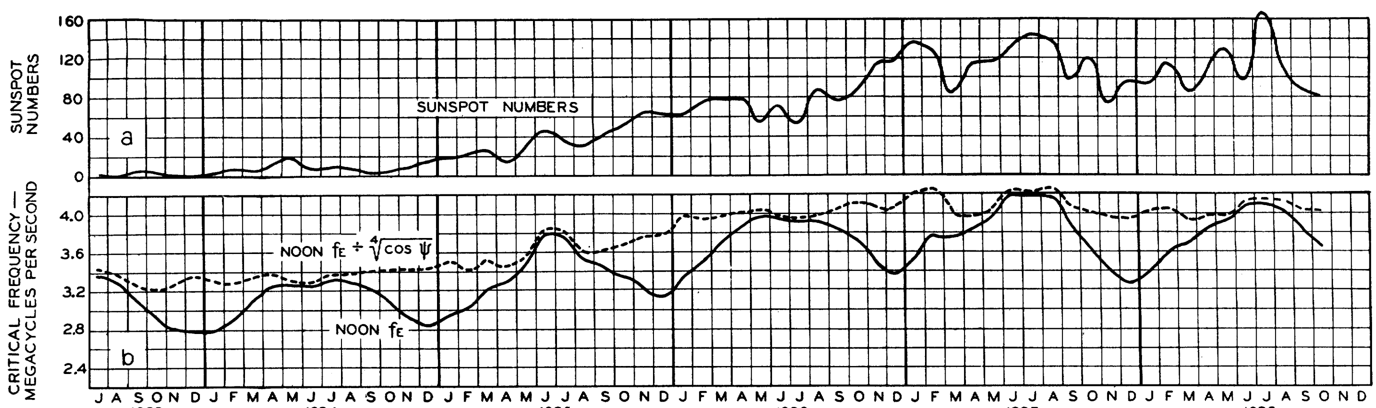


Figure 8. Variations of monthly average sunspot numbers and critical frequency of the E layer, 1933 to 1938

The solid curve (b) is the E-layer critical frequency as observed at Washington, and the dashed curve is the calculated E-layer critical frequency at the point where the sun's radiation is perpendicular

There is also fair week-to-week but not day-to-day correlation.

The F_2 layer presents a more complicated picture, since purely terrestrial influences, such as the temperature of the layer, play an important part. No such simple method therefore exists for eliminating diurnal and seasonal variations as may be done for the E layer. Figure 9 shows the variation, by average for each month, of the F_2 and F critical frequency for various parts of the day. The "midmorning" curve is for the time halfway between midnight and sunrise, and the "diurnal minimum" curve is for the time when the F critical frequency reaches its lowest value, about 30 minutes before ground sunrise. The midforenoon, midafternoon, and midevening curves are for times halfway between sunrise and noon, noon and sunset, and sunset and midnight, respectively. All of these curves show the general rise in critical frequencies superposed on the seasonal variation. The amplitude of the seasonal variation, as well as the average value, has risen each year, and is thus greater for greater solar activity. The high daytime values in winter, and high night values in summer, stand out. The midwinter dip in the daytime curves should also be noted. The diurnal minimum curve shows the least seasonal variation; this might be more or less expected, since the influence of the sun is least at this time.

Twelve-month running averages, which eliminate the seasonal variations, are given as dotted lines for the noon and the diurnal minimum values. They show more clearly the trend from year to year, and also make it easy to see the amplitude of the seasonal variations. It is interesting to compare them with the curve of sunspot numbers in figure 8.

The long-time trend of the ionosphere critical frequencies is shown more clearly in figure 10. The general trend parallel to that of the sunspot numbers is striking. All had their minimum in 1933 and maxi-

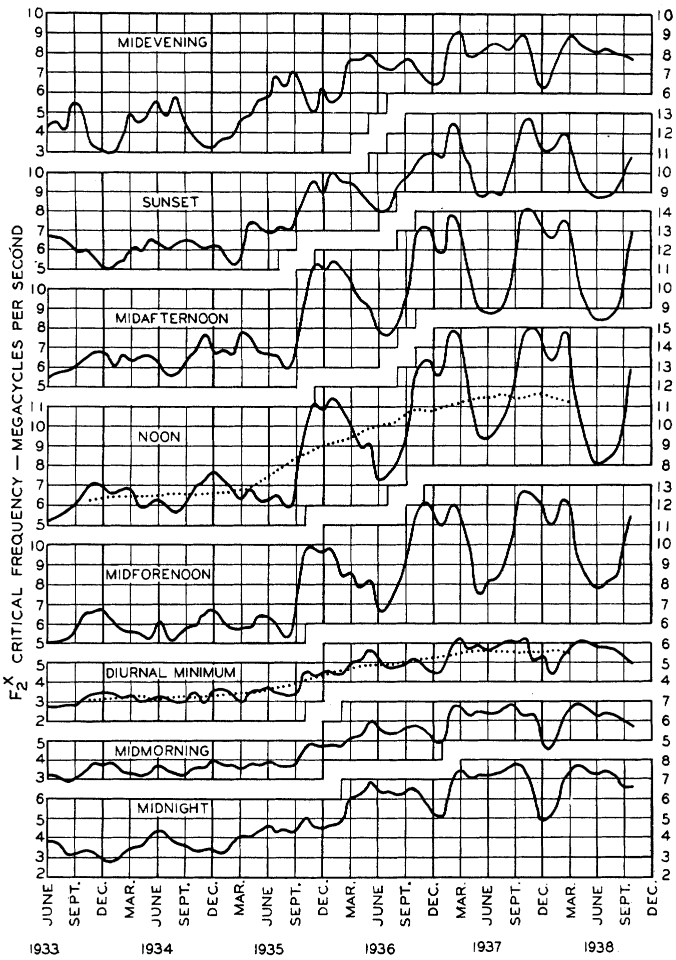


Figure 9. Variations of monthly average F_2 and F critical frequency, 1933 to 1938

mum in 1937. Again we see evidence that the sunspot numbers and the ultraviolet radiations which ionize the ionosphere have a close interrelation. The situation revealed in figure 10 is one of the most interesting of all the results yielded by radio-wave-transmission research.

Variations With Longitude and Latitude

The data presented in this paper are mostly those obtained in and near Washington, D. C., where the latitude is 39 degrees north. It is important to inquire how representative these are: are the ionosphere characteristics the same at other longitudes and latitudes? There is no a priori reason to expect any variation with longitude other than the large variations characteristic of the local time of day prevailing simultaneously at different longitudes. Ionosphere data from different longitudes in the same general range of latitude are limited. However, some information has been published in the past year, particularly from Japan,[10] and no material differences from the Washington data are found.

Latitude is quite another matter. It is to be expected that the varying angle of incidence of the solar radiations with latitude would lead to greater ionization in equatorial than polar regions. Data are limited; the available information[11] indicates that changes with latitude, other than those occasioned by the difference of season as the equator is approached or crossed, are not great. Measurements at latitude 12 degrees south indicate that the annual average critical frequencies (and thus the ionizations) are somewhat higher, and the virtual heights somewhat lower, than at Washington, 39 degrees north. There is relatively little seasonal variation near the equator. Measurements at latitude 70 degrees north, on the other hand, show lower average critical frequencies, and greater variations with season, than at Washington.

The variation with latitude is bound up with another effect, the occurrence of ionosphere disturbances, different types of which produce effects which are distributed differently in latitude. In particular, disturbances called "ionosphere storms," lasting a day or more and causing great changes in the F_2 layer height and ionization, are more intense and more frequent as the magnetic pole is approached. It is therefore important to separate days of this type from the undisturbed days, particularly when considering data from higher latitudes. This is discussed further in section IV.

III. Normal Radio Transmission

As explained in section I, some of the important facts of radio wave transmission over great distances are determinable from the characteristics of the ionosphere. We shall now inquire more carefully into the way in which this is done, and explain the application of these facts to practical radio-communication problems.

Relation of Oblique to Vertical Incidence

To compute long-distance radio data from ionosphere data is essentially to determine the relation between radio transmission incident obliquely on an ionosphere layer and transmission incident vertically on the layer. This is because ionosphere data are an expression of, and are determined by, radio transmission incident vertically on the ionosphere. When radio waves are transmitted over some horizontal distance, they are sent obliquely into the ionized layer, and as we have seen the ionized

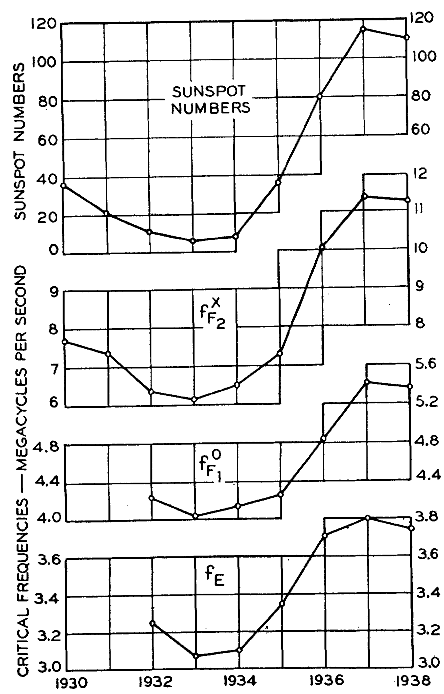


Figure 10. Comparison of annual averages of critical frequencies and sunspot numbers

10. "On the Long-Period Variations in the F_2 Region of the Ionosphere," K. Tani, Y. Ito, H. Sinkawa, *Report of Radio Research in Japan,* volume 7, 1937, page 91; also IRE *Proceedings,* volume 26, 1938, page 1340.
"Annual Variations in Upper-Atmospheric Ionization," K. Maeda, T. Tukada, T. Kamoshida, *Report of Radio Research in Japan,* volume 7, 1937, page 109.

11. "Annual Variation of the Critical Frequencies of the Ionized Layers at Tromso During 1937," L. Harang, *Terrestrial Magnetism and Atmospheric Electricity,* volume 43, 1938, page 41.
"The Ionosphere at Huancayo, Peru, November and December 1937," H. W. Wells and H. E. Stanton, *Terrestrial Magnetism and Atmospheric Electricity,* volume 43, 1938, page 169.

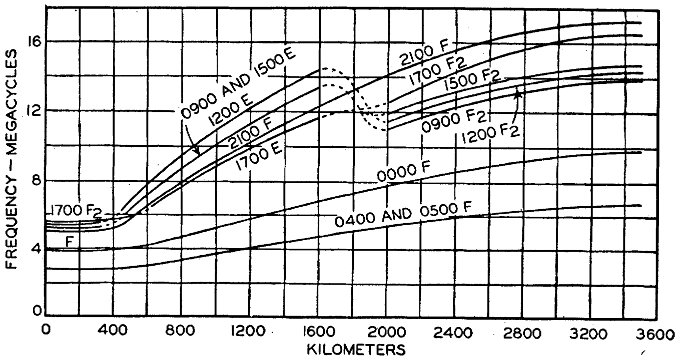


Figure 11. Average for June 1933 of maximum usable frequency at various distances and times of day

medium is able to reflect such waves of frequencies considerably in excess of the vertical-incidence critical frequency. As indicated by the approximate equation 2, $f_m = f_c \sec \phi$, a layer of given ionization density returns waves of frequencies which are higher, the greater the angle of incidence upon the layer. Thus, except as limited by the curvature of the earth, higher frequencies can be reflected the greater the distance of transmission and the lower the height of the reflecting layer.

In figure 1, for example, the E layer is shown reflecting a ray (1). Supposing the frequency to be the highest frequency which can be reflected by the E layer at that angle of incidence, then the ray (2) with a slightly smaller angle of incidence would not be reflected by the E layer but would penetrate it and go on up to the F_1 or F_2 layer and be reflected by one of them because of its greater ionization density. As the angle of incidence is decreased (angle of take-off increased), we arrive at a ray (3) which would not be reflected by the F_2 layer but would penetrate it and be lost to outer space. This is suggested by the arrow at the end of ray 3.

The approximate relation, $f_m = f_c \sec \phi$, gives a rough idea of the relation between f_c, the vertical-incidence critical frequency, and f_m, the maximum usable frequency, at any oblique angle of incidence, that is, at any distance. The exact relation[12] has to take into account a number of other factors. In the first place, the angle ϕ is itself a function of frequency, because the virtual height of the reflecting layer is a function of frequency. The higher the frequency the farther into the ionized layer do the radio waves penetrate. Thus, in figure 3 the virtual height rises slightly with frequency; this figure is for vertical incidence, and a similar rise appears in the relation for oblique incidence.

Secondly, the simple relation is modified because the earth's surface and the ionosphere layers from which the reflection takes place are not plane but curved, altering the geometry. Thirdly, the presence of the earth's magnetic field alters the relation, depending upon the length, direction, and location of the path of transmission.

All of these effects together make the calculation of maximum usable frequencies somewhat complicated. The National Bureau of Standards therefore publishes[13] curves of maximum usable frequencies as well as of critical frequencies in its monthly and other publications on the ionosphere. Examples are given in figures 11 to 14. Comparison of figure 14 with figure 6 shows how the simple secant relation or its equivalent, equation 1, is inadequate to represent the actual facts of radio transmission.

Maximum Usable Frequencies and Skip Distances

Figures 11 to 14 give the averages (for the days free from ionosphere storms) for the months shown, of the upper limits of usable frequencies, as a function of distance, for various times of day. The characteristic differences between summer and winter are shown, and also the striking increase of frequency with the advance of the cycle of solar activity during 1933 to 1937. This is what would be expected from the ionosphere changes depicted in figures 8 to 10. In fact, the diurnal, seasonal, and other effects shown in these curves follow directly from the ionosphere data of the type presented in section II of this paper.

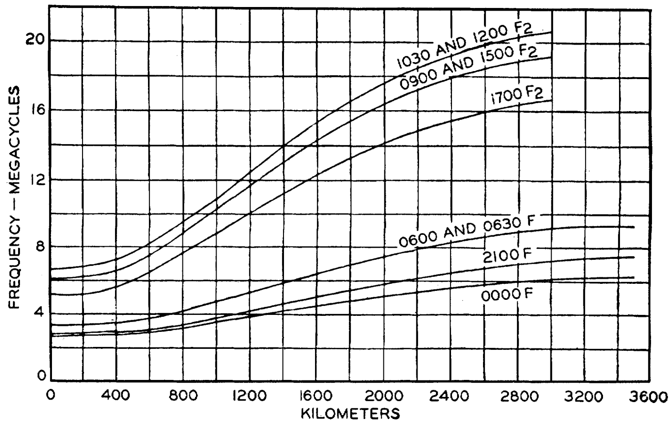


Figure 12. Average for December 1933 of maximum usable frequency at various distances and times of day

In figures 11 to 14, the layer which determines the maximum usable frequency is marked on each curve. As shown, the F and F_2 layers determine it most of the time; this is because their ionization so much exceeds that of the lower layers. In the daytime in summer, however, the E-layer ionization is so great, and beyond a distance of about 500 kilometers the lower height of the E layer results in so much more oblique transmission, that the E layer is controlling. The curvature of the earth and of the E layer limits the reflection distance to about 1,700 kilometers; beyond that the F_2 layer is controlling as the curves show. The F_1 layer is seldom (summer day, for 2,000 to 3,000 kilometers) enough lower than the F_2 layer, and enough more ionized than the E layer, to determine the maximum usable frequency.

There are times when transmission does take place at frequencies higher than these critical frequencies, either because of "sporadic E" layer or scattered reflections. These special phenomena are discussed in section IV.

Each point on these curves not only gives the upper limit of frequency which is usable over the distance, but also conversely gives the lower limit of distance over which the frequency gives satisfactory sky-wave transmission. The curves thus serve the additional purpose of giving data on skip distance. Where skip distances are of primary interest, the relation between frequency, distance, and time of day may be more conveniently shown by curves

12. "Application of Vertical-Incidence Ionosphere Measurements to Oblique-Incidence Radio Transmission," N. Smith, Bureau of Standards *Journal of Research*, volume 20, 1938, page 683 (research paper 1100).
"The Relation Between Atmospheric Transmission Phenomena at Oblique Incidence and Those at Vertical Incidence," G. Millington, Physical Society (Great Britain) *Proceedings*, volume 50, 1938, page 801.

13. "Maximum Usable Frequencies for Radio Sky-Wave Transmission, 1933 to 1937," T. R. Gilliland, S. S. Kirby, N. Smith, and S. E. Reymer, Bureau of Standards *Journal of Research*, volume 20, 1938, page 627 (research paper 1096); IRE *Proceedings*, volume 26, 1938, page 1347.
"Characteristics of the Ionosphere at Washington, D. C.,....," published each month in IRE *Proceedings* for the second month before, starting with the September 1937 issue.

of skip distance as a function of time of day, for various frequencies. Such a curve is shown in figure 15; it is for June 1937, and thus gives the same data as figure 13. Curves of skip distance against time, as in figure 15, are useful in selecting frequencies for a mobile radio station, such as an airplane or a ship, where the distance of transmission continually varies and it is necessary to determine a time and distance at which to change from one frequency to another.

Another useful form in which the data may be put is a set of curves of frequency as a function of time of day, for various distances. The data for June 1937 are shown in this form in figure 16. The curve for zero distance is identical with the curve of vertical-incidence critical frequency. Curves of this type are given in the monthly ionosphere reports of the National Bureau of Standards.[13] Such curves are useful in the operations of radio stations communicating between fixed points, as they are an aid in selecting the frequencies to be used during various times of day, for communication over a given distance.

Since all the curves change with time of day, and since on any long radio transmission path (except north-south) the time of day is different at all parts of the path, it is important to select the proper time of day in using these curves. The proper time is that of the locality where the waves reach, and are reflected by, the ionosphere. For paths in which the waves go from transmitter to receiver by a single reflection from the ionosphere, the place of reflection is halfway between transmitter and receiver.

Transmission by Multiple Reflections

It may be noted that figures 11 to 16 give no distances greater than 3,500 kilometers. The maximum possible distance of transmission by a single hop, that is, reflection from any ionosphere layer, is limited by the geometry of the earth's surface and the layers, and also by absorption at the ground of those waves which are nearly tangential. It is found in practice that the minimum angle with the ground of the radio waves transmitted or received averages about $3^1/_2$ degrees over land. From the geometry it results that the maximum distance along the earth by a single hop is ordinarily about 3,500 kilometers for the F_2 layer, and about 1,700 for the E layer. These may be exceeded in particular cases.

While the curves are drawn for single-hop transmission, they are nevertheless available for solving problems of multihop transmission, that is, multiple reflections from the ionosphere, with intermediate reflections from the ground. Calculation of the maximum usable frequency for multihop transmission is necessarily somewhat complicated. In the first place, we have to consider the time of day of the locality where each reflection from the ionosphere layer takes place. To a first approximation, the maximum usable frequency is the lowest one of the several corresponding to the times of day at the several localities where reflection takes place, for the distance on the curves equal to the transmission distance divided by the number of hops.

For very long paths, if very widely different latitudes are involved, it may be desirable to use curves appropriate to the latitude of each hop. As mentioned in section II, changes of ionosphere characteristics with latitude are not great, and it follows that the effect of latitude is of minor importance in radio transmission, except for extreme differences of latitude. In particular, for transmission within an area the size of the United States, or Europe, differences of latitude do not need to be taken into account. In transmission between the United States and Europe, however, although the difference of latitude may be no greater than within either area, the ionosphere characteristics vary materially along the transmission path from another cause. This is the effect of propinquity to the north magnetic pole. Certain ionosphere disturbances, as mentioned in section II, are more intense and more frequent as the magnetic pole is approached. There may also be a steady or permanent set of ionosphere effects having a maximum at the magnetic pole. See in this connection section IV.

For very long paths in which widely different longitudes (that is, times of day) are involved, it sometimes happens that the waves travel different parts of the way by different layers. For such cases, it is necessary to take account of the

Figure 14. Average for December 1937 of maximum usable frequency at various distances and times of day

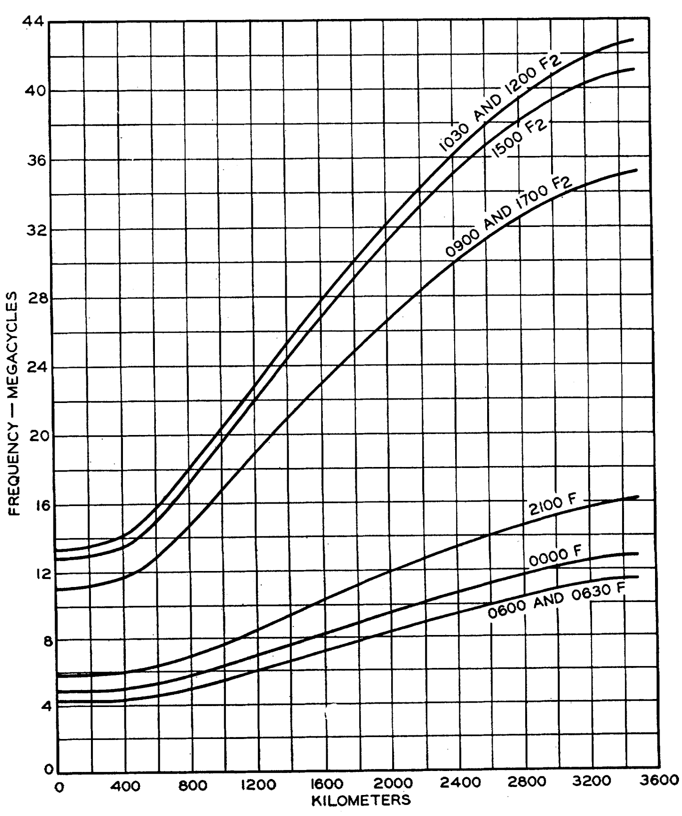
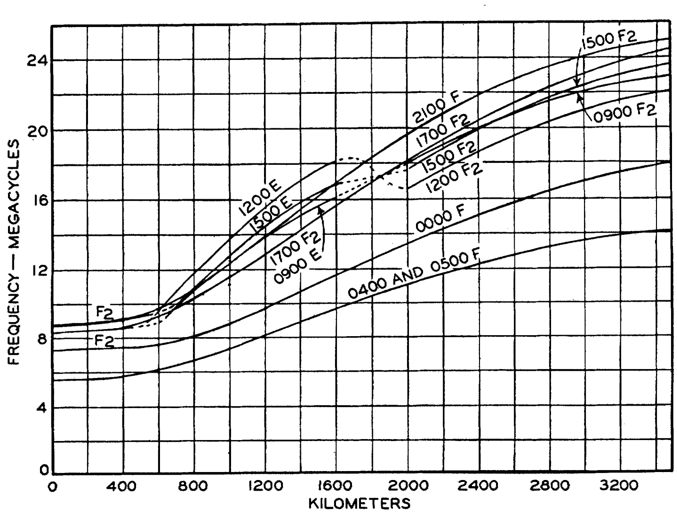


Figure 13 (below). Average for June 1937 of maximum usable frequency at various distances and times of day



heights of the different layers to determine the lengths of the several hops, and also to employ separate curves of maximum usable frequency for each layer.

The waves reaching a given point may be a combination of waves having traveled by different numbers of hops. For each of these it is necessary to take account of the time of day, and to consider which layer is effective, for the locality where each reflection from the ionosphere occurs.

Optimum Frequencies

It is necessary to inquire whether the maximum usable frequency is the optimum frequency. It is found in practice that in general (especially in the daytime) the absorption is greater, that is, received intensities are less, as the frequency is lowered below the maximum usable frequencies. Thus it requires much greater power to get communication through on frequencies very much below the maximum usable. On the other hand, a frequency should be chosen somewhat below the values indicated in the curves of monthly averages because of the variability from day to day, which is generally within 15 per cent.

Fair efficiency of communication is usually provided in the daytime by frequencies down to about 50 per cent of the maximum usable frequencies, and at night by frequencies down to somewhat less than 50 per cent of the maximum usable frequencies. Definite limits cannot be set because there are large irregular variations of absorption with time over both short and long periods. At frequencies near the maximum usable frequencies there is relatively little difference between night and day absorption. As the frequency is lowered, however, the daytime absorption increases relatively much more rapidly.

A simple rule is to use a frequency between 50 per cent and 85 per cent of the monthly average maximum usable frequency for the given distance and time. It is not ordinarily possible to keep changing frequency continuously so some such range of choice is necessary. Below 50 per cent, the received waves are likely to be too weak for use, and above 85 per cent communication will be impossible on some days.

Interesting examples of the calculation of frequencies to be used for practical radio transmission, from the graphs of maximum usable frequency, have been published[14] recently.

Experimental confirmation of the validity of these calculations and these considerations of the relations between ionosphere characteristics and radio transmission conditions is given by extensive experience of radio stations and also by special experiments. The National Bureau of Standards continuously records the intensity of radio waves received from a number of stations at various frequencies and distances, in order to investigate this.

An example of the type of evidence given by these recordings is given in figure 17. The upper part of the figure is a continuous record of the field intensity of W6XKG, Los Angeles, Calif., on 25,950 kilocycles per second, at a distance of 3,700 kilometers. The lower part of the figure gives graphs of the calculated maximum usable frequency over this distance, for both one- and two-hop transmission. This distance is slightly beyond the ordinary maximum distance for good one-hop transmission, and the one-hop transmission is thus weak. The changes between one- and two-hop transmission are well marked; the ratio of average received intensity is as great as 100 to 1. This difference in intensity is due partly to the unfavorable angle of take-off for the single hop, and partly to the increased absorption over the flatter trajectory. It should be noted how the times of beginning and ending of two-hop transmission agree with the times the calculated maximum usable frequency passed through about 26 megacycles per second at the western and eastern hops, respectively: that is, the two-hop transmission began as soon as the western hop permitted and ended when the eastern hop failed. After the failure of two-hop transmission at 1,850 (evening), the station came in weakly by one hop. As the evening progressed (after 1,850) the intensity increased, owing to the departure of the daytime absorption and the rise in height of the layer, with a consequent more favorable angle of take-off. The failure of single-hop transmission at 2,030 is seen to agree with the time of diminution of the calculated maximum usable frequency through 26 megacycles per second. (The very low intensity after 2,030 is electrical noise, that is, atmospheric disturbances or "static.")

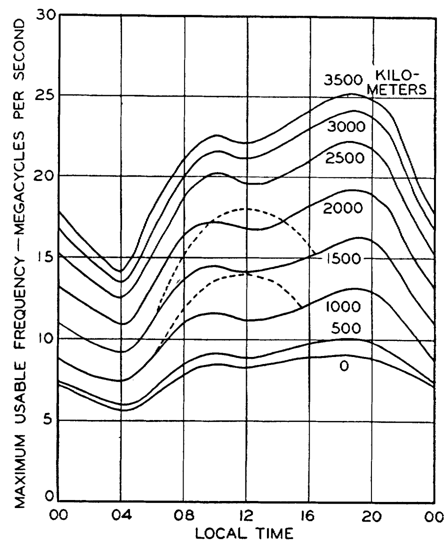


Figure 16. Maximum usable frequency at various times of day and distances, June 1937

Solid curves—*F* layer
Dashed curves—*E* layer

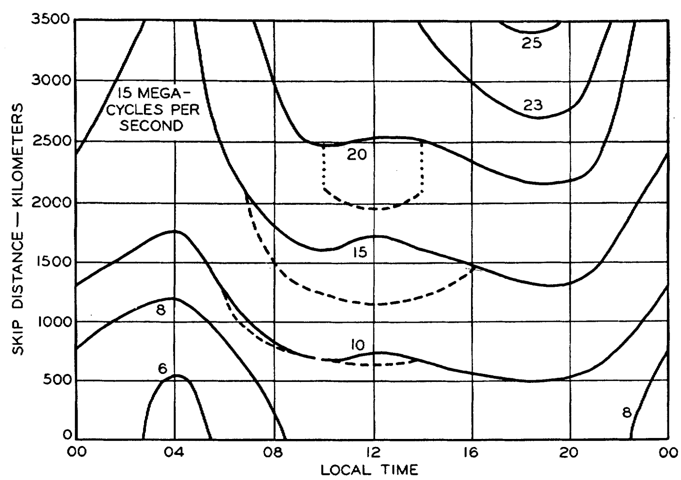


Figure 15. Skip distance at various times of day and frequencies, June 1937

Solid curves—*F* layer
Dashed curves—*E* layer

Received Field Intensities at High Frequencies

The received field intensity of waves propagated via the ionosphere depends on so many factors as to defy expression in any simple summary or any set of graphs. Data on this are constantly

14. "Maximum Usable Frequencies for Radio Sky-Wave Transmission, 1933 to 1937," T. R. Gilliland, S. S. Kirby, N. Smith, and S. E. Reymer, Bureau of Standards *Journal of Research*, volume 20, 1938, page 627 (research paper 1096); IRE *Proceedings*, volume 26, 1938, page 1347.
"The Application of Graphs of Maximum Usable Frequency to Communication Problems," N. Smith, S. S. Kirby, T. R. Gilliland, Bureau of Standards *Journal of Research*, volume 22, January 1939 (research paper 1167).

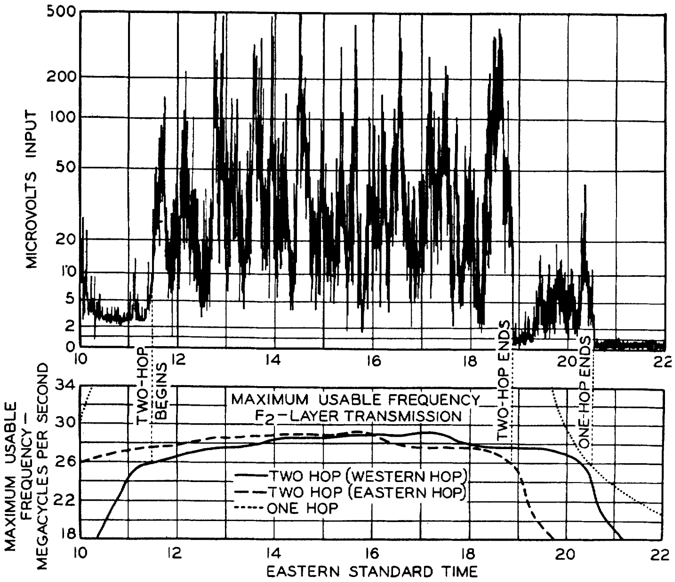


Figure 17. Example of effect upon received intensity of change of transmission from one to two hops and back to one hop

January 26, 1938; W6XKG, 25,950 kilocycles per second, Los Angeles, Calif.; 3,700 kilometers from Meadows, Md.

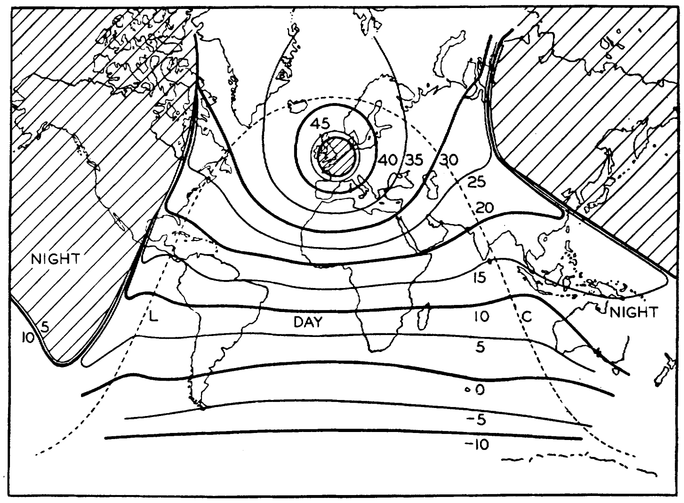


Figure 18. Sample diagram of lines of equal average received intensity for a particular frequency, time of day, season, epoch of solar cycle, and location of transmitter

The numbers are quasi-maximum field intensities expressed in decibels above one microvolt per meter, for one kilowatt radiated. The dotted line is the locus of sunrise (*L*) and sunset (*C*). The shaded areas are regions of zero reception

f = 18.8 megacycles per second (16 meters); winter, 1200 local time

being accumulated in the daily operations of radio stations and in such special recording laboratories as the National Bureau of Standards. As such data are analyzed they are found to be consistent with the ionosphere facts such as are reported in this paper, but detailed connections are limited because of the lack of direct measurements of ionosphere absorption. However, the facts of received field intensities are adding to our knowledge of the ionosphere, as just illustrated in the discussion of figure 17.

Data on the received field intensities of waves propagated via the ionosphere, as obtained in connection with the operations of radio stations, have been published in numerous articles.[15] Since sky waves exhibit great fading, the intensities constantly fluctuating over a great range, it is necessary to deal with averages. A convenient form of average for this purpose is the quasi-maximum, which is the value exceeded by the instantaneous value of the field intensity five per cent of the time.

An attempt was made to summarize a fairly large body of such information in the 1937 London Report of Committee on Radio Wave Propagation.[16] In a series of 18 figures, it gives contour lines of field intensity on a world map for two frequencies (8,600 and 18,800 kilocycles per second), for three times of day, for one epoch of the solar cycle, 1929–32, and for the transmitter (or receiver) located at London. For other locations of transmitter (or receiver) the data would be different, both because of different latitude and different distance from magnetic pole. The data given do not apply to times of ionosphere storms or other vagaries. One of these figures is reproduced here as figure 18. It gives received intensities for one kilowatt radiated from the transmitter. The curves are averages of values varying over a wide range. No set of graphs or tables would be adequate to give the facts of transmission for all frequencies, times, and other conditions.

Received intensities are largely dependent on the absorption of the wave energy in the ionized parts of the atmosphere below the ionized layer which reflects the waves. The absorption in general increases as frequency is decreased below the maximum frequency transmissible via a given layer. The absorption determines the minimum usable frequency and the maximum distance of communication, just as the ionization density or critical frequency determines the maximum usable frequency and the minimum distance (skip distance).

The absorption is found to vary with time of day, season, frequency, and length of path. It is usually greater during the day than during the night. It is greater during the summer day than during the winter day. A given frequency at large angles of incidence (long distance) behaves, with respect to absorption, like a much lower frequency at small angles of incidence (short distance).

Definite information about absorption and other characteristics of transmission via the different layers is given by continuous graphical records of received intensities. To illustrate the type of information given by many thousands of such graphs which have been recorded, two examples are given in figure 19. The top pair of graphs shows morning and evening intensities of reception from a station on 6,060 kilocycles per second at a distance of 648 kilometers. Shortly before 0700 (morning) there is a sharp increase in the general level of received intensity, and shortly before 1800 (evening) there is a sharp decrease and change in character of the received intensity. These are the times at which the radio transmission changed from F to E layer and vice versa. It follows that at these times the maximum usable frequency via the E layer for a distance of 648 kilometers was 6,060 kilocycles per second. The vertical-incidence critical frequency corresponding to this is 2,450 kilocycles per second. Determination of the times of change of layer from such records gives a means of determination of the vertical-incidence

15. A few examples are:
"Some Measurements of Short-Wave Transmission," R. A. Heising, J. C. Schelleng, G. C. Southworth, IRE *Proceedings*, volume 14, 1926, page 613.
"Short-Wave Wireless Telegraphy," T. L. Eckersley, IEE (London) *Journal*, volume 65, 1927, page 600.
"The Propagation of Short Radio Waves Over the North Atlantic," C. R. Burrows, IRE *Proceedings*, volume 19, 1931, page 1634.
"Attenuation of Overland Radio Transmission in the Frequency Range 1.5 to 3.5 Megacycles per Second," C. N. Anderson, IRE *Proceedings*, volume 21, 1933, page 1447.

16. "Report of Committee on Radio Wave Propagation," IRE *Proceedings*, volume 26, 1938 page 1193.

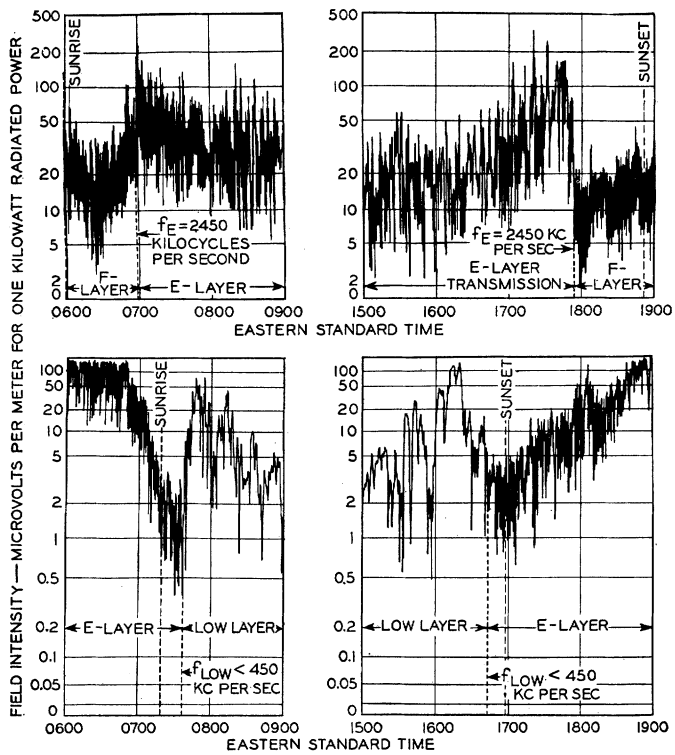


Figure 19. Typical changes of layer at a high frequency and a broadcast frequency

Top — W8XAL, Mason, Ohio; frequency 6,060 kilocycles per second, distance 648 kilometers; April 7, 1936

Bottom — WTIC, Hartford, Conn.; frequency 1,040 kilocycles per second, distance 480 kilometers; January 16, 1937

critical frequencies at particular times. This is the reverse of the process by which the graphs of maximum usable frequency described above were obtained.

Broadcast-Frequency Sky Waves

Referring further to figure 19, the lower pair of graphs shows the morning and evening intensities of reception from a broadcast station, on 1,040 kilocycles per second at a distance of 480 kilometers. At about 0740 in the morning there is a change of character and the start of a rise in intensity, and at about 1640 in the evening there is a marked change of character followed by a rise in intensity. These may be interpreted as the times at which the radio transmission changed from the E to a lower layer, and vice versa, and thus the times at which a layer below the E layer had a maximum usable frequency of 1,040 kilocycles per second for the distance of 480 kilometers. From this it may be calculated that the vertical-incidence critical frequency of the low layer is something less than 450 kilocycles per second; the exact value cannot be computed from this one observation because the height is not known. This shows that there is a layer below the E layer by which waves of broadcast frequency are reflected in the daytime. (In the summer the absorption is so great there is no transmission via the ionosphere at broadcast frequencies greater than about 1,000 kilocycles per second.)

This low layer may be called the D layer. Little is known about it, as ionosphere measurements have not been made directly upon it. Its properties are inferred from data of the kind shown in figure 19.

There is considerable empirical information on received sky-wave intensities at night on broadcast frequencies (which in Europe are from 150 to 1,500 kilocycles per second), for distances out to about 4,000 kilometers, and limited information for distances out to about 15,000 kilometers. For these frequencies the known facts can be expressed very simply, since to a first approximation the received intensity is the same for different frequencies and times of year, and at great distances is practically independent of ground conductivity. A curve of quasi-maximum values of received intensities at night on broadcast frequencies, for one kilowatt radiated from the transmitter, is given in figure 20. The same information on an expanded scale, for a part of the distance range, is given in figures 21 and 22. These two figures are for two typical cases of ground conductivity σ, stated on the figures in electromagnetic units.

The divergences of the curves in figures 21 and 22 for distances less than about 400 kilometers arise merely from the fact that at those distances the ground wave contributes appreciably, while at greater distances the received intensity is due entirely to the sky wave, that is, the wave propagated by reflection from the ionosphere.

The curve splits into two for distances beyond about 1,400 kilometers, one for paths far from the magnetic pole and one for paths relatively near the magnetic pole. This is because of the special effects in the vicinity of the magnetic pole mentioned above. The effect is very large. For instance, for transmissions between the United States and Europe, a distance of say 6,000 kilometers, the intensities are less than a tenth of what they are for transmissions over the same distance between the United States and South America.

There are variations from year to year. The evidence indicates that sky-wave intensities at broadcast frequencies are less in years of high sunspot numbers.

The data expressed by the curves represent only a beginning of effort to reduce

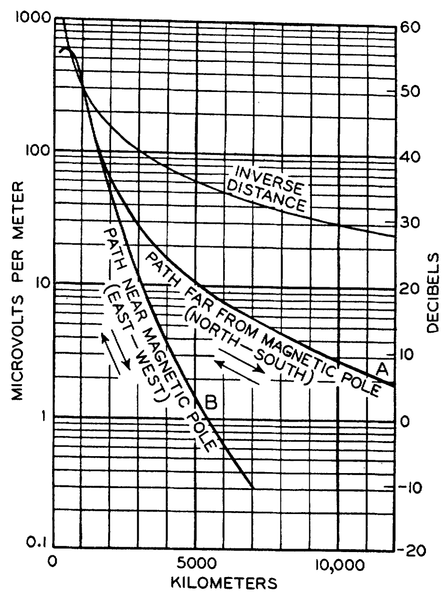


Figure 20. Received field intensity at night over great distances on frequencies from 150 to 1,500 kilocycles per second

Quasi-maximum for one kilowatt radiated

For the significance and limits of application of these curves, see text

this complex subject to quantitative form. The data for distances from 4,000 to 12,000 kilometers are based on measurements which although extensive, and averaged for several years, were all made at one time of year, early morning in the winter; it is not known to what extent they are representative for other times of the night and year. Incidentally, the practical importance of this is reduced, as far as east-west transmission is concerned, because night conditions remain only a short time over the whole of a very long path, particularly in the summer; furthermore there is little broadcasting after midnight at any one place, and thus time differences in the broadcasting tend to prevent interference.

The Contributions of the Several Layers to Practical Transmission

It is extremely hazardous to generalize about the intensity or the satisfactoriness of radio reception at different seasons, different ranges of frequencies, etc. A simple summary is impossible because (*a*) the constant changes of conditions in each layer are much too complicated, (*b*) radio transmission changes from one layer to another, and (*c*) the facts of absorption are largely unknown. Further complications are the various irregularities described in section IV. Furthermore, the satisfactoriness of reception is determined by other factors besides the received field intensity, notably the intensity of received noise or atmospheric disturbances. Noise itself exhibits complicated variations with time of day, season, place, etc.; a discussion of noise and its effects on radio reception is outside the scope of this paper.

Some information on what layers are effective in radio propagation at various ranges of frequencies is given in the curves of maximum usable frequencies, such as figures 11 to 16, but not full information.

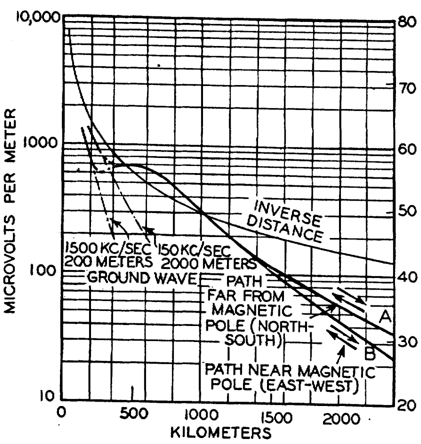


Figure 21. Night field intensity on frequencies from 150 to 1,500 kilocycles per second, for propagation over sea water

$\sigma = 4 \times 10^{-11}$ electromagnetic units

Quasi-maximum for one kilowatt radiated

For the significance and limits of application of these curves, see text

When the maximum usable frequency is transmitted by the F_2 layer, for instance, the curves give no information as to how effectively a lower layer may be serving at the same time as the means of transmission for lower frequencies. Thus in fact most long- and medium-distance communication in the summer is via the *E* layer; this is in part occasioned by the frequent occurrence of sporadic *E*. The F_1 layer is normally useless for radio transmission except for narrow ranges of distance and frequency, and need seldom be considered; this is because it is seldom enough lower than the F_2 layer and enough more ionized than the *E* layer to serve as the transmitting medium, particularly for considerable distances.

Daytime sky-wave transmission at broadcast frequencies is via the *D* layer. Night-time transmission at broadcast frequencies is principally via the *E* layer. At frequencies higher than about 1,000 kilocycles per second and particularly for the shorter distances, it is commonly via the *F* layer; this is particularly true in winter, sporadic *E* transmission replacing the *F* during much of the summer.

Specific information as to the layer which is operative in transmitting a particular frequency over a particular distance at a particular time is given by actual records of received intensities such as figures 17 and 19; there is, of course, as previously mentioned, no way of compressing all the information from such records into a simple summary. When the facts of radio transmission are examined, it is found that in general they follow very nicely from the facts of the ionosphere and the interpretations given here.

IV. Effects of Ionosphere Irregularities

The primary effects of the ionosphere on radio-wave propagation are those already described, which are due to the normal or regular characteristics of the ionosphere. The modes of variation of those characteristics have been shown to be of a regular and fairly predictable nature. There are some other ionosphere phenomena which are irregular in their nature and make radio phenomena in general much less predictable. Five types of such phenomena have been identified; sporadic *E*-layer transmission, scattered reflections, sudden ionosphere disturbances, prolonged periods of low-layer absorption, and ionosphere storms. The last three are probably due to irregular radiations of various types from the sun. The nature and origin of the first two are less well known; study of them must consider diffusion processes in the ionosphere as well as emanations from the sun and stars, meteors, and perhaps other agencies. The last three are primarily due to irregularities in time, while the first two are primarily due to irregularities in space; the space irregularities are patches or "clouds" in the ionosphere.

It is only recently that these irregularities have been well enough identified to be distinguishable from one another and from some of the regular ionosphere variations such as changes of critical frequency and consequent change of layer in radio transmission. This is another reason, besides the one cited in section II, why the present is a good time to summarize the facts of the ionosphere and their effects in radio transmission.

Sporadic E

It sometimes happens that waves are reflected by the *E* layer on frequencies higher than that at which the *E*-layer waves normally disappear and the reflection of waves by higher layers begins; for instance, in the example shown in figure 3 waves may sometimes be reflected at the *E*-layer height of 110 kilometers by frequencies higher than 3,000 kilocycles per second. These reflections are due to a different process than the normal reflection in the ionized layer; the process is probably one of reflection from a sharp boundary of stratified ionization. The existence of these "sporadic *E*" reflections necessitates a redefinition of the term "critical frequency," previously defined as the highest frequency at which signals are received back from the layer. When sporadic-*E* reflections occur they may be received simultaneously with reflections from higher layers; thus, for example, in the case shown in figure 3, vertical-incidence reflections might be received at 8,000 kilocycles per second from both the *E* and the F_2 layers. The *E*-layer critical frequency, more precisely defined, is the value (3,000 kilocycles per second in

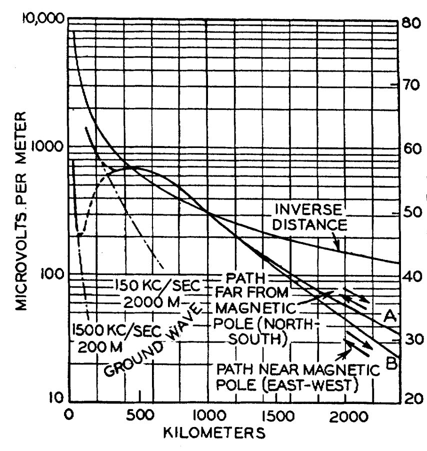


Figure 22. Night field intensity on frequencies from 150 to 1,500 kilocycles per second, for propagation over land of average conductivity

$\sigma = 4 \times 10^{-13}$ electromagnetic units

Quasi-maximum for one kilowatt radiated

For the significance and limits of application of these curves, see text figure 3) at which the observed virtual height shows a sudden rise to very large values as the frequency is increased. Except for the occurrence of sporadic-E reflections, all waves of higher frequency pass through the E layer and are not reflected by it.

The action of sporadic E may be analogous to partial reflection in optics while the regular ionosphere layer action is analogous to total reflection in optics. The sporadic E is confined to limited regions, like clouds or patches, in the E layer which have a very sharp boundary. These patches may be from perhaps one kilometer to several hundred kilometers in extent. Waves reaching them are reflected, even when of frequencies much higher than the E critical frequency. Sporadic E is thus patchy in space as well as sporadic in time, so that its name is well justified.

Sporadic E leads to interesting results in radio transmission. It accounts for long-distance transmission up to higher frequencies than by any other means. The maximum vertical-incidence frequency for which strong reflections by sporadic E have been found is about 12 megacycles per second. By reason of the large angles of incidence possible with the E layer, this has made long-distance communication possible on frequencies as high as 60 megacycles per second. Such communication is generally for only a short time and for restricted localities. For example, on a particular day a patch of sporadic E intense enough to reflect 56-megacycle transmissions was found to be somewhat larger than the state of South Carolina, as shown in figure 23; transmissions from distant points by a single hop, at such distances from a receiving point that they were reflected from this patch over the general vicinity of South Carolina, were received on frequencies as high as 56 megacycles per second. Transmissions reflected from outside this patch were limited to lower frequencies determined by the normal ionosphere ionization.

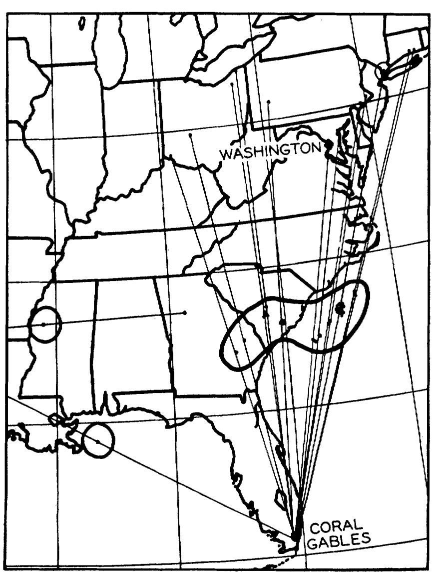


Figure 23. Location of large patch of sporadic E by reception of numerous transmitting stations over long distances

Points of reflection for 56-megacycle transmissions on May 15, 1938, reported by W4EDD, Coral Gables, Fla.

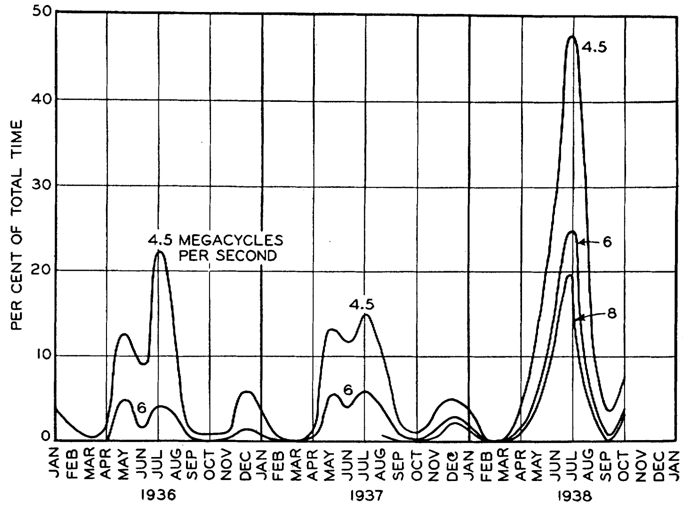


Figure 24. Variation of the prevalence of sporadic E, 1936 to 1938

Per cent of total time of observations on 4, 5, 6, and 8 megacycles per second; 6-megacycle observations begin April 1936, 8-megacycle observations begin August 1937

Sporadic E occurs most commonly in the summer, particularly in the morning and evening but may occur any time of day or night. It occurs occasionally at all seasons, particularly in the evening. Detailed information on its prevalence is published each month in *Proceedings* of the Institute of Radio Engineers by the National Bureau of Standards. A summary of known data is given in figure 24, expressed in terms of vertical-incidence transmission. The figure shows its year-round occurrence, its greater prevalence in May, June, July, and August, and its variation from year to year. The prevalence of sporadic E has not increased uniformly with the advancing solar cycle, like the regular characteristics of the ionosphere, but it is believed to be more prevalent at sunspot maximum than sunspot minimum. Its occurrence is not correlated with other types of ionosphere irregularities nor with thunderstorms or other known phenomena. There is evidence that it occurs more at high than at equatorial latitudes.

Scattered Reflections

An irregular type of reflection from the ionosphere occurs at all seasons and both day and night. These reflections are most noticeable within the skip zone, or at frequencies higher than those normally receivable from the regular layers. Like sporadic E, they occur at frequencies which may exceed the F_2 critical frequencies, but are unlike sporadic E in that they are complex thus causing signal distortion. They are almost useless for communication purposes. Some types of them are of very weak intensity. The scattered reflections are characterized by very great virtual heights, usually somewhere from 400 to 1,500 kilometers. Their occurrence was for a time thought to indicate the existence of another layer above the F_2 layer which might be called the G layer. It is now, however, thought that they are of several types, and that some of them are due to complex reflections from small, ephemeral, scattered patches of ionization in or between the normal ionosphere layers. It has been suggested that some types of these ephemeral patches of ionization may be due to irregular radiations from the stars. Scattered reflections are shown in portions of figures 27, 28, and 30, as explained in the text referring to each.

Sudden Ionosphere Disturbances

The most startling of all the irregularities of the ionosphere and of radio wave transmission is the sudden type of disturbance manifested by a radio fade-out. This phenomenon is the result of a burst of ionizing radiation from a bright chromospheric eruption on the sun, causing a sudden abnormal increase in the ionization of a portion of the ionosphere below the E layer, frequently with resultant disturbances in terrestrial magnetism and earth currents as well as radio transmission. The radio effect is the sudden disappearance of radio signals received on high frequencies.

The diminution of the radio signals to zero usually occurs within a minute. The effects occur simultaneously throughout the hemisphere illuminated by the sun, and do not occur at night. The effects last from about ten minutes to an hour or more, the occurrences of greater intensity in general producing effects of longer duration. The effects are more intense, and last longer, the lower the frequency in the high-frequency range (that is, the range from about 1,500 kilocycles per second up). The radio, magnetic, and other effects are markedly different from other types of changes in these quantities. The effects are most intense in that region of the earth where the sun's radiation is perpendicular, that is, greater at noon than at other times of day and greater in equatorial than in higher latitudes.

The effect is most striking. Frequently all "static" as well as the radio signals disappear. Many a radio operator has taken his radio receiver apart, thinking that some wire had become disconnected, and many a time it has been thought that a fuse had blown in the station, when one of these sudden fadeouts occurred.

An example showing both radio and magnetic effects is shown in figure 25. The four records of received field intensity from distant stations show that the radio intensity suddenly dropped from normal intensity to zero at 1758 Greenwich meridian time, that is, 12:58 p.m., Eastern standard time. This completely wiped out radio transmission throughout the hemisphere; reports to that effect were received from many points in the United States, also Europe and Japan. As shown, the effect lasted much longer on 6,060 kilocycles per second than on 9,570 kilocycles per second at about the same distance. It did not last longer on 9,570 than on 13,525 or 15,625 kilocycles per second because the distance (and the angle of incidence) was greater in the latter cases. As has been noted previously in other regards, effects on a given frequency for a short-distance path correspond to those on a higher frequency for a long-distance path.

Taking due account of the variation of the effects with frequency and distance, varying effects in differing directions can be explained. Reception in the United States from stations in the southern hemisphere usually exhibits greater effects than reception from other directions (because of passing the equatorial regions). Similarly, when the disturbance occurs at a time when it is morning at the receiving point the effects are usually greater in reception from the east than from the west, and vice versa for the afternoon (because of passing the region where it is noon). A radio fade-out sometimes occurs when it is night at the receiving point, but only when the path of the wave is somewhere in daylight.

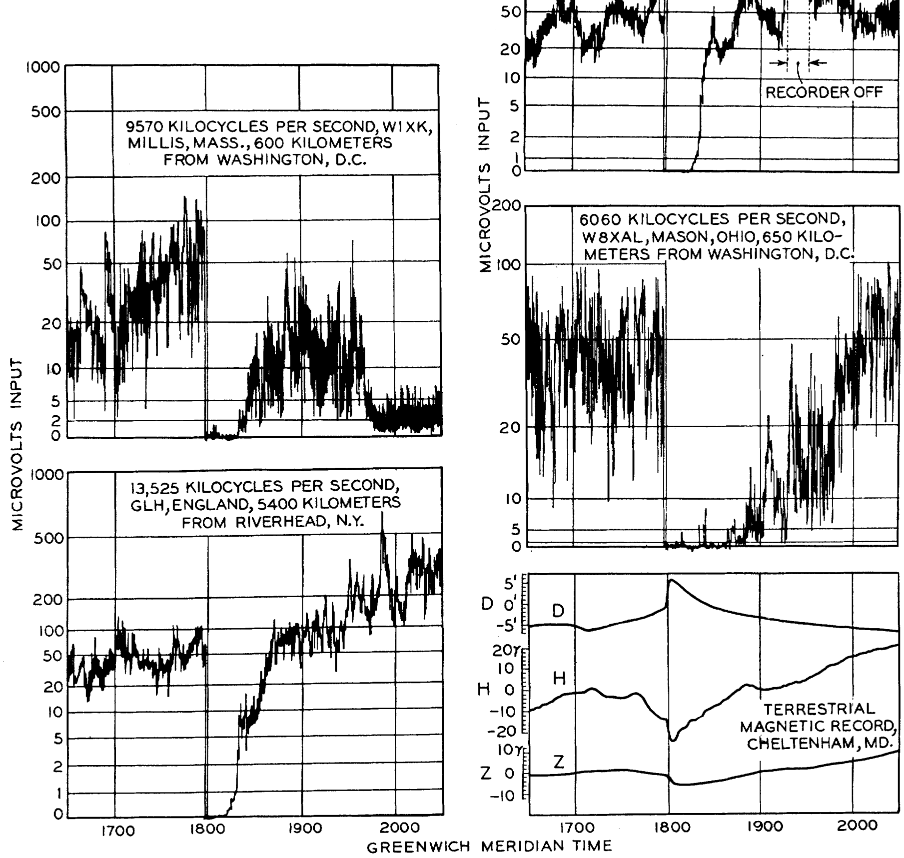
The cause of the sudden disappearance of radio signals is the sudden production, by a burst of ionizing radiation from the sun, of abnormally great ionization below the E layer. This causes abnormally great absorption of radio waves passing through this ionized region on their way up to and down from the regular reflecting layers. The ionization of the regular reflecting layers is not affected.

There is some evidence that received waves of broadcast and lower frequencies increase rather than decrease in intensity during one of these occurrences. As such waves are reflected by, instead of passing through, an ionized layer of the atmosphere below the E layer, this is consistent with the explanation of the disturbance as due to increase of ionization below the E layer. It is not known at just what height the ionization is produced, but it may be in the D layer, mentioned in section III as the layer responsible for daytime broadcast transmission in the winter. The height of the D layer is probably from 50 to 90 kilometers at different times.

The reason this abnormal ionization is produced low in the atmosphere and not in the regular reflecting layers (E, F, etc.) is doubtless because the solar eruptions emit some radiation of different frequencies from those of the steady radiations which maintain the ionization of the E and higher layers. This abnormal radiation is probably of such a frequency as to pass readily through the E and higher layers and be absorbed by the ozone which exists at heights from about 15 to 60 kilometers. The frequency of this radiation is presumably in the ultraviolet, nearer to the optical frequencies than those which produce the regular ionization of the E and higher layers.

Study of these effects is arousing great interest and focusing new effort upon the study of the sun. The sudden ionosphere disturbance is the only known instance in which a specific happening on the earth follows directly from a specific random happening on the sun or other heavenly body. The ionosphere gives information about some of the radiations from the sun which can be

Figure 25. Effects of sudden ionosphere disturbance on May 28, 1936, as revealed by radio fade-out and terrestrial magnetic perturbation



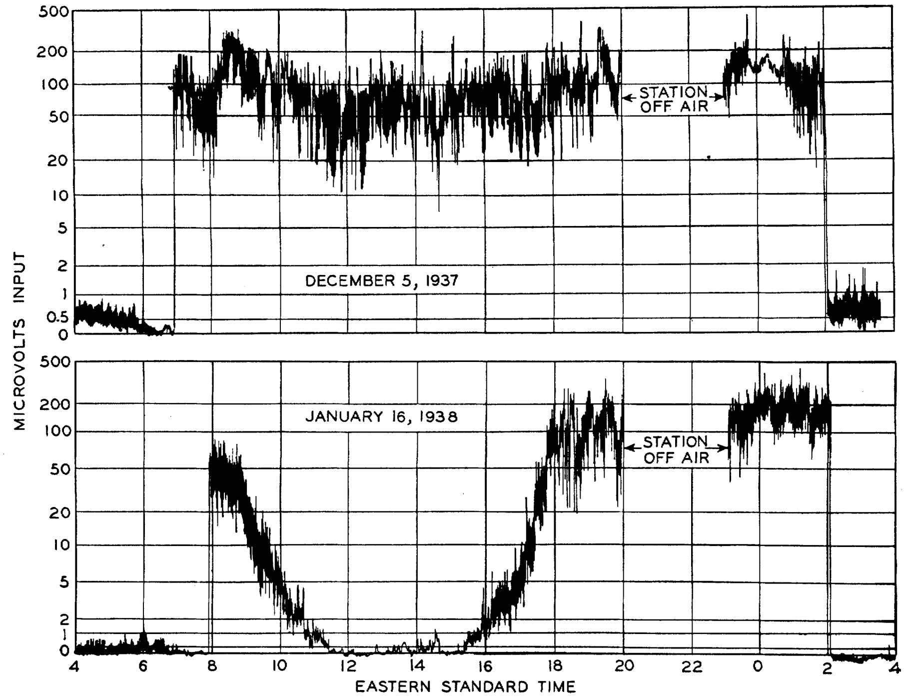


Figure 26. Comparison of recorded field intensity on normal undisturbed day and on day of prolonged period of low-layer absorption

W8XAL, Mason, Ohio, 6,060 kilocycles per second; 650 kilometers from Washington, D. C.

studied in no other way because they are wholly absorbed in the ionosphere and do not reach the earth's surface. The solar eruptions in particular can be studied very well by the aid of the sudden ionosphere disturbances which they cause. Such study may eventually elucidate the nature of the eruptive processes within the sun and the causes of sunspots and the 11-year cycle.

While the evidence indicates that every sudden ionosphere disturbance is accompanied by a solar eruption, the converse does not appear to be true. And there is no reason to suppose that every solar eruption would emit radiation of the particular frequencies which penetrate through the earth's ionosphere to the *D* layer. Probably many eruptions rise high enough in the solar atmosphere to permit the escape of visible light but not high enough to permit the escape of this ultraviolet radiation.

There is no seasonal variation in the occurrence of the sudden ionosphere disturbances or the solar eruptions which cause them. The solar eruptions produce the effect regardless of location on the sun's surface. An eruption usually, but not always, takes place near an active sunspot group. Most of the eruptions which produce sudden disturbances of the ionosphere are much brighter than the average eruption.

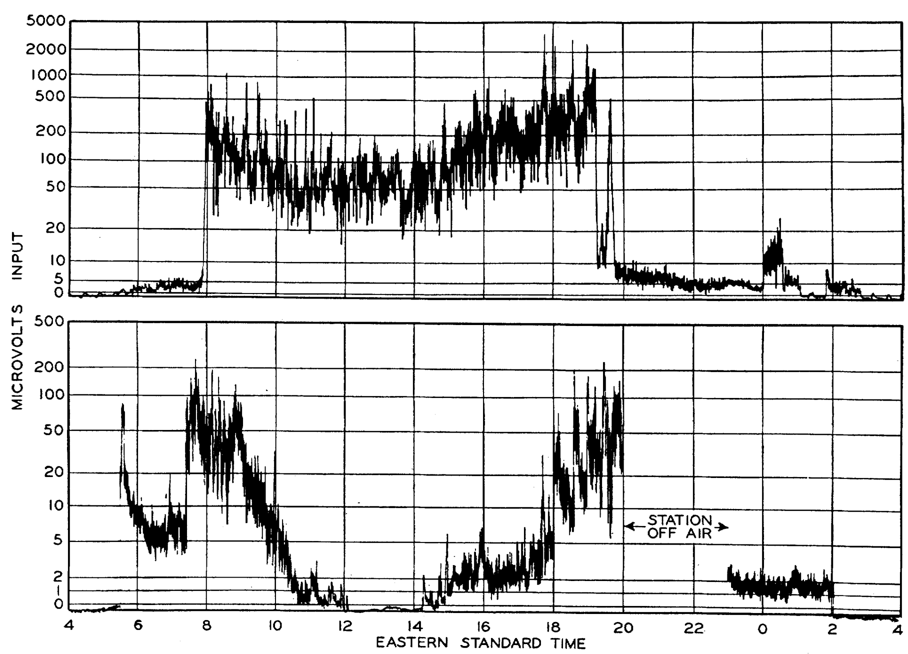
There were 17 known instances of the sudden ionosphere disturbance in the year 1935, 103 in 1936, and 220 in 1937. There was a similar increase of solar eruptions in these years, which were also years of increasing sunspot numbers. The variation of number of sudden ionosphere disturbances from week to week, and from month to month, corresponds fairly well with the number of solar eruptions, but not with the number of sunspots. Thus, just as we found in the case of the ionosphere critical frequencies (figure 10), the sunspot numbers show a year-to-year but not a short-time correlation.

Figure 27. Effects of prolonged period of low-layer absorption on two different frequencies

January 27, 1938

Top—W1XK, Millis, Mass., 9,570 kilocycles per second; 600 kilometers from Washington, D. C.

Bottom—W8XAL, Mason, Ohio, 6,060 kilocycles per second; 650 kilometers from Washington, D. C.



Prolonged Periods of Low-Layer Absorption

This phenomenon is similar to the sudden ionosphere disturbance in its effects and characteristics except that its beginning as well as recovery is gradual and it has a longer time duration, commonly several hours. The intensity diminution is in general not as severe as in the more intense fade-outs, but sometimes the intensities at the medium high frequencies fall to zero.

The phenomenon is illustrated in figure 26, which shows two field-intensity records of transmissions from a station on a frequency of 6,060 kilocycles per second, distant 650 kilometers from the point of reception. The upper graph, for December 5, 1937, is a record of a normal day. The average intensity goes down slightly during the middle of the day, indicating somewhat more absorption during the middle of the day than during the morning and evening hours. (The very low intensities at each end of each graph represent "static.") The lower graph, January 16, 1938, shows a prolonged period of low-layer absorption. Here the intensity falls gradually from the normal value at 0800 (a.m.) to zero, and is zero or very low for several hours, rising again to normal value at 1800 (6 p.m.). Except for the gradual beginning the effect is similar to the sudden ionosphere disturbances.

The different effects of this phenomenon at different frequencies are shown in figure 27. Both records are for the same day. The upper record, for a station on 9,570 kilocycles per second, shows some absorption, that is, reduction of intensity in the middle of the day. The lower record, for a station on a lower frequency (6,060 kilocycles per second) at about the same distance, shows very much greater absorption in the middle of the day. Thus, just as in the sudden ionosphere disturbances, the effects are less at higher frequencies if distance and other conditions are the same.

(Other phenomena shown in this figure are as follows. In the top graph, for W1XK: scattered reflections in the early morning until about 0800, then abrupt beginning of *F*-layer transmission as the ionization increases, abrupt failure of *F*-layer transmission as the ionization decreases at about 1910, then scattered reflections. The bottom graph, for W8XAL, shows: a burst of *F* layer at 0530, then scattered reflections until 0730, transmission until 1600, then F_2-layer transmission until some unknown time after 2000, then scattered reflections until 0200.)

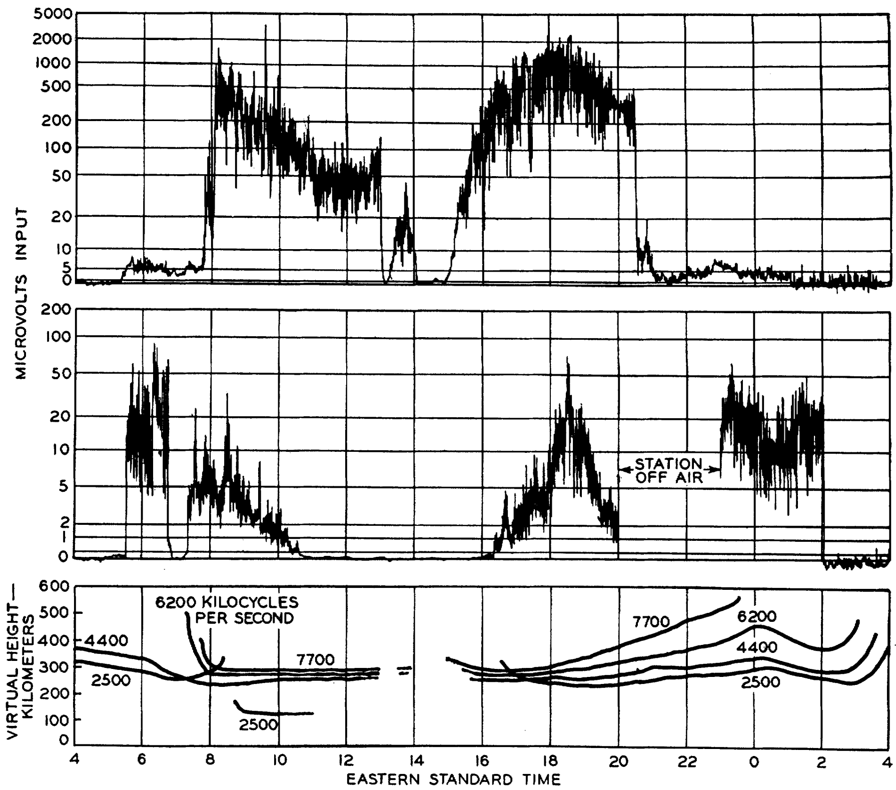


Figure 28. Occurrence of sudden ionosphere disturbances during a prolonged period of low-layer absorption

January 20, 1938

Top—W1XK, Millis, Mass., 9,570 kilocycles per second; 600 kilometers from Washington, D. C.

Center—W8XAL, Mason, Ohio, 6,060 kilocycles per second; 650 kilometers from Washington, D. C.

Figure 28 shows that the simultaneous occurrence of low-layer absorption and sudden ionosphere disturbances is possible; in fact it shows two severe fade-outs happening during a prolonged period of low-layer absorption. The lower field-intensity record, for W8XAL, shows a period of low-layer absorption which reduced the received intensity to zero for several hours. Because this had already gone to zero it could not indicate the two intense fade-outs which occurred during this period. They appear, however, on the upper record, which is for a station of higher frequency (W1XK), because the low-layer absorption did not reduce the intensity at this higher frequency to zero. The simultaneous occurrence of these two phenomena was fortuitous.

The equivalent vertical-incidence frequency for station W8XAL was about 2,300 kilocycles per second and for W1XK was about 6,500 kilocycles per second. The reason that the vertical-incidence critical frequencies are not proportional to the transmission frequencies over nearly the same distance is that the two transmissions were reflected from layers of different heights, *F* and *E* layers respectively. This figure illustrates the value of the concept of equivalent vertical-incidence frequency in evaluating absorption over long paths. Records of vertical-incidence reflections at several frequencies are shown at the bottom of the figure. The reflections at 2,500 kilocycles per second were lost between 1100 and 1630 Eastern standard time because of low-layer absorption, so the fade-outs could not be shown at this frequency. The reflections at the higher frequencies were not eliminated by the low-layer absorption, and the loss of reflections between about 1300 and 1500 Eastern standard time indicates the fade-outs.

(Other phenomena shown in figure 28 are as follows. In the top graph, for W1XK: "static" from 0400 to 0530, then scattered reflections until about 0800, then F_2 transmission until 2030, then scattered reflections until 0100, then "static." In the bottom graph, for W8XAL, are shown: "static" from 0400 to 0530, then *F* transmission until 0615, then *E* transmission until 1820, then *F* transmission until 0200, then "static.")

The absorption causing the low-layer absorption effect appears to be due to ionization in a part of the ionosphere below the *E* layer, exactly as for the sudden ionosphere disturbances. The ionization is caused by an abnormally great outpouring of ultraviolet light from the sun, but in this case it is not so sudden as in the eruptions which cause the sudden ionosphere disturbances. The variation of the effects with frequency, and other characteristics, are the same as for the sudden ionosphere disturbances.

Both phenomena occur at all seasons, but the prolonged periods of low-layer absorption have been found to occur in a group of several weeks duration at periods of high sunspot activity, the groups being separated by more or less quiet periods of several months. They frequently but not always occur during periods when sudden disturbances of the ionosphere are numerous. They seem to occur more during years of large solar activity than at sunspot minimum.

Ionosphere Storms

An ionosphere storm is a period of poor radio transmission (except for the low frequencies, below 500 kilocycles per second, which are sometimes improved) lasting a day or more, and usually accompanied by a magnetic storm, that is, a period of unusual fluctuation of terrestrial magnetic intensity. It has two phases, an initial turbulent phase and a following moderate phase. Usually only the second phase occurs in medium and low latitudes. The initial turbulent phase is the cause of the moderate phase which follows, but is confined to the auroral zone, that is, the region around the magnetic pole in which aurora is visible and which is usually limited to within about 20 degrees of the magnetic pole, which region is greatly extended in very severe storms.

The turbulent phase consists of a violent boiling or turbulence of the entire ionosphere in the auroral zone, resulting in irregularly moving small clouds of ionization and a disintegration of the normal stratification of the ionosphere from the E layer on up. Whatever causes the storm apparently plunges into the ionosphere at auroral-zone latitudes, and literally tears it up. On the rare occasions when the auroral zone has extended as far south as Washington, an increase in F-layer ionization has been observed to precede the turbulent phase. This is consistent with the idea that the carrier of the energy of the ionosphere storm, when it first entered the high ionosphere, caused an increase in ionization. No consistent increase in F-layer ionization has been observed to precede the ordinary less severe storms, when the auroral zone did not extend as far south as Washington.

During the turbulent period of the ionosphere storm, high-frequency transmissions are very erratic, both signals and "static" surge violently, being transmitted with good intensity for short intermittent periods, interspersed with periods of complete failure. This indicates severe turbulence in the ionosphere with small unstable patches or clouds of high ionization densities. Such clouds may not be directly over the midpoint of the great-circle path and the transmission may not even be by a great-circle path.

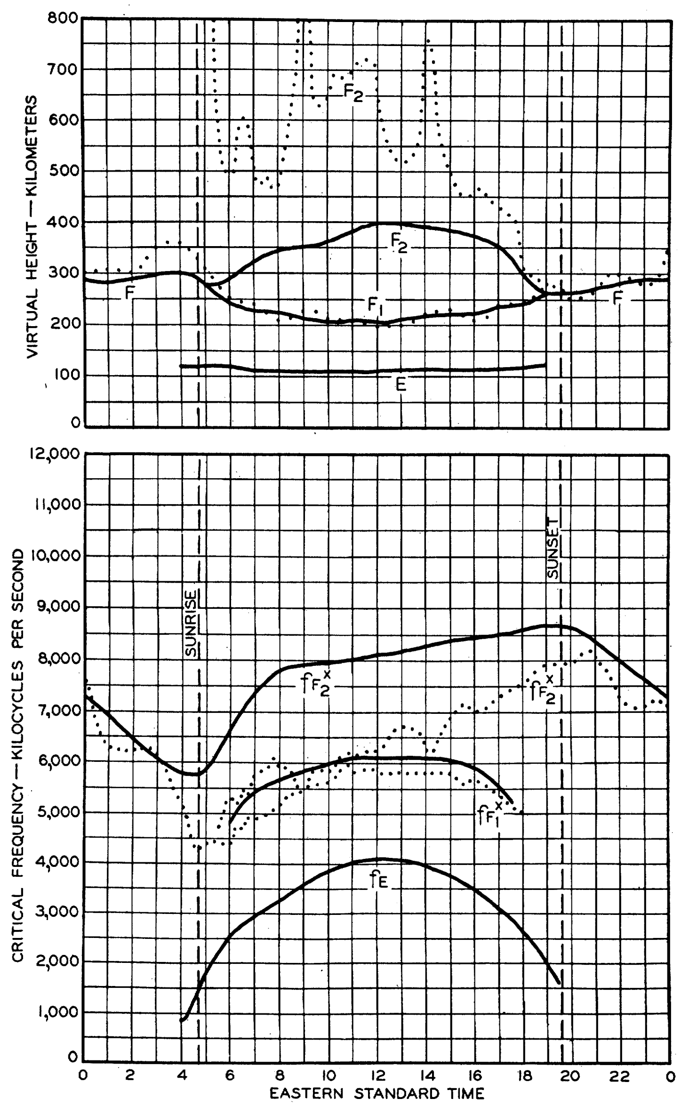


Figure 29. Comparison of ionosphere characteristics on a day of ionosphere storm and normal undisturbed days

June 1938

Solid curves—Average for undisturbed days

Dotted curves — June 8

The moderate phase, following the turbulent phase of an ionosphere storm, is characterized by an expansion and diffusion of the higher F region, extending into latitudes farther from the auroral zone, the greater the intensity of the storm. This expansion and diffusion of the ionosphere increases the virtual heights and lowers the ionization densities. This results in abnormally low critical frequencies and abnormally great virtual heights of the night F and daytime F_2 layers, and to a less extent of the daytime F_1 layer, and also increases the absorption, that is, reduces received intensities. (When the ionosphere storm occurs in a winter day, the normally unobservable F_1 layer appears.) The night F layer is more complex and turbulent than normal. Increased absorption of the x components of the daytime F_1 and F_2 layers is especially noticeable. The E layer is usually not appreciably affected. This moderate phase of the ionosphere storm extends to the latitude of Washington much more frequently than the turbulent phase. It lags behind the severe effects of the associated magnetic storm by several hours. During the moderate phase the ionosphere gradually returns to normal conditions, but this recovery also lags behind the associated magnetic storm's recovery.

The maximum usable frequencies for night F-layer and daytime F_2-layer transmissions are much reduced because of the lowered critical frequencies and increased virtual heights. Thus the higher frequencies are not usable. Frequencies low enough to be received are usually abnormally absorbed, especially during the daytime. There is usually increased fading and instability of transmissions over night paths. Sky-wave field intensities at broadcast frequencies rise much later at night and reach values much lower than normal.

Ionosphere storm effects diminish greatly with distance from the magnetic pole. Transmissions reflected from the ionosphere south of a radio receiving point may often be received satisfactorily while those reflected from the ionosphere farther north are not. Sometimes there appears to be a fairly sharp line of cleavage between the disturbed region (to the north) and the undisturbed (to the south). Transmissions which pass through the disturbed regions in the ionosphere are affected regardless of the direction of transmission.

In most ionosphere storms only the

second or moderate phase is experienced at any except very high latitudes. The principal effects of this phase may be summarized as: (*a*) increase of virtual heights of F, F_1, and F_2 layers, (*b*) decrease of critical frequencies of the same layers, (*c*) greater sharpness of F_1 critical frequency (*d*) decrease of maximum usable frequencies, (*e*) increase of skip distances, (*f*) increase of absorption (that is, decrease of received intensities). Effects (*a*) and (*b*) are illustrated in figure 29, which shows virtual heights and critical frequencies of a day of ionosphere storm (June 8, 1939) in comparison with the month's average of undisturbed days. The E layer was the same on June 8 as on the undisturbed days. Effects (*d*), (*e*), and (*f*) are illustrated in figure 30, which shows (at top) a graphical record of received intensity for a normal undisturbed day and (at bottom) a similar graph for a day of severe ionosphere storm; on the day of ionosphere storm the intensities were extremely low, even for the "static."

(Other phenomena shown in figure 30 are as follows. In the top graph: "static" from 0400 to 0600, then scattered reflections until 0700, then F_2 transmission until 2200, then scattered reflections until 0100, then "static" until 0400. In the bottom graph are shown: very weak "static" for a few minutes after 0400, then scattered reflections until 0700, then very weak "static" until 1130, then scattered reflections until 0100, then very weak "static." The top graph as recorded had values about three times higher from 1755 to 2205, because of a change of antenna at 1755; it is shown here as it would have been with constant antenna conditions.)

Ionosphere storms (and the magnetic storms that usually accompany them) have several characteristics the opposite of those of sudden ionosphere disturbances (and the magnetic perturbations that sometimes accompany them). The former are more intense the higher the latitude, while the latter are more intense the lower the latitude. The former occur both day and night, and the latter are confined to the day hemisphere. The former last one or more days, the latter usually last less than an hour.

These two types of ionosphere irregularity occur in general independently of one another but both are more likely to occur at times of great sunspot activity. They are more frequent and intense during years of sunspot maximum than sunspot minimum. A group of sudden ionosphere disturbances occurring on successive days is sometimes followed, after several days, by one or more ionosphere storms.

On account of their differences, different procedures are followed in practical radio communication to combat the effects of the two types of disturbance. When a sudden ionosphere disturbance occurs it may be possible to continue communication by shifting to a higher frequency; communication could also be accomplished by using a frequency lower than the broadcast range, but a change to such low frequency would in general be too cumbersome for the short time of the sudden ionosphere disturbance. During the turbulent phase of an ionosphere storm the only way to assure dependable communication is to use a frequency below the broadcast range. During the moderate phase of an ionosphere storm it may be possible to continue communication by shifting to a lower frequency within the high-frequency range.

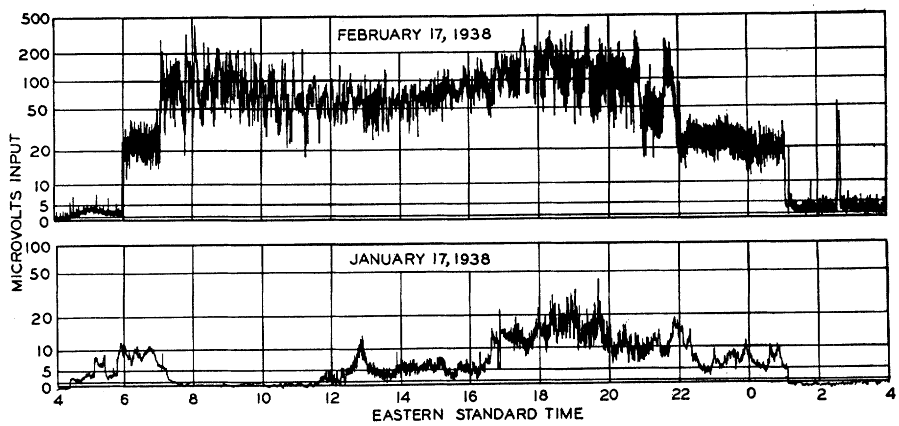


Figure 30. Comparison of received intensities on a day of ionosphere storm and on a normal undisturbed day

W1XK, Millis, Mass., 9,570 kilocycles per second; 600 kilometers from Washington, D. C.

V. Conclusion

The ionosphere is a new world to which radio research and radio operations have given us access in the past few years. A broad survey of this vast territory has been given in the present paper; many important features of the new territory have been neglected or barely mentioned.

I have summarized the work of many men and institutions. The nature of the paper forbade the individual crediting of most data and results to the originators. Splendid work has been done by the ionosphere pioneers (scientists and engineers) who have created the body of knowledge here reported. I am especially indebted to my colleagues in the National Bureau of Standards upon whose work I have drawn most heavily: S. S. Kirby, T. R. Gilliland, N. Smith, F. R. Gracely, and A. S. Taylor.

Short-distance radio transmission by means of the ground wave is not included in this paper, which is limited to the sky wave. Thus the important domains of very low frequencies and ultrahigh frequencies are not covered. The ground wave is calculable and is relatively unvarying with time; its phenomena are not so complicated or interesting as the sky waves. It is by use of the sky waves that nearly all long-distance radio communication is carried on.

This paper also does not include the subject of atmospheric disturbances (that is, natural electrical noise or "static"). They are themselves radio waves, originating principally in distant lightning flashes, and propagated by the same mechanisms as other radio waves. As their effects are largely produced by waves traveling in the ionosphere, much of the information presented in this paper is applicable to their study.

This paper has shown how sky-wave transmission is determined by, and calculable from, the heights and ionization densities and other properties of the ionosphere layers. The maximum usable frequencies at any distance, for instance, are directly determinable from the virtual heights and critical frequencies measured in vertical-incidence experiments. Optimum frequencies may be similarly estimated, though not as certainly as the maximum usable frequencies. The received intensities of the waves may be estimated to a certain extent from ionosphere data, but much more extensive data are needed for this purpose. Study of the behavior of the ionosphere during the five types of irregularities or anomalies discussed in section IV greatly clarifies our understanding of certain radio phenomena, and leads to knowledge of how to overcome transmission difficulties.

Examination of the facts and relations of radio wave propagation via the several layers ot the ionosphere should give ample warning that it is hazardous to generalize about good and bad radio reception. Any conclusion must take into account the heights and ionization densities of the several layers concerned, the absorption at the various levels through which the waves pass, the time of day, the season, the epoch of the sun-spot cycle, the distance of transmission and angle of take-off of the waves and angle of incidence at the ionosphere layer, the latitudes and propinquity of the transmission path to the magnetic pole, and the occurrence of ionosphere disturbances and irregularities.

The more one views the complexities of radio transmission via the ionosphere, the more he marvels that it provides any intelligible communication. However, as the facts of the ionosphere become better known, and the mechanism of reflection of radio waves from the ionosphere layers is more fully worked out, it becomes more nearly possible to assure long-distance radio transmission at all times. To this end it is fortunate that a beginning has been made on an ionosphere-data reporting service. As this is extended, it will be easier to predict radio-transmission conditions for a given time and path. The reliability of such prediction should surpass that of weather, for the controlling factors are better known and more uniform. Both weather and ionosphere phenomena are due primarily to the sun, but the sun's effects are more direct and more uniform over the earth for the ionosphere than for weather. The weather, incidentally, has no relation to ionosphere phenomena, being produced in much lower regions of the atmosphere.

Study of the ionosphere not only provides means of improving radio services, but is also advancing other branches of knowledge. It is furnishing an explanation for the variations of terrestrial magnetism, hitherto a great mystery. It supplies a way of studying various types of radiations from the sun, many of which do not reach the earth's surface because of being absorbed in the earth's atmosphere. These advances are believed to be only the beginning of gains which future exploration of the ionosphere will bring forth in abundance.

Resonant-Type Constant-Current Regulators

ROLAND R. MINER
MEMBER AIEE

Synopsis: This paper deals with the fundamentals of resonant regulators for constant current, and compares them with other kinds of regulators. It also discusses the various forms such regulators may take, their applications and control, and protective equipment required for them. A simple method of predicting performance by means of the voltage diagram is developed.

THE constant-current regulator is one piece of equipment that has been given little attention for some years. At a recent meeting of a Section of the Institute it was apparent, when the subject of constant-current regulators was mentioned, that many engineers present had only a very hazy idea of what they were. A description of the resonant type of regulator, now in its fourth decade, was received with considerable interest as an entirely fresh subject. Here, then, is a "Rip Van Winkle" who has come back to renew acquaintance with some of those other hoary oldsters that have reappeared in new clothes and increased stature. In that honorable company we find the step-type voltage regulator, the wound-core transformer, the overhead ground wire, and the spark-gap lightning arrester.

The term, constant-current regulator, as used here means a device that is used to supply a constant current to an electric system in which the receiving devices are in series, and in which the applied voltage must be varied in proportion to the impedance of the devices in use at any particular time. It is assumed that the regulation is automatic. The discussion is confined to a-c regulators.

Types of Constant-Current Regulators

A crude form of regulator can be made by inserting a large reactance in series with a number of resistance loads. The reactance consumes very little energy, but is such a large part of the total impedance that resistance loads can be connected into and out of the circuit without making large changes in the current. The fixed series reactance does not give current sufficiently constant for most purposes, and it is not widely used.

A variable reactance, either of inductance or capacity, inserted in series with the load and so designed that it automatically adjusts its reactance until the desired current flows in the circuit is the type of regulator in common use today. A new form of regulator of this type, sometimes known as a semiresonant or nonlinear network type, is one in which a saturable reactor in parallel with a capacitor of proper size behaves in the circuit like a variable capacitor, and automatically adjusts its effective reactance until a certain current flows.[1] The more common form is the moving-coil transformer that changes its leakage reactance. The moving coil is repelled from the other coil by the reaction of the currents in the windings. When properly counterbalanced, the coils will approach each other until the leakage reactance is just sufficient to pass the current for which the counterbalance weights are adjusted. In spite of the wide variance in physical form of these two regulators, they are essentially the same. Each is a variable reactance in series with the load.

The inherent characteristics of the regulators composed of a variable reactance in series with the load may be summarized for purposes of comparison as follows. The input current as well as the output current is constant. Consequently, the volt-ampere input to this type of regulator is constant and appreciably greater than the full load volt-ampere output. The power factor is comparatively low. The load circuit is reactive in character. In case of a break in it a very tenacious arc is drawn which frequently causes damage to equipment or contacts. The input circuit is also reactive in character, so switches and other control devices must be capable of breaking the reactive currents. The

Paper number 39-82, recommended by the AIEE committee on electrical machinery, and presented at the AIEE South West District meeting, Houston, Tex., April 17-19, 1939. Manuscript submitted February 4, 1939; made available for preprinting March 13, 1939.

ROLAND R. MINER is electrical engineer with the Kansas Gas and Electric Company, Wichita.

1. For all numbered references, see list at end of paper.