SECTION 2

ADVANCED
PRACTICAL
RADIO ENGINEERING

TECHNICAL ASSIGNMENT

RADIO WAVE PROPAGATION—RECEIVING ANTENNAS

PART I

Capitol Radio Engineering Institute
Washington, D. C.

2470

- TABLE OF CONTENTS -

RADIO WAVE PROPAGATION

RECEIVING ANTENNAS -- PART I

RADIO WAVE PROPAGATION

RECEIVING ANTENNAS.--PART I

Just as the telephone engineer must be thoroughly familiar with the electrical and physical properties of the wires and cables over which telephone conversations are carried from point to point, so must the radio engineer understand the electrical and physical properties of the space through which his signals are to travel from transmitter to receiver. This space--the earth's atmosphere--through which the radio wave is propagated is an extremely complex and variable medium. Without the atmospheric contents of the space, however, most radio communication as we know it would be impossible because of the curvature of the earth's surface.

Before Marconi proved that a radio wave could follow the curvature of the earth, it was believed that radio waves, like the shorter light wave lengths, traveled essentially in straight lines and could not follow the earth's surface beyond the horizon.

After his experimental results were announced, scientists began to investigate the question of propagation of radio waves more thoroughly, and in 1909 Arnold Sommerfeld, a German physicist, published a celebrated paper on he radiation from a short vertical antenna at the surface of the earth, and the effect of the finite conductivity of the ground upon the radiation. Sommerfeld's analysis indicated that (1) energy was directly propagated in a straight line from the transmitting to the receiving antenna (direct ray); (2) energy was propagated by means of a ray that was directed to the earth, reflected from it, and then impinged upon the receiving antenna (ground-reflected ray); (3) and finally, energy was propagated by means of a wave that moved along the surface of the earth, and hence was called a surface wave. *This wave could follow the curvature of the earth, and hence could explain reception beyond the line-of-sight.*

The direct and ground-reflected rays at the low frequencies then employed tended to cancel one another by the process of destructive interference (mentioned in previous lessons), so that at those frequencies it appeared that only the surface wave was instrumental in affording communication.

Marconi found that the lower the frequency, the farther the wave could travel before it was attenuated to a degree such that it was too weak to override the natural static of the atmosphere, and so lower and lower frequencies were employed for commercial purposes, and the higher frequencies were allotted to amateur use. Sommerfeld's analysis seemed to justify this allocation: the surface wave was attenuated by the resistance losses in the earth, but the effect was less at low frequencies than at high frequencies, and as a result low-frequency ground waves traveled over large distances before being attenuated to the minimum usable intensity. For example, Station NSS at Annapolis operates at about 17 kc. with a power of several hundred kilowatts. Its radiation is essentially a surface wave which gives reliable reception over thousands of miles practically day and night. In the early days the low-frequency wave was easier to generate at high power than high-frequency waves, and in view of the greater distance of reception, it is easy to see why the trend was to the lower frequencies, in spite of the greater difficulty in radiating and receiving the long waves with antennas of practical size, and in spite of the larger coils and condensers required in both the transmitting and receiving units.

Strangely enough, however, Sommerfeld made a mistake in sign in his formula, which error he subsequently corrected in 1926. The corrected formula was corroborated by Van Der Pohl, Wise, Burrows, Norton, and other scientists, and seemed to indicate that a surface wave did not exist, at least of the type previously considered. However, the last mentioned scientist, K. A. Norton, has shown that one can still consider the radiation from a short vertical antenna near the earth as being made up of the direct ray, the ground-reflected ray, and a surface wave. The result of all these investigations, as well as those of others regarding the ionized layer of air above the earth's surface *(the ionosphere)*, has been that although it is now more fully appreciated how complex radio transmission really is, its very complexity is recognized as an aid rather than as a hindrance in the art of communication, for different frequencies and accompanying phenomena permit different types of services to be utilized, such as broadcasting, point-to-point communication, long distance transmission, etc. Hence, although the second World War has accentuated the the use of micro-waves (very high frequency waves), one must not overlook that *all* frequencies are useful, and are by no means to be neglected. It is the purpose of this lesson to show that *the basic principles of radio communication are the same at all frequencies*, and that the apparently complicated phenomena fortunately simplify down to simple modes of behavior in each frequency range.

MODES OF TRANSMISSION.--The surface wave is theoretically present at all frequencies, and is a result of the reaction of the finite conducting earth upon the energy radiated into the atmosphere near its surface. It is guided by the curved surface of the earth, just as electrical energy is normally guided in an ordinary electrical circuit around bends by the two wires of the circuit. For this reason receiving antennas beyond the horizon of the transmitting antenna (to be explained in greater detail later on) can pick up energy: the surface wave will travel around the bulge of the earth and reach the receiving antenna.

However, the earth that guides the surface wave also absorbs a certain amount of energy per unit length, and ultimately attenuates it down to a non-usable degree. The distance that the ground wave can travel before it is attenuated unduly is greater, the greater the conductivity of the earth, and the lower the frequency. For example, for average earth conditions, a surface wave at 1 mc. (in the standard broadcast range) can travel as much as 100 miles or so before it is attenuated to a value comparable to the sky-wave intensity at that point, whereupon "fading" occurs. Hence, the so-called primary service area of a broadcast station may be as much as 100 miles.

On the other hand, at a high frequency of 46 mc. in the FM band, the surface wave is attenuated so rapidly with distance that it is of no value for even line-of-sight distances. In this frequency range one can disregard the surface wave and instead consider the direct and ground-reflected rays, a combination known as the *space wave*. It will be shown here that at ultra-high frequencies the space wave is the important means of communication, whereas at low frequencies, such as in the standard broadcast band, the surface wave is the important factor.

Under certain conditions, however, the space wave may be of importance even at low frequencies. The surface wave decreases very rapidly in intensity as one moves away from the earth. Thus, in plane-to-plane communication, the surface wave is very weak if the airplanes are flying very high, and are not separated by too great a distance. Fortunately, however, these conditions are favorable for the space wave, so that the latter is available for communication where the surface wave fails to be of use.

Finally, there is the question of long distance communication. It was stated previously that the amateurs were given the then high-frequency part of the spectrum (frequencies in excess of 1.5 mc.). The amateurs diligently began to exploit this frequency range, and soon began to report instances of remarkably long distance reception. Investigation indicated that this was due to *high-angle space waves* that were reflected from or refracted (bent) by an ionized layer of air known as the ionosphere, back to the earth at distances many thousands of miles from the transmitter. In many cases the reflections were concentrated in distant regions of the earth, with no reception on the intervening portions of the earth. This was known as *skip distance*. But one of the most interesting features of this short wave transmission was the low attenuation suffered by the refracted space wave, so that very little transmitter power was required. It was not long before important commercial uses began being made of the short waves, and today they are among our most useful frequencies.

THE GROUND WAVE.--From the foregoing discussion it is evident that transmission over short and moderate* distances is due solely to the *ground wave:* at low frequencies due to the *surface wave component;* at high frequencies due to the *space wave component*. Long distance transmission, on the other hand, is due to reflection or refraction of the *space wave* from the upper regions of the atmosphere (the ionosphere), although *very low* frequency surface waves can

---

* A tabulated list of distances for various frequencies is given farther on in this lesson.

also travel long distances, and are therefore used for this purpose. In the latter case it appears that the surface wave is "guided," as it were, between two concentric conducting spheres, the earth and the surrounding ionosphere.

The lower portions of the atmosphere, known as the troposphere, are also capable of bending or refracting radio waves, particularly in the u.h.f. range, and have an effect upon ground wave propagation. Further mention of this will be made later on in this lesson.

The ground wave will be considered first, particularly the space wave component. In Fig. 1 is shown a short transmitting antenna T, h_1 units above the earth's surface, and a receiving antenna R, h_2 units above the surface of the earth, The direct ray proceeds in a straight line between T and R, as shown, while the ground-reflected ray proceeds to the earth at an angle of incidence ψ_2, and is reflected to R at the same angle ψ_2 according to the law of reflection that the angle of reflection equals the angle of incidence.

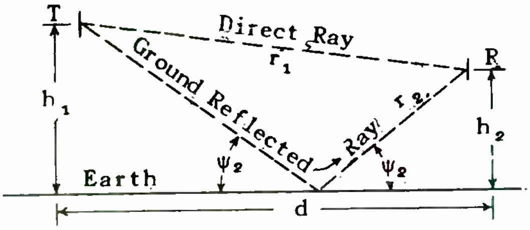


Fig. 1

At the point of reflection from the earth, however, the electric field component can undergo a change in the phase that may affect the magnitude of the resultant at R of the direct and ground-reflected rays. This matter is so important that some mention at this point must be made of the characteristics of reflection, although the matter will be discussed in more detail in a later lesson.

The reflection depends upon the angle of incidence ψ_2, the conductivity and dielectric constant of the earth, the frequency, and the polarization of the wave. In most cases the earth's conductivity is the important factor at low frequencies, and the dielectric constant at high frequencies. As regards polarization, for a horizontally polarized wave the electric field vector undergoes a 180° reversal of phase for all values of ψ_2, but a vertically polarized wave undergoes a 180° phase reversal if the angle of incidence ψ_2 is small, but tends to remain in phase if ψ_2 is large. Thus, if in Fig. 1,

ψ_2 is near "grazing" (close to zero) then upon reflection, the electric field vector of the ground-reflected ray is shifted through 180° in time phase, and hence tends to meet the direct ray at R in *phase opposition.*

Normally distance d is many times h_1 or h_2, so that ψ_2 is small--on the order of a few degrees--and a phase reversal upon reflection takes place. However, there is a further factor that must be taken into account, and that is the difference in path length between the two rays. Thus, the path length for the direct ray is r_1; that for the ground-reflected ray is r_2. The latter wave is delayed in phase with respect to the former by an amount depending upon how much r_2 exceeds r_1 *when measured in wave lengths.* For example, if r_2 exceeds r_1 by $\lambda/2$, then the ground-reflected ray is delayed by 180° with respect to the direct ray. This must be added to the 180° phase shift experienced by the ground-reflected ray owing to reflection, so that the total phase shift is 360°, *or the two rays meet in phase and are additive.*

At the lower frequencies, h_1 and h_2 are small when measured in wave lengths. As a result, the difference $(r_2 - r_1)$ is small when measured in wave lengths, and the phase delay is therefore negligible. Consequently, the ground-reflected ray differs from the direct ray only by the 180° owing to reflection, and so tends to cancel it at R. Hence, at low frequencies the space wave, which is the vector combination of these two components, is negligibly small, and only the surface wave is of importance.

At ultra-high frequencies the surface wave is attenuated so rapidly by the ground losses that its intensity is negligible at but a short distance from the transmitter. Normally it can be disregarded. However, in this frequency range it is possible to elevate antennas by several wave lengths above the earth's surface, (at least the transmitting antenna), and thus it is possible to make r_2 exceed r_1 by an appreciable part of a wave length. This in turn, as shown above, tends to counteract the 180° phase shift due to the reflection and bring the two components of the space wave more nearly into phase, so that the resultant becomes appreciable in magnitude. Thus, in the frequency range where the surface wave fails the engineer, the space wave steps in to provide communication.

In actual practice, large objects such as buildings, hills, etc., may produce several reflections, so that there are several ground-reflected rays of different path lengths to be combined with the direct ray. The resultant is not so easy to compute, but it is evident that it would rarely turn out to be just zero. More important in such cases is that some one of the reflected rays may be appreciably delayed compared to the others, and--in the case of television images--give rise to echo effects (multiple images) on the television screen.

In point-to-point communication, where both the receiver and transmitter antenna heights can be controlled, and a knowledge of the intervening terrain, or possibly even its choice, is possible, it is possible to calculate the resultant voltage, and choose a receiver height, for example, such that the resultant will be a maximum. For ordinary television and FM broadcasting, however, such control of the transmission conditions is not available, and the receiving antenna is located in a region that will avoid, as far as possible, reflections that will produce echo effects and other undesirable conditions.

At this point it is well to stress the fact that in ultra-high frequency transmission there is an even more important reason for elevating the transmitting and receiving antennas, and that is the increase in the line-of-sight transmission distance. Since in this frequency range the surface wave, which can follow a curved surface, is of negligible importance, and since the space wave components travel in straight lines (like light energy), they cannot be intercepted by a receiving antenna that lies much beyond the horizon and is therefore in the shadow of the bulge in the earth owing to its curvature.

Transmission is practical therefore only over distances not much greater than line-of-sight, and such distances are increased by elevating the antennas.* In Fig. 2 is shown a transmitter T at a height h_1, and several receiver posi-

* At frequencies around 40 m.c. reception considerably beyond the line-of-sight distance is possible because of diffraction (spreading of energy from a straight-line path). Diffraction is less in actual distance of spread at the higher frequencies, and hence frequencies around 1000 m.c., for example, are limited practically to line-of-sight paths.

tions, all denoted by the letter R with the appropriate subscript. Any receiver that can be joined to T by means of a straight line without intercepting the earth's surface is on the line-of-sight path from T. Even though the space wave consists of two components, both can reach the receiver over such a line-of-sight path. Thus R_2 can easily receive the space wave, as can other receivers *on the earth* farther and farther away until position R_1 is reached. Evidently this is the greatest line-of-sight path, and a receiver at R_3, for example, is in the shadow of the bulge in the earth's surface, and will not receive energy from T through line-of-sight transmission. This is clearly shown in the figure.

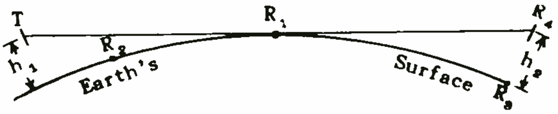


Fig. 2

The distance *along the earth's surface* from T to R_1 is called the horizon distance. It is determined by drawing the tangent from T to the earth's surface. Evidently a whole cone of such tangents can be drawn all around T, and they trace out a circle on the earth's surface passing through R_1 and having T as a center. This circle is known as the horizon line. Its size evidently depends upon the height h_1 of T above the surface of the earth; if h_1 is zero, the horizon is zero. However, even a grounded vertical Marconi antenna has some equivalent height above the earth's surface, and hence some horizon distance. The latter may be so short, however, that the surface wave will be appreciable or even predominating up to that distance.

From Fig. 2 it is evident that the line-of-sight path is unchanged if the receiver and transmitter antennas are interchanged. Hence one can speak of a horizon distance for an elevated *receiving* antenna. Thus, R_4 at the elevation h_2 has a horizon distance up to R_1. Furthermore, for transmitter T and receiver R_4, the total line-of-sight path is clearly the sum of the two horizon distances. The higher the elevation of either or both antennas, the greater is the maximum distance of transmission, and *this is the principal reason for elevating the antennas.*

Nevertheless, if the receiving distance is well within the transmitter's horizon distance, as for example, R_2 in Fig. 2, there is still an advantage in raising R_2 as far above the ground as possible. This is because of the considerations previously presented concerning the increase in path length difference between the direct and ground-reflected rays as the two antennas are elevated, and the consequent veering around of the ground-reflected ray into phase with the direct ray. For example, in the "Standards of Good Engineering Practice Governing High Frequency (FM) Broadcast Stations: 43-50 mc.," adopted by the FCC, June 28, 1940, we find the statement: "Doubling the height of the antenna is equivalent to increasing the power by four times." Within the line-of-sight path, and for the usual small values of ψ_2 (Fig. 1), doubling the transmitter height roughly doubles the difference in path length, $(r_2 - r_1)$, (normally a small quantity). This doubles the small angle of phase shift produced by the difference in path length, and this in turn approximately doubles the *resultant* of the direct and ground-reflected rays. But the resultant is the observed field strength at the receiving antenna, and if the resultant is doubled, the received energy is *four* times as great, since it varies as the *square* of the field strength. If instead of raising the transmitting antenna, the transmitter power were quadrupled, the receiving antenna would also receive four times as much energy, hence the above statement by the FCC.

The formula for the maximum line-of-sight distance between a transmitting and a receiving antenna is given by

$$d_t = 1.414(\sqrt{h_t} + \sqrt{h_r})$$

where d_t is in *miles*, and h_t and h_r are the transmitting and receiving antennas' heights, respectively, above the earth's surface in *feet*.

For example, the Empire State television antenna is approximately 1200 feet above ground. How far away can a receiving antenna be located if it is 30 feet above the earth's surface?

$$d_t = 1.414(\sqrt{1200} + \sqrt{30}) = 1.414(34.7 + 5.48) = 56.8 \text{ miles}$$

Thus, receiving antennas 30 feet high can pick up signals within a radius of approximately 60 miles from the Empire State Building.

THE TROPOSPHERIC WAVE.--The troposphere is the lower portion of the earth's atmosphere. In this region is produced the weather--storms, snow, rain, etc. The portions from about one-half mile above the earth up to about six miles is the region of interest to radio engineers, for it is in this region that the water vapor content of the air varies from a maximum near the earth to a minimum six miles up or so. The dielectric constant of water vapor is much higher than that of air (taken as unity) and hence will make the dielectric constant of the atmosphere near the earth appreciably greater than that six miles up.

The effect of a dielectric constant that decreases with height is to produce a *refraction* or *bending* of the space wave that enables it to overcome the curvature of the earth to some extent and thus to act upon a receiving antenna beyond the ordinary line-of-sight. Since a space wave is involved, it is evident that this bending is important mainly at ultra-high frequencies, where the space wave is the principal mode of transmission.

The bending effect of the troposphere can be taken into account most simply by assuming that the space wave *travels in straight lines, and that the earth is flatter than it actually is*, i.e., by assuming that the radius of the earth is about one-third greater. Precisely that assumption has been incorporated in the previous formula for d_t, the maximum line-of-sight distance between two elevated antennas.

However the above bending effect does not represent a *tropospheric wave*, but rather the *bending effect of the troposphere* upon the space wave component of the ground wave. The *tropospheric wave* is assumed to be a wave distinct from the ground wave. It has been found that occasionally quite strong signals can be picked up at ultra-high frequencies considerably beyond the line-of-sight distance. While there is normally some signal in this region owing to diffraction, nevertheless the signal occasionally picked up is much greater than can be accounted for by diffraction alone.

One explanation that has been offered is that occasionally the troposphere is so disturbed that in an upper layer of it the dielectric constant varies so abruptly as to constitute a *reflecting*, rather than a refracting layer. Accordingly, a space wave is reflected down abruptly into the region of the earth appreciably beyond the line-of-sight, and gives rise to the occasional strong signal there. This reflected wave is called the *tropospheric wave*. The action is similar to that of the ionosphere to produce ionospheric or sky waves (to be described later).

There is some question at the present time as to whether actual reflection takes place, or merely very marked bending so that the ordinary space wave *refracts* to a greater degree and is able to extend to a greater distance than normally around the curvature of the earth. Whatever the exact explanation may be, such effects are of interest in explaining unusual reception in the ultra-high frequency range, as well as fading phenomena similar to those produced by the sky wave at the lower broadcast frequencies.

IONOSPHERIC WAVES.--It was mentioned previously that at low frequencies, such as those employed by Marconi, propagation is by means of the ground wave and the ionosphere, and that at the lowest frequencies, in the order of 10 to 30 kc. the ionosphere and the earth act as two concentric spherical shells between which the radiant energy is guided.

At higher frequencies, such as from 100 kc. and up, the transmission is somewhat separated: the ground wave and the space wave directed to the ionosphere and reflected from it act more independently of one another. The student may wonder why a space wave directed toward the ionosphere is not negligibly small at the standard broadcast frequencies in view of what was said about the cancellation between the direct and ground-reflected rays in this frequency range.

The reason is that the reflection of high-angle waves is different from that of low-angle waves. This has been indicated previously, and will be discussed in greater detail here. In Fig. 3 is shown a high-angle direct ray, and the accompanying ground-reflected ray. Actually the distance between the ionosphere and the earth is so great that the angle α between the two rays is practically zero, so that they form essentially one ray, and after reflection from the ionosphere constitute the so-called *ionospheric* or *sky* wave.

It is clear from the figure that ψ_2 is on the order of 45° or more. For such large angles of incidence the electric field vector of the ground-reflected ray is *not* shifted 180° in phase with respect to that of the direct ray, but instead is in phase with it. The difference in path length at the standard broadcast frequencies is negligible when measured in wave lengths, so that the only phase angle between the two rays is that due to reflection, and as just stated, is zero for large values of ψ_2. Hence the two rays add at their point of contact with the ionosphere, instead of subtracting, as they do for low-angle rays. Moreover, at such heights the surface wave is of negligible intensity and may be disregarded. However, along the surface of the earth the surface wave is not normally negligible, and reception at the distant point R may be a vector combination of the surface and sky waves. The resultant may be greater or less than the two component waves, depending upon their phase at this point. If the height of the ionosphere varies, the path length of the sky wave will vary, and this in turn will vary its phase at R. The result will be fading at that point, and this will be discussed in greater detail farther on in this lesson.

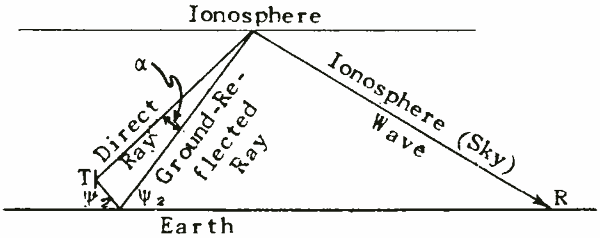


Fig. 3

One further point will be brought out concerning reflection. The variation in phase of the electric field in the ground-reflected ray with ψ_2 *occurs only with vertically polarized waves*. However, at the lower frequencies only vertically polarized waves can be radiated with any efficiency, since the conducting earth tends to short out the horizontal electric field vector of a horizontally polarized wave.

As the angle of incidence ψ_2 is varied from 0 to 90°, the phase of the reflected electric field vector varies from 180° to 0°. At some intermediate value of ψ_2 the *phase* angle is 90°. The *space angle* ψ_2 at which the *phase angle* is 90° is known as the *Brewster* angle, after the name of a physicist who discovered a similar angle in optics.

At this angle of incidence the direct and ground-reflected rays neither add nor subtract arithmetically; instead they add (geometrically) in phase quadrature (90°). Hence even in the broadcast band the space wave may not be negligible *at this particular angle of incidence*. The result is that in calculating the directive pattern of an antenna array on the basis of a ground image, some error will be involved in neglecting this effect at and around the the Brewster angle of incidence, but this error is fortunately not serious and need be considered only when high precision is required in the calculations.

The Brewster angle depends upon the frequency and the conductivity and dielectric constant of the earth. For average values of earth conductivity and dielectric constant, the Brewster angle at 1 mc. is about 6°, and at 46 mc. it is about 14.5°. The significance of this is that in the standard broadcast range, space waves directed towards the ionosphere even at low angles (but greater than 6°) are appreciable in strength, and may interfere with the surface wave as described above. That is one reason why so much effort is directed to having the broadcast antenna radiate as much energy as possible along the horizontal: to extend the interference (fading) effect as far out as possible, as well as to increase the signal strength along the surface of the earth for a given amount of radiated power.

In the u.h.f. range it might at first appear that by elevating the transmitting and receiving antennas to a sufficient degree, the angle of incidence could be made to exceed the Brewster angle, and thus cause the electric field vector of the ground-reflected ray to meet that of the direct ray in phase instead of 180° out of phase. (Additional effects of the difference in path length of the two rays would tend to counteract this additive effect).

However, in view of the large value that the Brewster angle has at these frequencies (14.5° at 46 mc.), it will be evident that for all *practical* antenna elevations and all but the shortest distances between the two, the angle of incidence will be less than the Brewster angle, and hence the two rays will inherently tend to cancel, except for the effect of difference in path length. Furthermore, high angle space waves in the u.h.f. range (for which ψ_2 exceeds the Brewster angle) are of no concern to the radio engineer because they are not reflected back to the earth by the ionosphere, but instead pass through it and are lost into space.

In the year 1902 two scientists, Kennelly and Heaviside, working independently, developed the theory that the upper regions of the earth's atmosphere contain an ionized stratum and that this ionized region, being a better conductor than the region below, reflects or refracts the radio wave and causes it to bend downward, forcing the energy to follow the curvature of the earth's surface. This assumed layer was named the Kennelly-Heaviside layer.

Over the years scientists of the Naval Research Laboratory, the Bureau of Standards, the Carnegie Institute, Radio Corporation Of America, Bell Telephone Laboratories, and others have done a vast amount of experimental work to determine the composition of the ionized region and its effects on radio transmission. Since the ionized region has been found to be very deep and to consist of several continuously changing layers, the entire region which affects radio transmission has been named the "Ionosphere." Before considering the effects of the ionosphere it is first necessary to know something of its characteristics.

The ionization of the upper atmosphere apparently is caused by ultra-violet radiation from the sun. Thus the degree of ionization over any portion of the earth's surface at any instant is determined by all the factors which affect the sun's radiation reaching that position at that time. Some of these factors are: time of day; season of the year; relative time along the sunspot cycle, and hour-by-hour variations in any unusual solar activity. From the data obtained over a number of years, it is now possible to predict from current measurements the probable degrees of ionization and their effect on radio transmission a considerable time (several months) in advance--with the exception of those effects caused by unusual and sporadic solar activities.

While the ionized regions of the upper atmosphere which affect radio transmission have the effect of reflecting layers, this term may be somewhat confusing. Actually there is some ionization of all the earth's atmosphere, the degree of ionization in general increasing with altitude within limits. The effect is somewhat as shown in Fig. 4. Three layers are shown, designated E, F_1 and F_2. These are the layers that have the greatest effect on radio transmissions and are shown at typical noon heights in summer. It will be seen that the increased degree of ionization at certain heights with lesser degrees of ionization between produces the effect of *layers* of abnormal ionization at several heights above ground.

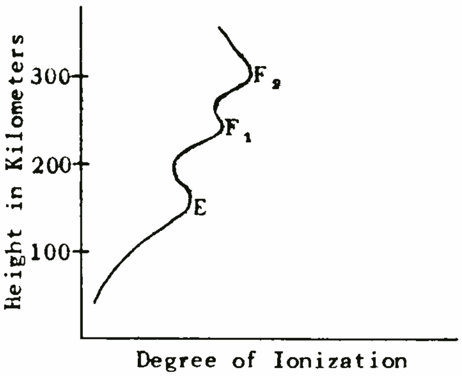


Fig. 4

Investigations of the ionosphere are made by transmitting short pulses vertically and receiving the reflected pulses at a receiver located either at or near the transmitter. The receiver normally picks up at least two pulses, one along the ground directly from the transmitter, the other or others reflected from the ionosphere. The received pulses are recorded by means of an oscilloscope equipped with a continuous photographic recorder. By carefully measuring the time along the photographic record between the direct pulse and the reflected pulse, the virtual height of the reflecting layer is easily calculated.

As the degree of ionization of a layer varies with variations in the sun's rays, the reflective and refractive characteristics of the layer vary. Further, these characteristics are different for the different frequencies. As the radio frequency of the signal is increased, for any given degree of ionization, the greater is the tendency for the signal to penetrate the layer rather than to be reflected by it. From this has been developed a term "critical frequency." The critical frequency is defined as "*the lowest radio frequency of a wave which penetrates a layer at normal incidence.*"

Another term that should be understood is the "virtual height" of an ionized layer. This term is defined as follows: "*The height at which reflection from a definite boundary surface would cause the same time of travel as the actual reflection, for a wave transmitted from the ground to the ionosphere and reflected back.*" Virtual height depends on the wave components and the frequency; the value usually stated is for the ordinary wave and for the lowest frequency at which reflection occurs.

In certain frequency ranges under certain conditions simultaneous reflections may occur from more than one layer. This is owing to the fact that while some of the energy is reflected from a lower layer, some energy penetrates and is reflected from an upper layer. An example of this is shown in Fig. 5, which shows the virtual heights of the E and F layers for a frequency range of 1.6 kc. to 2.5 kc. taken over a period 1615 to 1652, October 6, 1933.

Fig. 6 (left) shows the virtual heights of the E and F layers for the almost identical period three months later over the same frequency range. Note that reflections from the E layer occur only at certain frequencies with complete penetration at other frequencies and that F layer reflections occur over this entire band of frequencies.

Fig. 6 (right) illustrates how rapidly ionosphere conditions can change during late afternoon. This frequency run, differing for equivalent frequencies from the preceding by not more than 40 minutes, shows both E and F reflections over the entire frequency band. Such rapid changes are typical of early morning and early evening conditions.

Figs. 5 and 6 were redrawn from curves plotted by Kirby and Judson and used in Figs. 8 and 9 in "Recent Studies of the Ionosphere," *I.R.E. Proceedings*, July 1935. These curves also show multiple reflections, which occur when the original wave returning to earth is reflected upward to the layer and back to the earth a second or third time, registering each time it strikes the earth.

Measurements taken over a number of years have established several pertinent facts. In the latitude of Washington, D.C., there are at all times two major ionosphere layers, the E layer from 100 to 120 kilometers above the earth's surface and the F_2 layer with approximate virtual heights from 230 to 350 kilometers. A third layer, present in the daytime and especially in summer, is termed the F_1 layer and exists in heights from 180 to 240 kilometers.

The F_2 layer is a daytime continuation of the F layer and varies

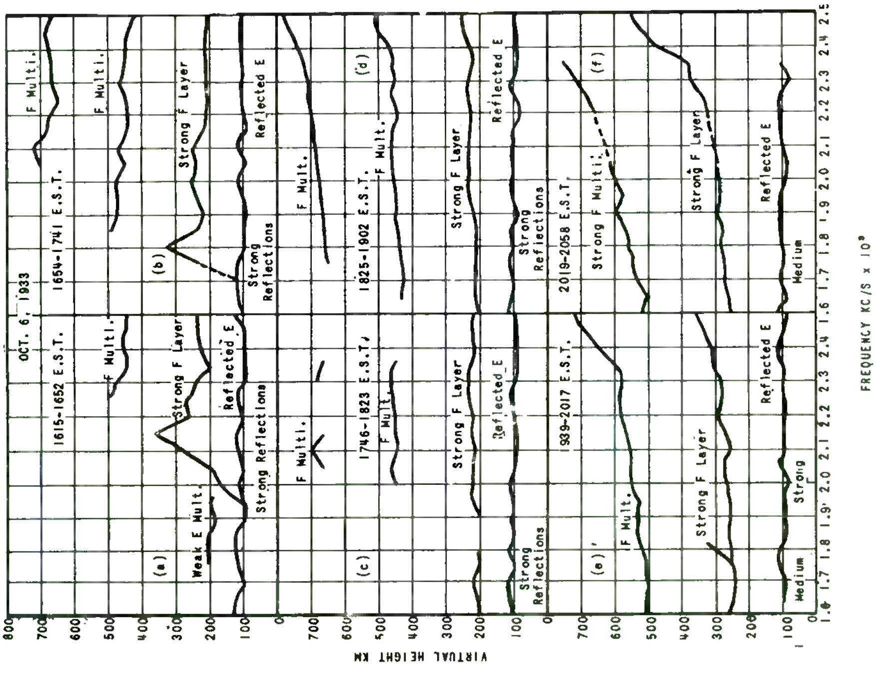

OCT. 6, 1933
(a)
1615-1652 E.S.T.
Weak E Mult.
Strong F Layer
Reflected E
Strong Reflections
F Multi.
(b)
1654-1741 E.S.T.
F Multi.
F Multi.
Strong F Layer
Reflected E
Strong Reflections
(c)
1746-1823 E.S.T.
F Multi.
F Mult.
Strong F Layer
Reflected E
Strong Reflections
(d)
1825-1902 E.S.T.
F Mult.
F Mult.
Strong F Layer
Reflected E
Strong Reflections
(e)
1939-2017 E.S.T.
F Mult.
Strong F Layer
Reflected E
Medium
Strong
(f)
2019-2058 E.S.T.
Strong F Multi.
Strong F Layer
Reflected E
Medium
VIRTUAL HEIGHT KM
800 700 600 500 400 300 200 100 0
FREQUENCY KC/S x 10³
1.6 1.7 1.8 1.9 2.0 2.1 2.2 2.3 2.4 2.5

Fig. 5

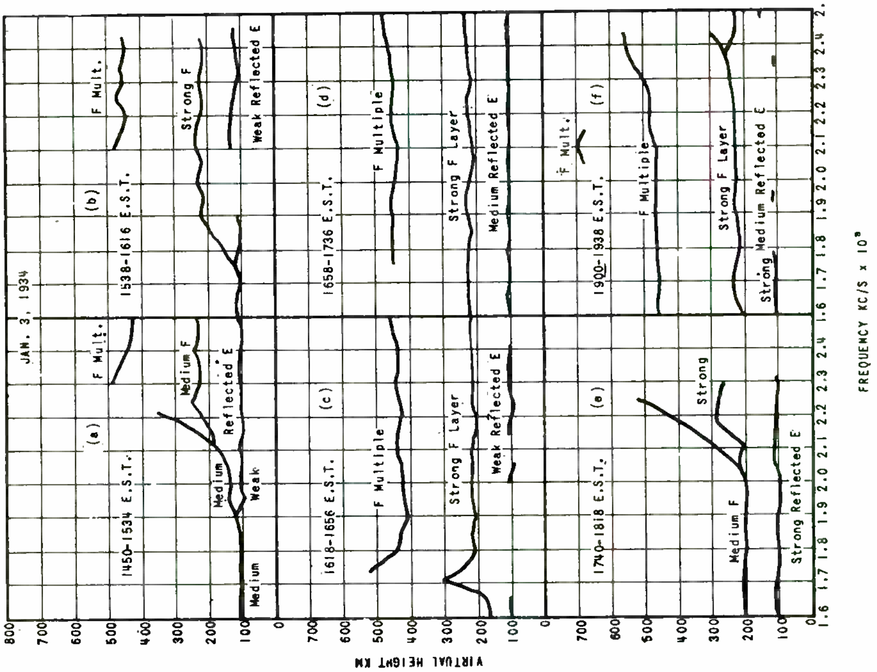

JAN. 3, 1934
(a)
1450-1534 E.S.T.
F Mult.
Medium F
Reflected E
Medium
Weak
Medium
(b)
1538-1616 E.S.T.
F Mult.
Strong F
Weak Reflected E
(c)
1618-1656 E.S.T.
F Multiple
Strong F Layer
Weak Reflected E
(d)
1658-1736 E.S.T.
F Multiple
Strong F Layer
Medium Reflected E
(e)
1740-1818 E.S.T.
Strong
Medium F
Strong Reflected E
(f)
1900-1938 E.S.T.
F Mult.
F Multiple
Strong F Layer
Strong
Medium Reflected E
VIRTUAL HEIGHT KM
800 700 600 500 400 300 200 100 0
FREQUENCY KC/S x 10³
1.6 1.7 1.8 1.9 2.0 2.1 2.2 2.3 2.4 2.

Fig. 6

symmetrically around noon. The F_1 layer disappears during the night. This is shown in the upper section of Fig. 7. (Fig. 7 is taken from "High Frequency Transmission" by the National Bureau of Standards, *I.R.E. Proceedings*, December 1941.

CRITICAL FREQUENCY.--This frequency, also called the "penetration frequency," is that at which the virtual height for a wave component at vertical incidence has a maximum value caused by penetration of the wave through the layer.

At any given time of day under conditions of ionization, as the frequency is increased from some low value, the apparent virtual height of a layer will have some fairly constant value until the penetration frequency is approached; then as the frequency is further increased the measured virtual height will increase rapidly and reflections may cease for a short frequency interval, then again appear from the next higher layer, etc. A typical frequency run is shown in Fig. 8 with the E, F_1 and F_2 layers and their critical frequencies shown. The E layer has a virtual height of approximately 115 kilometers until a frequency of about 3.3 mc. is reached, at which point it rises sharply and reflections disappear, apparently owing to complete absorption at that frequency; a strong reflection is resumed from the F_1 layer at about 3.6 mc., the virtual height of this layer being approximately 200 kilometers. At about 4.2 mc. the F_1 reflections rise sharply reaching about 450 kilometers at 4.4 mc., this being the F_1 critical frequency. The F_2 virtual height is about 415 kilometers until the F_2 critical frequency is reached at about 5 mc. This is

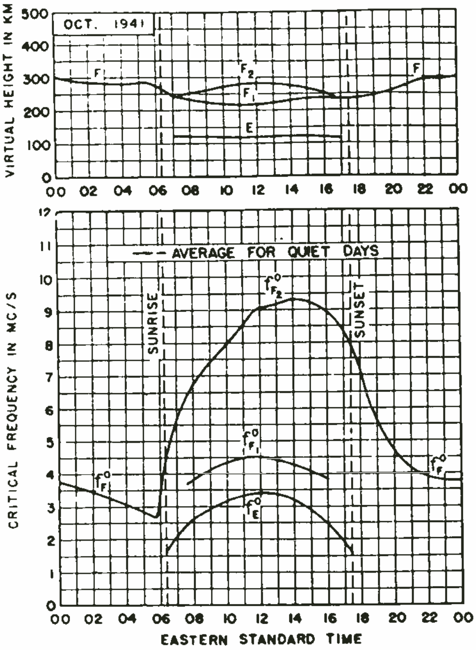


Fig. 7

a typical late spring condition shortly before noon.

f_E, f_{F_1} and f_{F_2} denote respectively the critical frequencies of the E, F_1 and F_2 layers.

The critical frequency is a function of the layer height and the degree of ionization, and hence is affected by all the factors which control the conditions in the ionosphere. The lower section of Fig. 7 shows average curves for critical frequencies of the three major layers plotted against time of day for the month of October, 1941. Note that the curves of f_E and f_{E_1} vary quite symmetrically around noon. The critical frequency f_E actually drops just about into the upper limit of the broadcast band at sunrise and sunset, rising to about 3.4 mc. at noon.

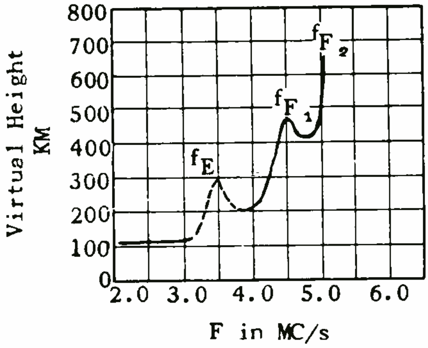


Fig. 8

The F_1 critical frequency starts in the morning shortly after sunrise at about 3.7 mc., reaches 4.5 mc. at noon, and drops back to 3.8 mc. in the late afternoon. This is strictly a daytime condition.

f_{F_2}, the critical frequency of the F_2 layer, varies over a very wide range. Dropping from 3.7 mc. at midnight to 2.7 mc. just before sunrise, it rises sharply to a peak of 9.3 mc. at 2 p.m., then drops gradually back, the drop continuing until the following morning but levelling off somewhat about 9 p.m.

It particularly should be noted that the most rapid *variations* of ionization in all the major layers take place around sunrise and sunset; *hence the variations in the medium of propagation are the most rapid at those times.* The effects of those variations on the transmission of a radio signal will be discussed later. It should be pointed out that the curves of virtual height and critical frequency are simply typical and deviations from the values shown may be considerable under varying conditions of season, sunspot cycle and solar activity.

The critical or penetration frequency is a criterion of the height and degree of ionization of the layer under consideration, and as such is important in that it permits the determination of more valuable effects of the layers of the ionosphere. In conventional radio transmission it is desired to propagate the signal not vertically but from point to point along the surface of the earth. Consider the conditions illustrated in Fig. 9 and the manner in which a wave is affected in passing into an ionized layer.

In the discussion above the wave is considered as being *reflected* from one or more of the ionized layers and thus returned to earth. The actual process, however, is a combination of reflection and *refraction.* Where the entrance of the wave into the layer is perpendicular, the return of signal energy back to the transmitting point probably is mostly due to reflection and depends upon the frequency. Where the angle of propagation with respect to the plane of the layer is substantially less than 90°, the effect on the wave is due mostly to refraction.

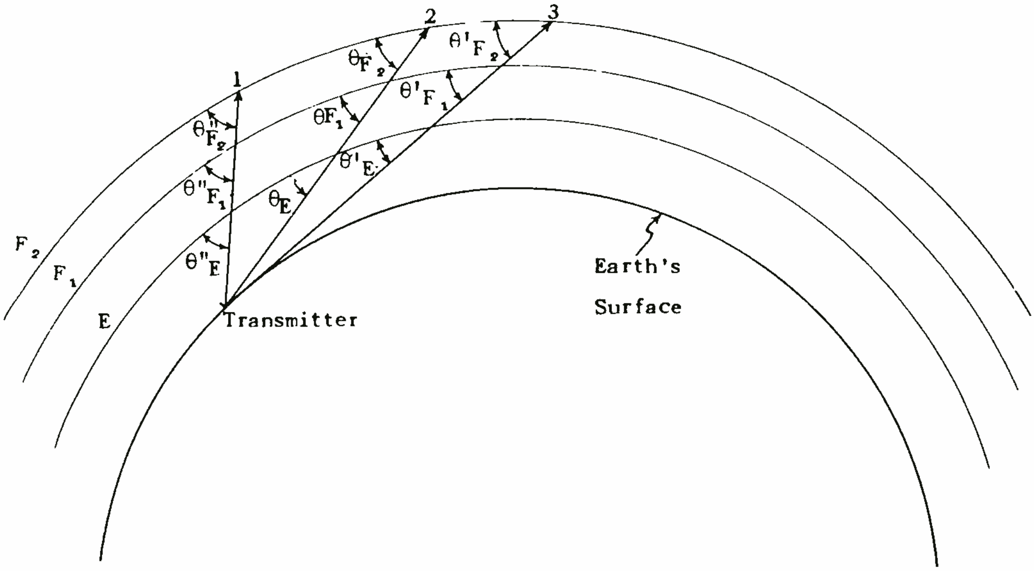


Fig. 9

Refraction is defined as the deflection from a straight path suffered by a ray of light, heat, sound, or the like, in passing obliquely from one medium to another in which its velocity is different, as from air into water or from a denser to a rarer layer of air.

Thus as the radio wave passes into a layer of greater ionization it is bent. Further, *the degree of refraction is a function of frequency.* Everyone has seen the familiar "rainbow" effect produced when a ray of ordinary "white" light is passed through a prism. The various frequency components making up the white light are refracted to different degrees, the higher frequencies (blue) more than the lower frequencies (red) so that a narrow beam of white light entering one side of the prism leaves the other side of the prism in the form of a wide beam in which the color components are separated.

In the case of radio waves travelling through the ionosphere, the high frequency radio wave is refracted less than a lower frequency wave; i.e., the bending is less at the higher frequency. Thus a higher frequency wave entering a lower layer at a given angle is more likely to penetrate the lower layer and be refracted by a higher layer than is a lower frequency wave.

From Fig. 9 it will be seen that the angle of incidence at which the radio wave enters an ionized layer is determined by two factors: first, *the angle of propagation with respect to the earth's surface;* second, *the height of the layer.*

Consider, in Fig. 9, the three components of the wave leaving the transmitter. No. 1 leaves at a rather high angle with respect to earth and hence θ''_E, θ''_{F_1} and θ''_{F_2} all are large. However, θ''_{F_1} is greater than θ''_E and θ''_{F_2} is greater than θ''_{F_1}. No. 2 component is transmitted at a lower angle and θ_E, θ_{F_1} are all smaller than the corresponding angles for No. 1. All of the layer angles of incidence are smaller for component No. 3 than for corresponding angles for the other components.

For a given degree of ionization of the layer, the smaller the angle of incidence the less the degree of refraction required to bend the wave back toward the earth. This is easily seen from inspection of Fig. 9. Fig. 10 shows the process by means of which a radio wave is returned to earth by an ionospheric layer. It will be noted, however, that *the lower the angle of incidence, the smaller the degree of bending and hence the further the wave travels through the layer before being refracted back to earth.*

The facts explained above may be summarized in general terms and their broad

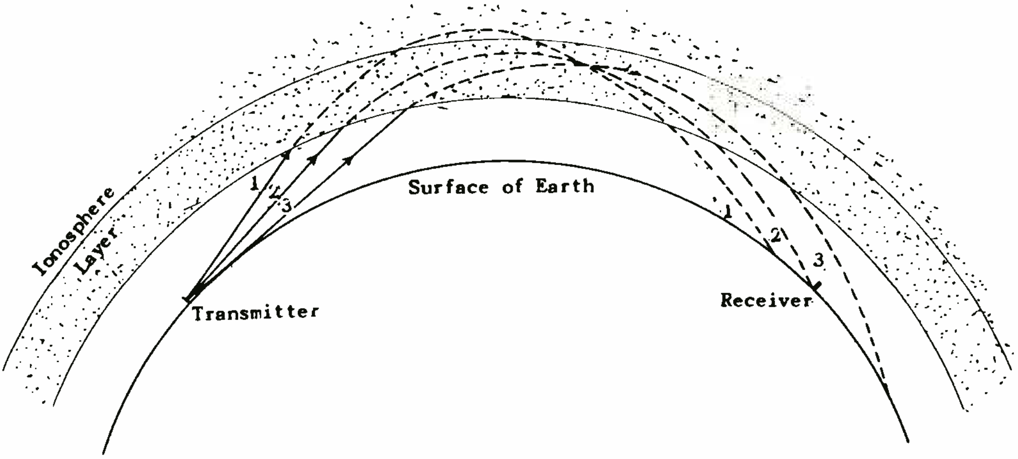


Fig. 10

effects studied:

First, the ionosphere does not consist of a single highly ionized stratum in contact with a non-ionized stratum, but instead consists of several layers at various heights above the earth's surface. Ionization in a degree effective to radio waves becomes apparent about 50 km. above the surface of the earth and increases as the distance above the earth increases. (See Fig. 1.) This means that there is no sharp difference in the ionization of the atmosphere at adjacent levels but rather a gradual bending of the wave in which the law of reflected energy, which states that the angle of reflection equals the angle of incidence, does not hold true except in the case of waves that enter the layer at angles approaching 90°.

Second, the height of any degree of ionization varies with the time of day and the season of the year, normally being lower during the middle of the day and higher at night, averaging somewhat lower in the summer and higher during the winter months. Measurements conducted by the Naval Research Laboratory in conjunction with the Carnegie Institute at Washington and by the Bureau of Standards have shown that the degree of ionization at which lower level refraction takes place varies between heights of about 100 km. at noon on a summer day to 250 to 400 km. on a winter night, and even higher at times; also *that the degree of* ionization is continually changing, sometimes rapidly, sometimes slowly. This continual change causes many undesirable phenomena.

Third, the several principal ionization layers affect various frequencies differently. At any instant, definite refraction of one frequency may take place in the lower layer while a higher frequency signal may pass through that layer and not be seriously refracted until it strikes a higher layer. The critical frequency at which refraction by a given layer begins may be quite sharply defined, and as the degree of ionization varies the critical frequency of refraction may change quite suddenly. This accounts for the sudden fading in or out of a distant signal when the frequency happens to be in the vicinity of a critical value.

SKIP DISTANCE.--As experience was gained in high-frequency communication many peculiar phenomena were observed, one of the most important being that of "skip distance." It has been found that for given power output, as the frequency is increased above about 2,500 kc., the reliable communication range is increased. However, as the frequency is increased above about 6,000 kc. there will be certain distances at which communication cannot be maintained although at much greater distances the signal strength is excellent.

It is perhaps well to tabulate some of the average results at different frequencies and then analyze them. It must be understood that these average figures are subject to wide deviation with season, sunspot cycle, etc.

Frequency in KC.	Ground Wave in Miles	Skip Distance in Miles		Max. Range in Miles	
		Day	Night	Day	Night
4,000	100	0	0	---	7,000
8,000	50	70	200	---	7,000
12,000	15-25	285	575	2,000	7,000
16,000	5	495	Inf.	3,500	0

Analyzing the tabulated figures, it is seen that at a frequency of 4,000 kc. the ground wave extends about 100 miles, and at this frequency the reflected and refracted energy from the sky wave overlaps that of the ground wave and there is no skip distance. The daylight range of this frequency is not very great, but the reliable night range is about 7,000 miles. (All of the tabulated figures are for *reliable telegraph communication* with about five kw. in the antenna. The occasional distances obtained are of course much greater than those tabulated. Greater power will be necessary to produce equivalent results with modulated wave transmission.) This frequency would be a very good reliable night frequency *for either long or short distances.*

At 8,000 kc. the ground wave reaches only about 50 miles, and from that distance before the signal is again heard there is a gap of about 70 miles during the day and 200 miles at night. Beyond this gap or "skip distance" the communication is again very similar to that of the 4,000 kc. frequency--not very good during the day, but reliable up to 7,000 miles at night. This is also a good night frequency for long-distance communication, but is not particularly desirable for daytime work. An intermediate frequency band around 6,000 kc. is extensively used for high-frequency broadcasting at night.

At 12,000 kc. the ground wave dies out at from 15 to 25 miles. Beyond that point, for about 285 miles during the day and about 575 miles at night, no signals will be heard, but beyond the skip distance the communication is again good. It will be observed that the reliable daytime distance is now about 2,000 miles while the nighttime distance is still about 7,000 miles. This is a very good frequency for moderately long distances during the day and all long distances at night. This frequency, however, would be useless for distances less than 300 miles during the day and less than 600 miles at night. A band around 11,500 kc. is used for daytime high-frequency broadcasting and later in the evening a band around 9,500 kc. is used.

At 16,000 kc. the ground wave dies out at about 5 miles and the daytime skip distance extends out to about 500 miles. During the night the skip distance extends to infinity, i.e., the sky wave normally is not refracted down within the limits of the earth's surface and so is completely lost. This is an excellent daytime frequency giving a reliable range of about 3,500 miles, but cannot be used at all for reliable night communication. Frequencies in this region are extensively used for high-frequency broadcasting, particularly in the morning and where the entire path is in the daylight zone.

It must be understood that all of the above figures are averages over a period of many months and that there is no sharp dividing line between the day and night distances. Also, conditions will vary from year to year with annual positions along the sunspot cycle. The daytime distances are taken at noon and night distances at midnight. At other times the conditions are continually changing, gradually blending from one condition to another. This is clearly shown in Fig. 7. Also the day and night figures will be different at the different seasons of the year. The conditions during a winter day may be similar to those during a summer night, etc. Conditions east and west will be different from north and south transmission. Operating conditions in the tropics will not be the same as in the more northern regions.

The figures show that for moderate distant daylight communication, 4,000 kc. may be used without danger of missing the signal due to skip distance. The same will be true for night conditions except that the reliable range will be extended. At 8,000 kc. for shipboard or aeronautical work, it would be necessary to know approximately the desired ship or plane position, if comparatively close, to be sure it was not in the region of skip distance. 12,000 kc. and 16,000 kc. are strictly long-distance frequencies. Both are very good in the daytime, particularly 16,000 kc. The latter frequency, however, is not at all dependable for nighttime communication.

With these figures in mind it should be a simple matter, if a very wide choice of frequencies is available, to select a frequency that will give the most reliable results at any particular time of day and season of the year. In selecting the frequency it should be remembered that night conditions may prevail over part of the distance between the two stations and day conditions over the other part. Also, when communication is between stations in a north and south direction, summer conditions may prevail at one station and winter conditions at the other. All of the various factors will affect the reliability of long-distance communication.

In commercial communication work and broadcasting at high frequencies over long distances such as in transatlantic services, stations may start out in the morning at a high frequency, shift early in the afternoon to a lower frequency, again shift to a lower frequency later in the afternoon and to a still lower frequency at night. In the case of short-wave broadcasting a station may broadcast simultaneously on two or more frequencies so as to serve the greatest number of listeners in different parts of the world in which different conditions of day and night prevail.

Based on many years of experience in measuring ionospheric and transmission conditions, the National Bureau of Standards over several years published in the *I.R.E. Proceedings* "High-Frequency Radio Transmission Conditions, (*month*), (*year*), with Predictions for (*month*), (*year*)." Fig. 11 shows a set of curves published in the June, 1941 issue of the *I.R.E. Proceedings*, each curve representing a stated skip distance and showing the maximum frequency vs. time of day that can be used at that distance from the transmitter to avoid skip signal.

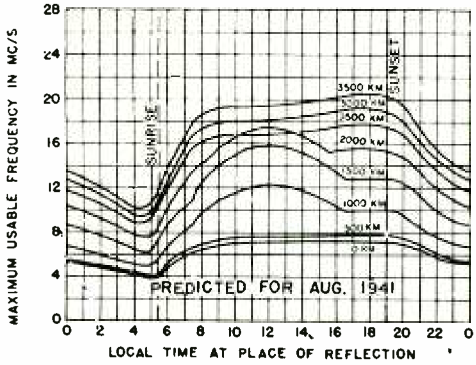


Fig. 11

For example, to operate at a distance of 1,000 kilometers at noon it is possible to use any frequency up to 12 mc., while for the same distance at 4 a.m. it would be necessary to drop down to about 5 mc.

For reliable communication at 3,000 km. in the late afternoon, 4 p.m. to 7 p.m., any frequency up to 19 mc. could be used but at midnight to avoid being in the skip zone it would be necessary to drop to below 13 mc.

Particularly note that for the shorter distances during the day the maximum usable frequencies occur about noon while for the longer distances the maximum usable frequencies occur later in the afternoon. These should be compared with the curves in the lower section of Fig. 7. Evidently the lower frequencies are being transmitted through the E and F_1 layers *until late afternoon*, the higher frequencies through the F_2 layer all during the day and night.

The reason for skip distance and for the long distances obtained with comparatively low power, is the fact that the refraction of the wave when it enters the ionosphere causes it to travel for long distances above the surface of the earth where the absorption losses are almost negligible. Refraction is owing to the fact that the ionization of the medium at different heights is not uniform but is greater at the higher levels. Energy in the form of a radio wave travels more rapidly through a highly ionized medium than through a medium in which the degree of ionization is less. Therefore, when the radio wave travels upward at an angle, the top of the wave hits the highly ionized medium before the lower portion of the wave arrives; the wave thus bends gradually downward because the top of the wave is travelling more rapidly than is the lower portion of the wave. If the wave is travelling at an angle approaching the vertical, a large portion of the front of the wave may enter the ionized region at almost the same time and some reflection will take place instead of refraction. A large portion of the wave which enters the layer at angles greater than the critical angle of refraction travels straight into the ionized strata and does not return to the surface of the earth.

As the frequency is increased, the degree of refraction of the wave becomes less and less, and the energy therefore returns to the earth's surface at greater and greater distances from the transmitter. At the very high frequencies the ground wave is practically useless. Since the reflected energy comes down at very sharp angles, and as the frequency is increased the refracted energy comes down at greater and greater distances, it is evident that as the frequency is increased beyond a certain point there will be a gap between the last of the reflected energy plus the ground wave and the first of the refracted wave, and as the frequency is made higher the gap will become wider. This gap is the skip distance and is shown in Fig. 12.

The greater the height of the effective layer, the less the degree of refraction of the wave because the angle at which the wave strikes the layer will become greater (see Fig. 9, for example) and only the comparatively low angle radiation will be refracted; the other energy will be partly reflected but most of it will penetrate the layer and not return to earth. This will cause the distance before the refracted energy reaches the surface of the earth to be-

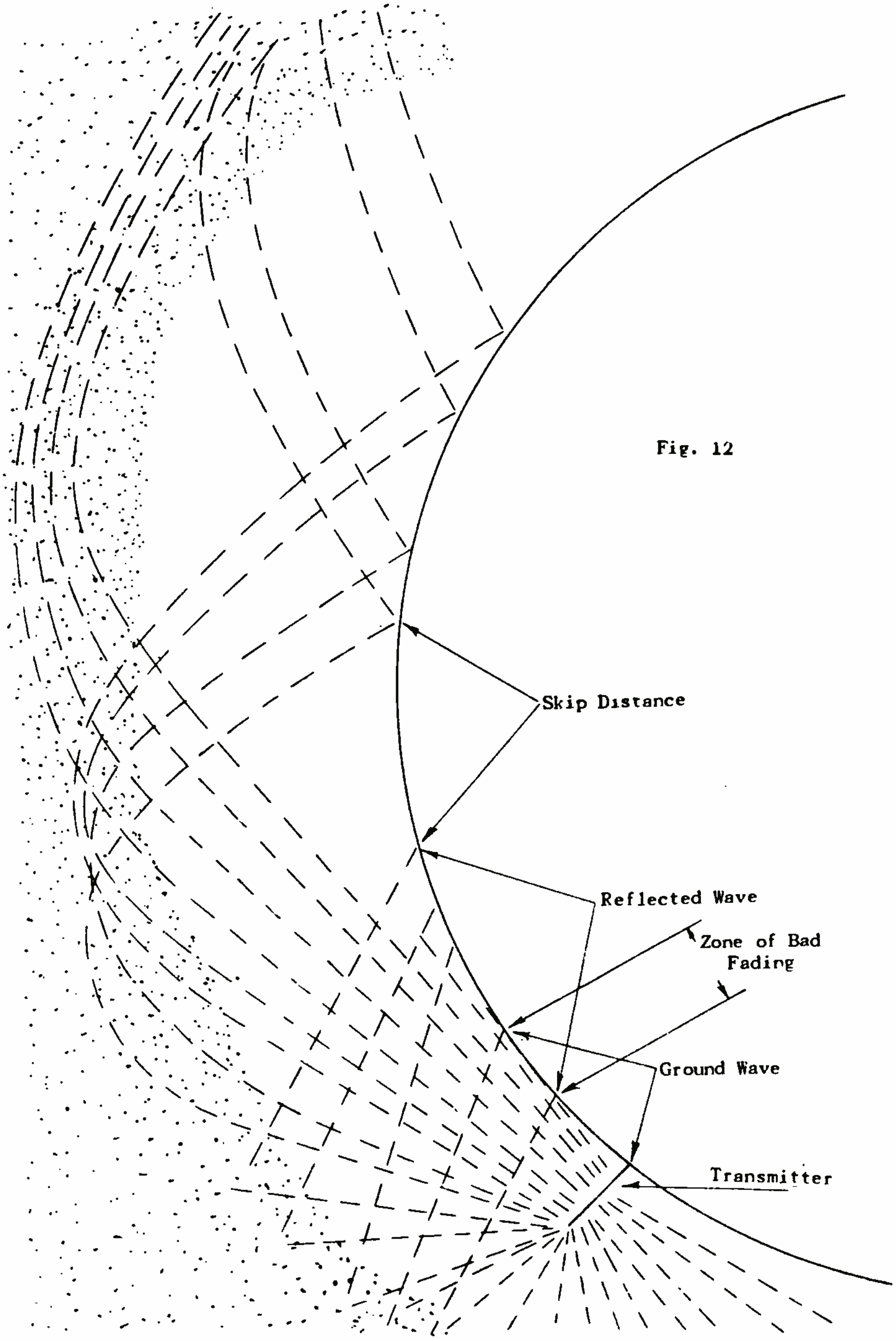

Skip Distance
Reflected Wave
Zone of Bad Fading
Ground Wave
Transmitter

Fig. 12

come greater. Thus when the layer is higher the skip distance (and also the reliable communication range), will be greater.

For reliable long-distance communication it is also necessary to use adequate power. While amateurs operate over long distances with very low power, reliable communication as required for commercial operation and good broadcasting could not be maintained under such conditions. The power required, however, owing to the decrease in absorption losses and to high radiation efficiency, is much less than that required to maintain equivalent long-distance communication with the low and intermediate frequencies. For transatlantic communication and for similar distances, 20 to 40 kw. with a directional transmitting antenna is the commonly accepted power requirement for reliable communication and broadcasting. The directional antenna produces the equivalent of an increase of power. Also a properly designed directional antenna permits the vertical angle of propagation to be kept low so that the wave strikes the ionized layer at the proper angle for long-distance transmission. This is demonstrated on Fig. 9, where the low angle radiation enters the layer at a much smaller angle than does the high angle radiation and hence travels to greater distances before returning to earth.

It is a common belief that the reception of energy from the sky wave is confined to the higher frequencies. That is not correct. All unusual long-distance reception is due to the sky wave. Often at night some low-power station on an intermediate frequency will be heard at a distance of several thousand miles when the station's reliable daytime range is a hundred miles or less. This usually occurs on a cold night when the effective layer is very high. All long-distance broadcast reception is due to the sky wave.

Whether the antenna system radiates a vertically polarized or a horizontally polarized wave, the effect of the ionized layer in conjunction with the earth's magnetic field is to produce polarization opposite to that of the incident wave, so that at the point of reception the wave contains more or less equal vertical and horizontal polarization components. As explained later, it appears preferable to employ a receiving antenna more responsive to the horizontally polarized components of the wave for long-distance reception.

FADING.--At certain distances within the limits of the ground wave fading is very pronounced, particularly in the early evening. At distances within the limits of the ground wave total energy from a transmitter at any given point will be made up of two principal components, the ground wave and the sky wave. At any point in space within the limits of the ground wave energy will reach a receiving antenna by way of at least two separate paths: first, in a direct line following the curvature of the earth; second, from the ground up to the effective ionosphere layer and back to ground. Now if the signals over both paths reach the receiving antenna exactly in phase, the total voltage developed at the receiving antenna will be the sum of the two. If, however, the signals over the two paths reach the receiving antenna exactly out of phase the total voltage developed at that point will be the difference between the two; and if they are equal in amplitude and opposite in phase the total voltage across the receiving antenna will be equal to zero. At any instant the total signal voltage at the antenna will be the vector sum of the ground wave and the sky wave.

The height of the ionized layer varies continually--sometimes slowly, sometimes rapidly. Each time the height of the layer varies, the length of the path for the sky wave between the transmitter and receiver varies. As the path varies in length the phase relation between the sky wave and the ground wave varies, the two waves sometimes adding and sometimes subtracting. The signal at the receiver varies accordingly, being strong when the two waves add and weak when the waves subtract. This will cause alternately strong and weak signals at the receiver, or *fading*.

Fading may be very slow or may be so rapid as to occur at an audible frequency, in which case bad distortion will result. High speed fading is most noticeable on broadcast programs, some of which are made entirely unintelligible. Practically all modern receivers have automatic volume control, which minimizes the effects of slow fading within quite wide limits. However, very rapid fading often occurs at a rate too fast for the a.v.c. to follow, in which case the effect is very bad.

Fading is most noticeable at night and particularly in the hours immediately after sunset. This is because then the de-ionization of the upper atmosphere is occurring most rapidly, thus rapidly changing the effective heights of the ionosphere layers and therefore the length of the path of the sky wave.

The radiated wave for long-distance transmission usually follows more than one path from the transmitter to the receiver. This is indicated in Fig.12, where a path of two hops, involving the skip distance is shown, as well as one long hop which will ultimately reach the earth at the same point as that for the two-hop path. If short waves are employed, and in any case if the distance is long, the ground wave will be attenuated before reaching the receiver, and fading as a result of interference between the ground and sky wave will not occur. But if one or other of the paths varies in length owing to variations in the ionosphere, *then fading can occur as a result of varying interference between two or more sky waves* arriving along different paths. Such fading may be different for the carrier and side band frequencies, and give rise to the phenomenon known as *selective fading*. This renders speech and music unintelligible.

Fading is independent of the power output, and depends solely upon the design of the antenna, the conditions in the ionosphere, and the conductivity of the earth. With these conditions fixed, if the power is increased, the field intensity of both components reaching the receiver will be increased in the same proportion so that the points where the two interfere with one another will not be changed. This is true whether the two components are the earth and a sky wave, or two sky waves that have traveled over different paths.

In the standard broadcast range, if the height or configuration of the antenna is changed so as to vary the vertical angle of radiation; if the conditions in the ionosphere vary so as to change the angle with which the sky wave strikes the layer; if the conductivity of the earth changes so that there is more or less absorption of the ground wave; then the distance at which any degree of fading takes place will be changed. Serious fading may start at anywhere from about 75 to 125 miles from the transmitter, depending upon the factors mentioned above.

All of these effects are being studied by leading radio engineers in an endeavor to minimize the undesirable phenomena, particularly that of fading.

It has been found that the fading of a given high-frequency signal is rarely simultaneously the same at two or more points separated by five to ten wave lengths. At the receiving stations of RCA Communications, this fact is made use of in one of the most practical solutions to the problem of fading yet to be developed.

The diversity antenna system in which three directional antennas separated by about 1,000 feet feed into a common receiving system will be explained later in this lesson. Experience has shown that these antennas should be spaced not less than 10 wave lengths for best results. For high-frequency operation (in the crder of 8 to 20 mc.) spacing of 1,000 feet is satisfactory. The diversity system operation is based on the fact that, for a given signal, the voltages developed in three antennas sufficiently well spaced will very seldom fade simultaneously.

For radio telegraph reception the signals from the three antennas are fed by untuned transmission lines to separate receivers, amplified, rectified, and the audio outputs amplified and fed into a common load resistor, the current through which varies the bias on a vacuum tube keying relay. The vacuum tube relay keys the output of a tone oscillator to a telegraph line. So long as the amplitude of the signal *from any one receiver* is above a minimum value as determined by the noise level, the vacuum tube relay operates and the signal to the line is exactly the same as if the received signal were much stronger. Thus the only time in which the signal reaching the operator or the recording device will show the effects of fading, is when the signal fades out completely and simultaneously in all three receivers.

Satisfactory diversity reception of radio telephone or broadcast signals is more difficult. In the case of radio telegraph reception it is simply a matter of operating a keying relay circuit for dots and dashes; out-of-phase signals are relatively unimportant so long as they do not destroy the spacing of the dots and dashes, and so long as the length of a dot or a dash is long as compared to the time difference between the arrival of out-of-phase components. Amplitude variation in the telegraph signal is unimportant so long as the signal will operate the vacuum tube relay.

In the case of a broadcast signal, the carrier and its sidebands occupy a space in the frequency spectrum in the order of 10 to 20 kc. The actual intelligence of the signal is conveyed by the variations in amplitude. High speed fading may be at an audible frequency thus producing the effect of modulation which may destroy the characteristics of the signal. Fading over a frequency band of 10 or 20 kc. is not always the same in all parts of the band; it may differ at frequencies separated by only a few hundred cycles/second. Thus it is possible for the one side band to fade out leaving the carrier and the other side band; the carrier may fade out leaving the two side-band frequencies; certain side-band frequencies may fade out leaving the carrier and other side-band frequencies. (Such selective fading is particularly likely to occur at a time when the operating frequency approaches the critical frequency of one of the ionosphere layers.)

For ordinary slow periodic fading a good automatic volume control circuit with the proper time constant will hold the level quite constant. However, when the diversity system is used to minimize the more rapid fading, other problem are introduced. The first problem is that of proper phasing. Where slight phase discrepancies may be unimportant in the reception of telegraph signals, they can make the reception of broadcast signals entirely unsatisfactory. The most satisfactory arrangement would be to have the output of only one receiver connected to the line at any one time, the one, of course, in which the signal-to-noise ratio is the highest. This would involve a quite elaborate system of relays.

When the outputs of three receivers, each equipped with automatic volume control, are fed into a common circuit, another undesirable condition arises. Even though the automatic volume control keeps the three signals at a quite constant level, if the signal tends to fade in two receivers but not in the third, the sensitivity of the two receivers in which the signal is fading is increased, this in turn increasing the noise level. Thus at an instant when one receiver may be adding little to the signal output, its noise level is increased and adds to the noise level of the output. Even if the signal does not fade sufficiently so that its signal contribution is a total loss, the increased sensitivity due to a.v.c. will increase the general noise level.

In the system finally adopted for use with the diversity antenna, the signal from each antenna is amplified by means of an individual superheterodyne circuit and individually rectified. The audio outputs of all three detectors are fed into a common resistance and a portion of the common resistance drop is used to control the gain of all receivers simultaneously. Thus so long as at least one receiver is delivering a satisfactory signal level, the gain of all receivers is reduced so as to reduce the noise level. This allows full advantage to be taken of the diversity system because as the output of any one receiver increases, the total gain is reduced; if one increases in about the proportion of the decrease in another, the decreasing signal is simply allowed to fade out, the other receiver carrying the load. A general decrease in all three receivers must occur before the total gain is increased to the point of excessive noise level.

ECHO SIGNALS.--With the common use of comparatively high power at high frequencies another troublesome condition has arisen. Since high-frequency energy encounters so little absorption in travelling through the ionosphere, a signal that is not returned to the surface of the earth reasonably soon too often can travel enormous distances through these regions. Thus it is possible for a signal to travel completely around the earth a number of times and still retain a considerable field intensity.

If the signal at the receiver arrives by way of the most direct route, and then proceeds completely around the earth and is returned to the same receiving antenna again, the second signal will be received at a short time interval after the original signal, even though both left the transmitting antenna together. The velocity of propagation is approximately 186,000 miles/second and the distance around the earth is approximately 25,000 miles so that the time lag of this second signal will be 25/186 or .13 second. At slow telegraph speeds this will sound to the operator like an echo but probably will not prevent copying. The second signal is called an "echo signal."

At high operating speeds the echo signal can render a circuit completely inoperative. At 100 words/minute, assuming three characters per letter and the time of the spacing between dots and dashes as equalling the time of the actual characters, the average time of one character is .02 second. This is a much shorter time than the time lag of the echo signal, and consequently the echo of a character will appear *several characters later*. It can easily be imagined what this will do to the signal. There may even be second and third echo signals. Thus a frequency at which an echo signal occurs will be useless for high-speed communication so long as this condition exists. A similar condition can occur in high-frequency broadcasting in which case the speech or music will be completely garbled.

Another cause of echo signal can be original reception of energy from two paths of dissimilar lengths. For example, a signal being received from a station 3,000 miles away may reach the receiver by the direct route and also by the long path around the earth, and approximate distance of 22,000 miles. This will produce an echo only slightly different in time lag from the one already described. An echo signal of this type, however, can be eliminated by the use of a uni-directional receiving antenna because it is coming from a direction approximately opposite to that of the direct signal. A directional transmitting antenna would not produce this type of echo signal.

DIRECTING THE ARRAY.--In erecting a directional transmitting or receiving antenna, it must be remembered that the surface of the earth is spherical and a radio wave in going from one point on the earth to another point ordinarily will follow a great circle course. Thus the direction in which the directional antenna array must be pointed for communication between two widely separated points must be great circle direction. The study of a globe in comparison with a flat map will make this point clear.

TYPES OF RECEIVING ANTENNAS

In its simplest form a receiving antenna is a vertical conductor connected to ground at the lower end through a coupling coil which energizes the first tuned r.f. circuit of a receiver. When intercepting a normal vertically polarized wave the voltage developed across such an antenna is directly proportional to its effective height. Thus if the signal intensity is 50 µv./meter and the antenna has an effective height of 6 meters, the voltage developed across the antenna will be 300 µv.

This signal voltage will cause a current to flow in the coupling coil and induce a voltage into the tuned grid circuit of the r.f. amplifier tube. See Fig. 13. In this case for most frequencies the current in L will be small because the antenna probably will not be near resonance and the reactance of the circuit will be large. This arrangement is practical only with modern high-gain receivers. Such an antenna is completely non-directional in the horizontal plane.

If it is necessary to develop maximum voltage at the input terminals of the receiver from such an antenna, condenser C may be added as shown in Fig. 14, and the antenna circuit tuned to resonance with the signal being received. This will very greatly increase the signal input to the receiver and will add to the selectivity of the system. Its disadvantage is the additional tuning control required. Since the receiver manufacturer ordinarily does not know the electrical characteristics of the antenna to be used, an antenna tuning condenser cannot be ganged to the other r.f. tuned circuit condensers. Therefore with modern receivers it usually is considered more practical to obtain the desired gain and selectivity by proper design of the r.f. and i.f. amplifiers than by tuning the antenna circuit. This, of course, does not apply to specialized receivers to be operated by experts and where the manufacturer can specify the antenna constants.

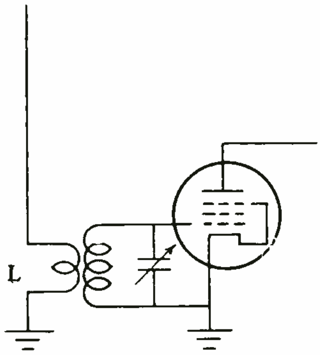


Fig. 13

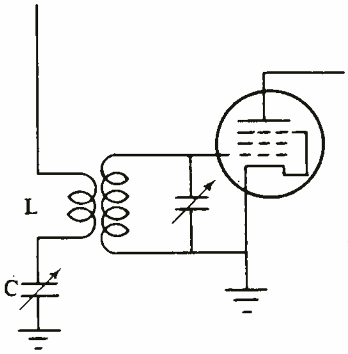
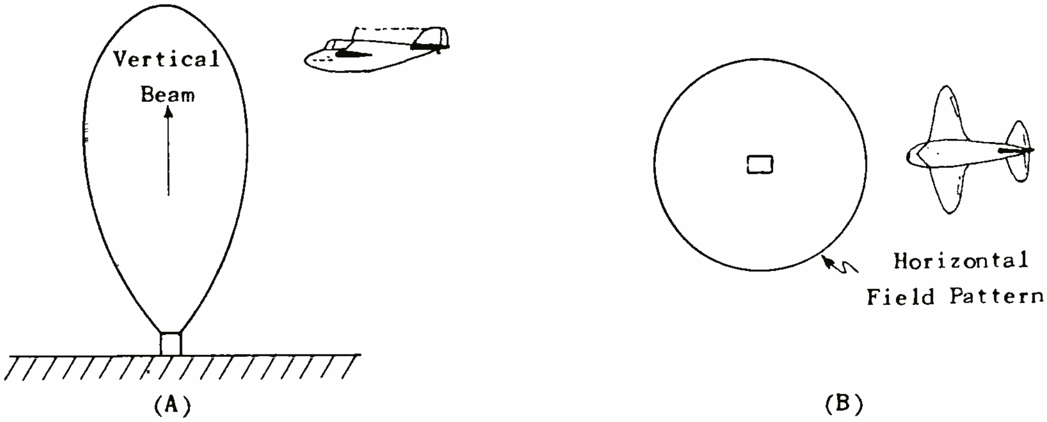


Fig. 14

In the expanding commercial applications of radio the use of directional antennas has become a frequent requirement, for reception as well as for transmission. One of the most rapidly expanding radio applications has been in aircraft navigation. Simple directional receiving antennas on the airplane combine with the use of directional transmissions for this purpose.

For example, for purposes of aircraft navigation the C.A.A. operates certain 75 mc. marker beacons. One such beacon is the "cone marker" which radiates a cone-shaped beam vertically above each radio range station to indicat definitely its location. Such a beam is shown in Fig. 15(A) with the horizontal pattern shown in Fig. 15(B). A directional transmitting antenna is used to produce this cone-shaped beam. It must be understood, however, that the pattern shown simply represents that for some arbitrarily selected field intensity. If a lower field intensity were selected the horizontal area covered by the beam would be larger.



(A) (B)

Fig. 15

Mounted beneath the airplane and extending *fore and aft* is a horizontal dipole. The receiving characteristics of such an antenna are shown in Fig. 16 and were explained in a preceding lesson. Note that reception from directly ahead or behind will theoretically be zero. Thus the signal in the receiver is very small until the airplane flies directly into the major portion of the beam, i.e., until it is almost directly over the station; and the signal

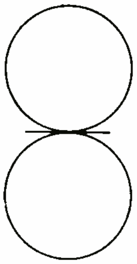


Fig. 16

drops out sharply as the airplane flies out of the vertical beam. Thus the use of a directional receiving antenna increases the overall directivity of the system.

Other uses for directional reception on an airplane involve the use of a loop for direction finding.

By the use of a directional receiving array at the receiving end of a communication circuit that employs a directional array at the transmitting end, a substantial increase in signal gain and overall directivity is obtained.

Increased signal at the receiver input and directional selectivity are important factors in dependable communication on long-distance circuits. Theoretically it is possible to design a receiver for any desired voltage gain and thus theoretically it might be thought possible by this means to operate with a signal voltage of any small amplitude if receiver cost were not a factor. This is not true, however, for two reasons: static and inherent receiver noise.

In the frequency range of long-distance communication, in the order of 5 to 20 mc., natural static often is quite directional. Thus by the use of a directional receiving antenna it is usually possible to obtain substantial discrimination against static interference except during the times when static happens to be coming from exactly the same direction as the received signal. In the northern latitudes much of the prevailing static comes from a southerly direction. Thus on east-west circuits such as between the United States and Europe, considerable static discrimination usually can be obtained by means of a directional antenna. In this case directional discrimination is more important than the signal gain obtained with certain types of directional arrays.

Inherent receiver noise owing to shot effect in the first r.f. tube and thermal agitation of the molecules in the grid circuit of the first r.f. tube set a more serious limitations to the amount of receiver gain it is feasible to use. As the receiver gain is increased to permit the reception of weaker signal voltages, the receiver noise also increases. Thus there is a definite limitation to practical receiver gain, and some means must be provided to assume that adequate signal voltage to override the inherent noise voltage is applied to the receiver input terminals if dependable communication is to be assured. One way of doing this is to design an antenna that simultaneously discriminates against extraneous signals and static, and at the same time provides signal *gain* in the desired direction.

Some types of directional antennas, such as a small loop, provide excellent directional discrimination but very poor signal pickup. Other types such as the wave and broadside antennas provide both directional discrimination and signal gain.

The term "gain," as referred to a receiving antenna, usually means the ratio of the signal voltage developed by a directional array to that developed at the base of a half-wave vertical wire. For example, if the signal voltage E delivered to the receiver terminals by a directional antenna is six times greater than that (E_a) delivered to the receiver by a $\lambda/2$ vertical wire, the directional antenna gain would be

$$\text{Gain in db} = 20 \log_{10} \frac{E}{E_a}$$

$$= 20 \log_{10} 6$$

$$= 20 \times .778 = 15.6 \text{ db}$$

Such an antenna would provide a 15.6 db improvement in signal/noise ratio over a vertical $\lambda/2$ antenna and at the same time provide directional discrimination against undesired radio signals and static. Thus directional receiving antennas are as important as directional transmitting antennas in reliable long-distance communication.

Directional Antenna.--Any directional array used for transmission may be reversed and used as a directional receiving antenna. Such practice has two primary disadvantages: first, the directional transmitting antenna is usually a quite elaborate and expensive installation, much more so than is ordinarily required for reception; second, in order to obtain high radiation efficiency, the transmitting array is designed for selective operation, and for reception would be efficient over only a narrow band of frequencies. In some cases the latter is desirable, but for most receiving installations it is desirable that the receiving equipment have considerable operating flexibility, and high frequency selectivity in the antenna is usually neither necessary nor desirable. The selectivity can better be built into the receiver with the antenna effective over a reasonably wide band of frequencies. This is made possible by the high gain of modern receivers. A principal object in the use of a directional antenna with a modern high-gain receiver is to minimize the pickup of stray noise and interfering signals. This is often more important than to increase the received signal voltage. Of course, any reduction in stray noise and interference permits the use of a correspondingly weaker signal in conjunction with a higher receiver gain to bring it up to the desired output level, with the same signal-to-stray noise ratio.

Loop Antenna.--One of the earliest forms of directional antennas was the loop antenna. One special form, the Alford loop, was described in a previous lesson. The older and more usual form of loop is shown in Fig. 17. It may be of square, round, octagonal, or other shape, and may consist of one turn or many turns. Assume, for simplicity, that it is square in shape and that a vertically *polarized* wave is approaching from the left, as shown in the figure. One can say that the vertical radiation electric field vector produces an electromotive force in vertical conductors 1, 1' and 2, 2' as it passes through them, or, alternatively, one can say that the radiation magnetic field vector (which is horizontal in this case) cuts through them inducing the e.m.f. in the conductors. From either viewpoint it is evident that voltages *in the same upward direction* are induced in the vertical conductors.

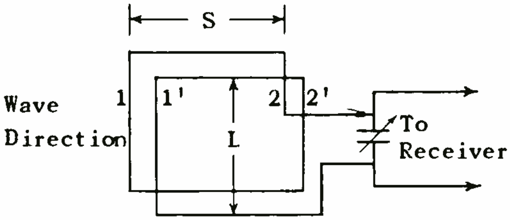


Fig. 17

However, since the wave passes first through conductors 1, 1' and then through 2, 2', the voltage in the latter, call it E_2, will lag that induced in the former, call it E_1, by an amount depending upon the width of the loop S, as compared to the wave length. The phase angle in radians will be equal to

$$\alpha = \frac{2\pi S}{\lambda}$$

This difference in phase produces a *net resultant voltage* E_R *around the loop* as shown in Fig. 18.

It is evident that the *magnitude* of the voltage set up in each vertical conductor is the same, and that no voltage is set up in the horizontal conductors by the vertically polarized wave. The latter simply serve as connectors for the vertical conductors. If ε is the field strength of the wave, say in volts per meter, and L is the length of a conductor in meters, i.e., the length of *one side of one turn* of the loop, then the voltage induced in the conductor is εL, and if there are N turns, then the total voltage induced in all the conductors in one side of the loop is εLN. In the illustration the total voltage induced in the two conductors 1, 1' is $2\varepsilon L$, which has been denoted by E_1, and the total voltage in the other two conductors 2, 2' is also $2\varepsilon L$, and has been denoted by E_2.

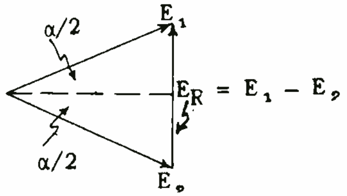


Fig. 18

From Fig. 18 it is evident that E_R is twice the projection of E_1 or E_2 upon it. The projection is clearly $E_1 \sin \alpha/2 = E_2 \sin \alpha/2$ or

$$E_R = 2\varepsilon LN \sin \frac{2\pi S}{2\lambda} = 2\varepsilon LN \sin \frac{\pi S}{\lambda}$$

E_R is small compared to E_1 or E_2 if S is much less than λ, as is usually the case, and hence loops are in general not very efficient antennas.

Suppose, however, the wave direction is not in the plane of the loop as shown in Fig. 17. Suppose it is as shown in Fig. 19, where a plan (top) view of the loop is shown, and the wave direction makes the angle θ with the plane of the loop. Now the difference in phase angle is produced by a shorter difference in path length, namely S cos θ, and the phase angle is therefore

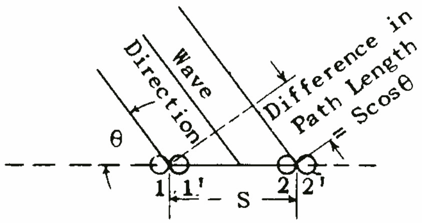


Fig 19

$$\alpha' = \frac{2\pi S \cos \theta}{\lambda}$$

The previous formula must be modified by the introduction of the term cos θ, so that the general formula for the voltage induced in the loop is

$$E_R = 2\varepsilon LN \sin \left(\frac{\pi S}{\lambda} \cos \theta\right)$$

When $\theta = 0$, we have the conditions of Fig. 17 and $\cos \theta = 1$, so that this formula reduces to the preceding one. A further simplification is possible if S is small compared to λ. The angle $\alpha'/2 = \pi S/\lambda \cos \theta$ is then small, and $\sin \alpha'/2$ is practically equal to $\alpha'/2$ in radians. The above formula can then be written as

$$E_R \cong \frac{2\varepsilon LN\pi S}{\lambda} \cos \theta$$

Since L x S equals the area A of the loop, we can finally write

$$E_R \cong \frac{2\pi\varepsilon NA}{\lambda} \cos \theta$$

This formula can be used for loops of shapes other than rectangular, and indicates that it is the area, rather than the shape of the loop, that determines E_R. Since a circle has the greatest area compared to its perimeter, it might be expected that a circular loop would have the greatest voltage pickup compared to its ohmic resistance. Although airplane loops do have a circular shape the above optimum relation of induced voltage to resistance is not too important in practice, and loops of various shapes may be encountered in the field.

In general, as the wave approaches the loop from different directions θ, the induced voltage will vary as shown in the polar diagram of Fig. 20. This is the familiar figure-of-eight pattern. Note that if the wave direction is perpendicular to the plane of the loop ($\theta = 90°$) the resultant voltage is zero. This enables the loop to be employed as a direction finder. Instead of varying the direction of the incident wave the loop is rotated on

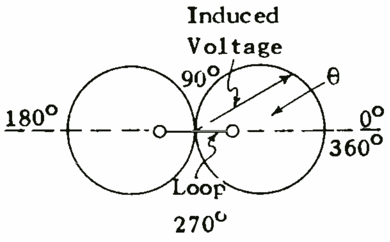


Fig. 20

its axis so that its plane can be set at any angle to the given wave. The direction of the wave can be told either by maximum output when the plane of the loop coincides with the wave direction, or by minimum output (plane of loop at right angles to the wave direction). The latter indication is almost always preferred because the variation in signal for a given small change in θ when θ is around 90° is *much* greater than the variation in signal with change in θ when θ is around 0°, so that the null indication is more sensitive than the maximum indication. This is clearly shown in Fig 20.

The loop finds its major application in direction finders and radio compasses for ship and aeronautical use and will be discussed in more detail in lessons on those applications. Of late the loop has also been employed as an antenna in broadcast receivers, where its compact structure enables it to be built into the radio cabinet, although its directional characteristic is hardly an advantage here and the small pickup is a definite disadvantage which must be compensated for by increased r.f. or i.f. gain in the receiver. However, such gain is feasible, and the loop itself is not as directional as might at first be supposed because of its capacity to ground.

Electrically the loop appears as an inductance if its dimensions are small compared to a wave length, and so it is often tuned by a condenser across its terminals. This in conjunction with the distributed capacity of the loop tunes it to series resonance. The maximum number of turns of the loop is such that it is self-resonant at its highest operating frequency, and in actual practice it is wound with somewhat fewer turns and then may be tuned as described above.

Finally, it may be noted that the loop, even for low-frequency, long-wave pickup, does not require a ground connection, and is very compact. As indicated previously, its signal pickup is low, but especially at the lower frequencies this is compensated for in present-day receivers by high gain with low inherent set noise, and so it can handle signals which in this range are limited by the static that is also picked up.

Beverage Antenna.--Another early and very successful type of directional antenna is the Beverage antenna shown in Fig. 21. This type of antenna was originally used for the reception of very long wave signals, 10,000 to 15,000 meters, in which case the antenna consisted of a copper wire eight or ten miles long supported on ordinary telephone poles. Such an antenna is entirely practical for short-wave reception, in which case it does not have to be so long. When used for short-wave reception the antenna is ordinarily five or six wave lengths long and supported six or eight feet above ground. Thus for reception of wave lengths in the order of 30 meters, an antenna 450 or 500 feet long pointing directly toward the transmitter from which reception is desired will give excellent results. Such an antenna has little frequency selectivity. The greater the length, in terms of wave lengths, the greater will be the directional effect. The operation of such an antenna is simple. As the signal voltage wave coming from the direction shown *away* from the receiver, sweeps along the wire toward the receiver, a moving voltage wave is built up along the wire owing to the slope of the wave front; the current owing to the voltage wave passes through coil L to ground. By means of L the signal voltage is induced across the receiver circuit, energizing the receiver. The slope of the wave front is owing to the fact that the lower portion of the wave is retarded as it travels over the surface of the earth. This slope also could be owing to the skywave approaching the surface of the earth at an angle and not directly parallel to the surface.

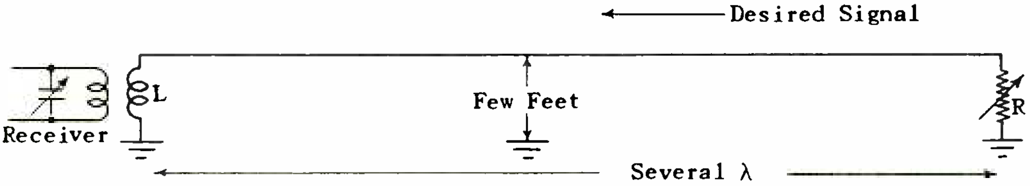


Fig. 21

Note resistor R between the extreme end of the antenna and ground. This resistance, when adjusted to equal the surge impedance of the line, makes the antenna unidirectional. (Such an antenna functions as a single-wire transmission line.) If the extreme end of the antenna were open a signal from the opposite direction would build up a voltage along the wire. When the voltage wave hit the infinite impedance of the open end of the wire it would be reflected back along the line and would energize the receiver just as did the desired signal. When the resistance is connected and *properly adjusted*, the undesired signal is dissipated in R and no reflection occurs; reception is then unidirectional. R is ordinarily in the order of 400 or 500 ohms. A variable resistor adjustable between 100 and 1,000 ohms should take care of any such installation. R is adjusted until reception from a station in a direction opposite to which the antenna points disappears. A wave coming from a direction broadside to the antenna does not sweep along the wire so that the voltage built up across the

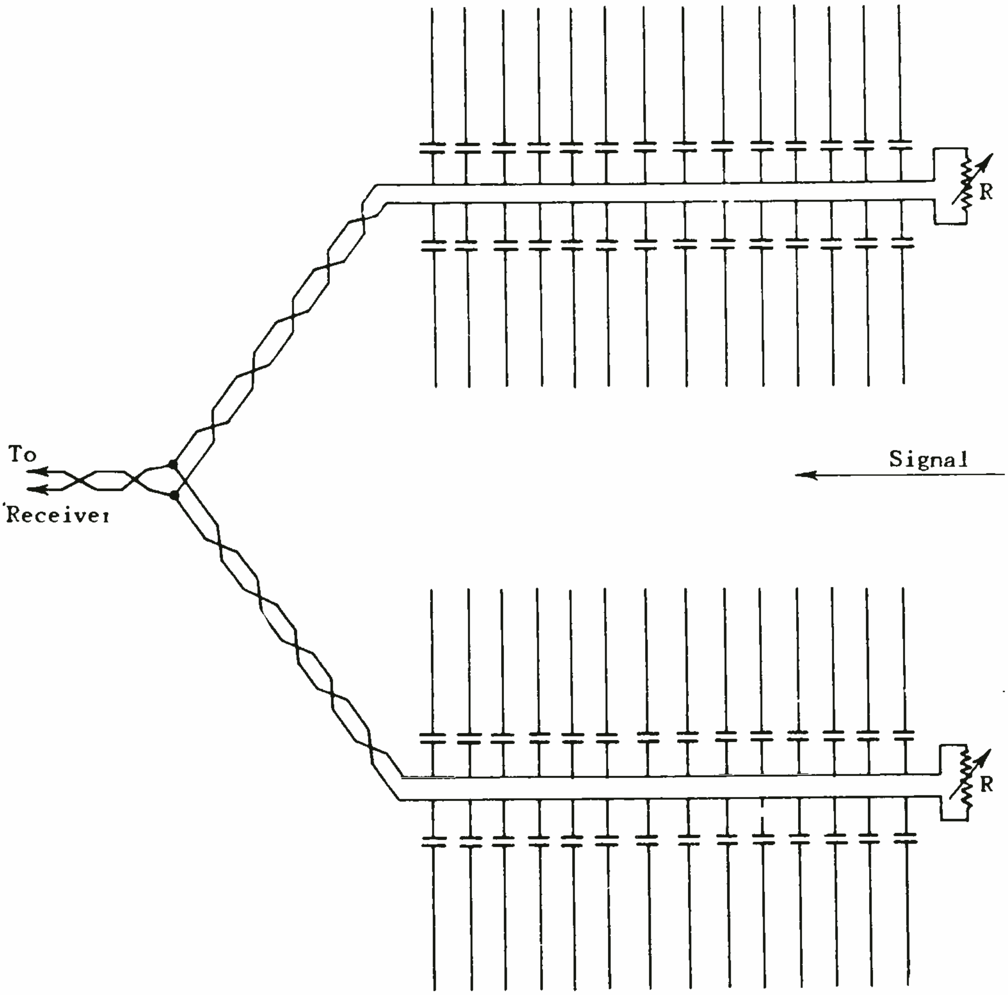


Fig. 22

wire by such a signal is ordinarily negligible.

If it is necessary to be able to receive directively from several directions, the required number of antennas may be built, radially from the receiver building, and the receiver connected to the proper antenna to receive any desired signal. It is desirable that the Beverage antenna be quite low because any effective height will result in a vertical component of voltage which may be received from all directions. Thus the greater the height of the antenna above ground, the less its directional selectivity. It should be noted that the operation of this antenna is very similar to that of a single-wire untuned transmission line.

RCA Broadside Antenna and Diversity Array.--Fig. 22 shows the RCA broadside antenna which is one section of a diversity array. While the Beverage antenna of Fig. 21 acts as a single-wire transmission line, each half of the array of Fig. 22 acts as a two-wire transmission line to which are coupled, to increase the pickup and properly load the line, a number of short horizontal members oriented at right angles to the main line. These members are short and are not resonant in the desired band of frequencies to be received so that the frequency response is quite broad. The connected wires should be spaced not more than 3/8 λ at the highest frequency to be received. As the signal wave comes from the direction as shown, it sweeps along the array and the voltages built up across the individual members are successively coupled to the central transmission line and then by means of an untuned transmission line, to the receiver which may be 1,000 feet or more distant.

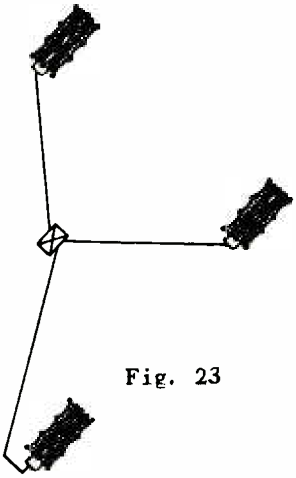


Fig. 23

The complete arrangement is as shown in Fig. 23. The view as shown in Fig. 22 is from above, the entire array being suspended horizontally about 50 feet above ground between wooden poles and supporting cables. Unidirectional reception is obtained as in Fig. 21, by terminating the otherwise open end of the line with a dissipating resistance. Two parallel sections are used in each array unit as this increases the directivity. A complete diversity array consists of three units as shown in Fig. 23, each complete in itself as in Fig. 22, and the three units separated about 1,000 feet from each other. Each unit is coupled by means of a transmission line to one unit of a diversity receiver as explained earlier in this lesson.

By close spacing of wires and transposition every few feet, the pickup by the long transmission line between the array and the receiver is made negligible. It is on diversity arrays such as this that foreign programs are picked up for rebroadcasting, and such receiving systems make possible high-speed, high-frequency, long-distance communication with an extremely high factor of reliability.

EXAMINATION *RADIO WAVE PROPAGATION—RECEIVING ANTENNAS*

1. (a) In what way did Marconi's demonstration of long distance transmission apparently contradict the laws of optics?

   (b) How is transmission over very great distances accomplished? Explain briefly.

2. (a) What does the ground wave consist of?

   (b) What component is effective in medium frequency transmission over moderate distances? Explain fully.

   (c) What component or components are effective in ultra-high frequency transmission?

3. (a) Describe the characteristics of the ionosphere.

   (b) Name the various layers and describe their positions at various times of the day and various seasons of the year.

4. (a) How is the virtual height of a layer measured?

   (b) What is a "critical" frequency?

   (c) At what time of the day do the greatest variations in the medium of propagation occur?

5. (a) What is meant by "Skip Distance"?

   (b) What is the cause of skip distance? Explain.

   (c) Why are the losses in the case of skip distance transmission so low?

6. In the eastern part of the United States, what frequencies of European short wave broadcasts should normally come through well in the morning; in the late afternoon; in the evening? Why, in each case?

7. (a) What is an "Echo Signal" and how is it caused?

   (b) What are its effects on communication? Why?

8. (a) Given a vertical Marconi antenna which is radiating at a frequency of 1 mc. Describe the polarization of the wave picked up at a distance of 25 miles; and at 500 miles.

   (b) What is meant by fading and what is one of the principal causes of fading? When is fading most likely to occur? Why? Explain fully.

   (c) Explain the "diversity" method of minimizing the effects of fading. Upon what fact is this system based? Why is the diversity system impractical for use with the ordinary broadcast receiver?

9. Given a rectangular loop whose sides are each 3 feet long, and whose top and bottom are each 2½ feet long. There are 10 turns in the loop. The frequency is 1.5 mc. and the field strength is 2 mv. per meter.

   (a) What is the voltage induced in the loop when the wave arrives in the plane of the loop, with the polarization parallel to the sides?

   (b) What is the voltage when the same wave arrives at an angle of 30° to the plane of the loop?

10. Explain the operation and construction of the Beverage receiving antenna. How is this antenna made unidirectional in operation? Why is such an antenna more practical for high frequency operation than for low frequency work? Explain. How high should it be? Why?