U H F RADIO

Simplified

MILTON S. KIVER

Registered Professional Engineer, State of Illinois

D. VAN NOSTRAND COMPANY, INC.
TORONTO NEW YORK LONDON

NEW YORK
D. Van Nostrand Company, Inc., 250 Fourth Avenue, New York 3
TORONTO
D. Van Nostrand Company (Canada), Ltd., 228 Bloor Street, Toronto
LONDON
Macmillan & Company, Ltd., St. Martin's Street, London, W.C. 2

First Published June 1945

Produced by
TECHNICAL COMPOSITION CO.
BOSTON, MASS.

04497a25
PRINTED IN THE UNITED STATES OF AMERICA

UHF RADIO
Simplified

BY THE SAME AUTHOR

TELEVISION Simplified

PREFACE

Ultra-high frequency radio was literally thrust upon us by the urgent necessities of war. Detection devices, which safeguard the country's inner defenses, were the primary reasons. For the moment these prevail. But it requires no great foresight to envision the future use of the higher frequencies. Already, Television is probing into these regions, seeking a suitable resting place where thousand-line colored images may be developed. Progress in this direction is far beyond the wishful-thinking stage. The F.C.C., in assigning frequency allocations, definitely indicates that this is the trend. We need no better proof.

With this and other new forms of radio elbowing sharply for room, there is only one direction to go — up, up to the higher frequencies, where space abounds. Hence it behooves those who follow the radio field, either as a hobby or for a livelihood, to understand the elements of the seemingly newer circuits. It was with this in mind — namely, to point out the underlying equality of all radio, that this book was written. The concepts of u.h.f. radio are presented as logical outgrowths of the more familiar low-frequency equipment. With what we know as a guide, the reasons for modifications become evident and fall more naturally into place.

At various points throughout the book, explanations are advanced with more of an eye toward easier comprehension than mathematical rigor. For that, no excuse is offered. However, at no point are the basic truths ever consciously distorted. In keeping with the character of the book, they are merely simplified. There are many excellent theoretical treatises available, and the more inquisitive reader is directed to them.

Liberal use is made of original diagrams published in the Pro-

ceedings of the Institute of Radio Engineers, especially in Chapter 7. The author is grateful to the Institute for permission to use these drawings. Due acknowledgment is also given to William Stocklin, associate editor of Radio News, who first suggested the need for such a book, to Morton Whitman, for his valuable criticisms, and to my wife, Ruth Ann, for her aid and encouragement throughout the entire writing. Thanks are also given to the General Radio Company, the RCA Manufacturing Co., the Sylvania Electric Products Co., the Western Electrical Instrument Corporation, and the Sperry-Gyroscope Company for the material they so readily furnished.

M. S. K.

February, 1945
Chicago, Illinois

CONTENTS

CHAPTER 1

INTRODUCTION TO THE HIGHER FREQUENCIES

Being concerned with the transition from the low frequencies to the ultra-highs

With the advent of the many detection devices which have been developed during recent years, the field of radio has expanded sharply. As a result, the communications spotlight now definitely points to the ultra-high frequencies. However, these devices and their associated apparatus are not entirely new but are based primarily on radio as we know it currently, supplemented by modifications applicable to the higher frequencies.

Effect of Increasing Frequency. In order to demonstrate what happens to ordinary radio apparatus when the frequency is raised, take a common oscillator and attempt to increase its output frequency. The relationship between frequency and coils and condensers is given by the formula

$$F = \frac{1}{2\pi\sqrt{LC}}$$

where L is in henries,
C is in farads.

In order to have the frequency F increase, therefore, either L or C or both must be decreased. This means using fewer turns for the coil and fewer plates for the tuning condenser. But there is a limit to this process and at the end nothing would remain of the condenser except two small plates, and of the inductance, a turn or two of wire. In Fig. 1.1 this process has been applied to the well known Hartley oscillator. Illustration A shows the usual form of this oscillator encountered at the low frequencies, and

B shows how a higher frequency (shorter wavelength) is obtained by utilizing a smaller capacitance condenser and a coil of few turns. With these modifications, frequencies up to 200 megacycles can be generated.

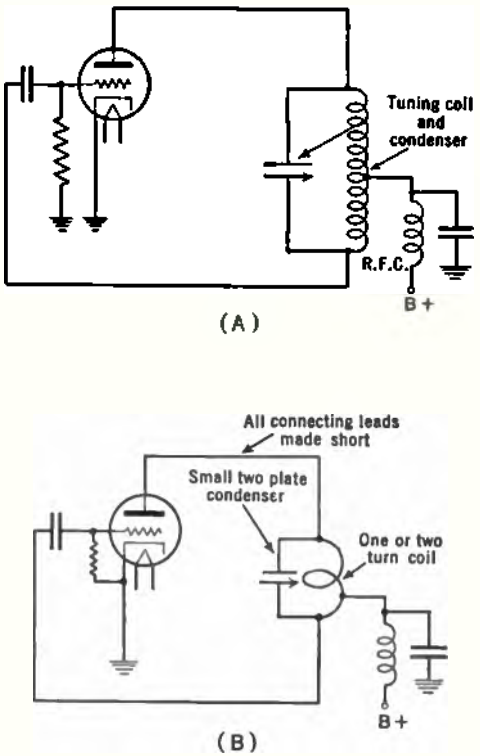


FIG. 1.1. The above illustrations show the transition in oscillator design from the low frequencies to the U.H.F.

In order further to reduce the wavelength obtainable and at the same time minimize losses due to poor insulation, the construction of the tube was next attacked. Dimensions were reduced until the tube had shrunk to the " acorn " and " doorknob " sizes. See Fig. 1.2 where the high-frequency tubes 9001, 9002, 826, 954, and 958 are shown. Leads from the elements are now brought directly through the glass envelope instead of using a base

as is done in the ordinary tubes. Isolantite and polystyrene insulators are used throughout further to minimize losses due to dielectric leakage.

By decreasing the area of the various elements, the interelectrode capacitance was reduced. By using very short, direct leads from the elements themselves, the inductance was likewise lessened. And the need for polystyrene and isolantite insulators arose when it was discovered that many substances which acted as insulators at the low broadcast frequencies became partial con-

Courtesy of R. C. A.

Fig. 1.2. Number of different tubes that have been designed for ultra-high frequency applications.

ductors at the ultra-highs. The ordinary tube socket is a good example of what is meant. It works well for the ordinary wavelengths but introduces large losses due to leakage currents at the very high frequencies.

Thus the final oscillator for the high frequencies, as described so far, will be dependent on a small tuning condenser and a coil of few turns for determining the frequency that can be generated. Not a very flexible arrangement, as you can see, since neither the tuning capacitance nor the coil is very large in physical size and hence not easily handled. What is needed is some sort of tuning circuit that would be large enough to handle and at the same time have good selectivity or, as it is commonly called, have a good Q.*

*This should not be confused with the same letter used for quantity of electricity.

Selective Tuning — *Q*. *Q*, at times, may become quite puzzling and so a little discussion about it may not be out of place. The purpose of the tube, in any ordinary oscillating circuit, is to apply energy to the resonant circuit in order to keep the currents circulating. The less resistance there is in a circuit the less energy will be dissipated as heat. Since, in a tuning circuit, almost all the resistance is associated with the inductance, the ratio of the inductive reactance of a coil to its resistance has been called *Q*. The ratio may be looked upon as a relationship between the amount of energy stored in the magnetic field of the coil to the energy lost in the coil's resistance. The greater this ratio, the sharper will be the resonance curve and the more selective the circuit. Adding resistance will usually flatten this curve and thus give a broad tuning effect. In oscillators it is generally more desirable to have sharp tuning. Consequently, coils having high *Q*'s are used. Another factor associated with having a circuit of high *Q* is that the circulating current within the resonant circuit is greater, giving rise to a much larger output. Hence the importance of having a circuit with a high *Q*.

In the modified oscillator just discussed, the *Q* of the inductance used was very small and gave rise to weak output power and poor frequency stability. It was found, however, that concentric lines and Lecher wire systems had the three desirable properties of a tuned circuit and so came into widespread use. The three requirements are:

1. A high *Q* or, what is the same thing, low ohmic, eddy current, and dielectric losses.

2. Ability to resonate accurately with the oscillator so as to build up large values of voltage and current.

3. And, lastly, to stabilize the oscillator and allow the least amount of energy to be wasted by unwanted radiation.

In the Lecher wire system shown in Fig. 1.3A, consisting of a pair of parallel wires, resonance is accomplished by means of a shorting bar which can be moved easily along the wires until the resonant point is found. The shorting bar changes the effective length of the system and hence changes the resonant frequency.

It will be shown in a later chapter on transmission lines that standing waves can be formed in this way. All that is necessary to know now is that the tube sees a large impedance where it joins the Lecher system. This effect is the same as in an ordinary circuit where, at the resonant frequency, a parallel tuned circuit presents a large impedance. Lecher systems are usually a quarter wavelength or some odd multiple long, because, with a shorting bar, this length is required in order to have the necessary large impedance values at the open end.

The loss due to radiation of the concentric resonant line shown in Fig. 1.3B is appreciably less than in the open wire arrangement of the Lecher wire system. The Q of the concentric line is very high, as much as 15,000 or more. In order to make this line flexible, a movable shorting stub is used and the effective length adjusted to a quarter wavelength or some odd multiple thereof. Furthermore, as will be shown in Chapter 4, where the concentric line is tuned so that there is a standing wave with a current node (a zero value) at the open end of the line, the high-frequency current is confined to the inside of the lines. Direct voltages can then be applied directly to the concentric lines without having the high-frequency currents flowing in the supply lines. Another advantage obtained by confining the electric fields inside the concentric lines is the absence of body capacitance effect on tuning.

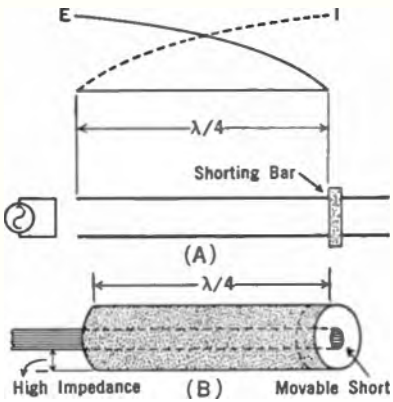


FIG. 1.3. (A) The voltage and current distribution on a Lecher wire tuning system, for that particular position of the shorting bar. (B) A concentric line resonant circuit.

High-Frequency Oscillators. Now that the concentric lines and Lecher wire systems have been discussed as possible tuning circuits in high-frequency sets, it would be instructive to see the differences between the conventional low-frequency oscillator and

the high-frequency modification using a Lecher wire system. This is shown in Fig. 1.4, where the ultraudion oscillator schematic is given. Fig. 1.4A is the modification of Fig. 1.4B. Two differences are noticeable. In part A, the tuning circuit consists of a quarter-wave Lecher wire system instead of a coil and condenser. The other change is in the filament circuit where a filter circuit has been inserted to keep the ultra-high frequency currents out of the filament supply lines. Instead of actual coils and condensers, it would have been possible to use a short length of concentric transmission line to obtain the same filtering effect (see Chapter 4). The oscillator in Fig. 1.6 has this arrangement. To transfer energy from this oscillator to an antenna or an amplifier, a small loop of wire is placed near the line to pick up the required energy.

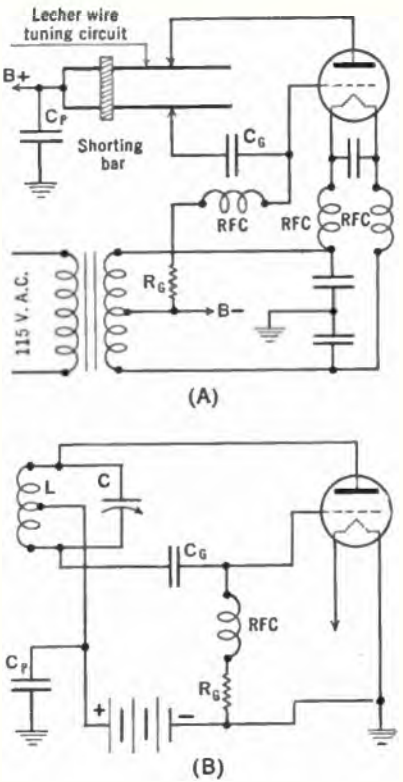


FIG. 1.4. Two arrangements of the same oscillator. (A) is suitable for U.H.F. while (B) is the low frequency version.

For a push-pull arrangement, Fig. 1.5 might be suggested. Here again quarter-wave transmission tuners and shorting bars are used as the resonant circuits; one for the plate and one for the grid. To tune this oscillator, the grid resonant circuit is first set, and then the shorting bar in the plate circuit is slowly moved until a point of resonance is found. The output is taken from a point of high r.f. voltage by means of a pair of condensers. All leads in these high-frequency oscillators must be as short as possible.

One of the shortest wavelengths reached by these oscillators (about 25 centimeters) utilizes the 8012 tube. A circuit diagram of an oscillator of this type is shown in Fig. 1.6. Due to the double-ended construction of the 8012 tube, it is possible to place it at the center of a half-wave concentric line having closed ends. This gives the equivalent of using two tubes in parallel, each tuned by a quarter-wave line. Because the lumped interelectrode

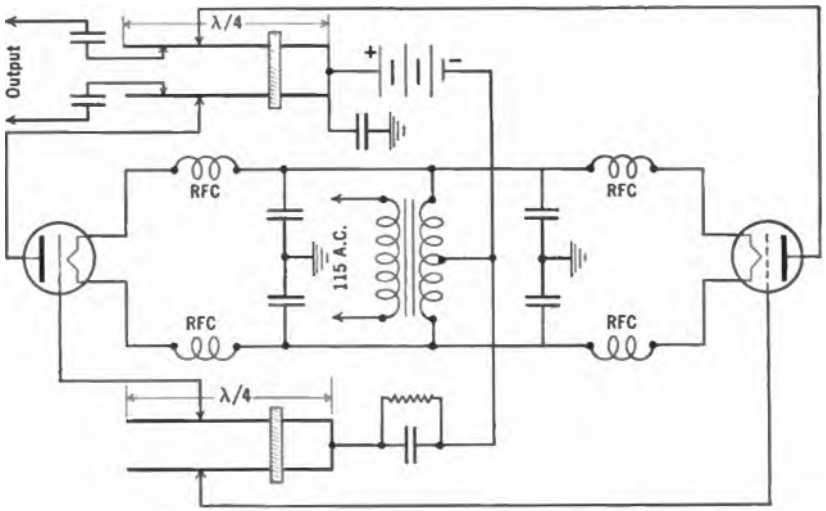


FIG. 1.5. An ultra-high frequency oscillator using two tubes in push-pull.

capacitance may be assumed to be divided between the two quarter-wave halves of the concentric line system, the frequency of oscillation is higher than in the single-ended case, where all the capacitance is associated with one quarter-wave circuit. Also, as only half the charging current flows through each set of leads, there is a decrease in the amount of power lost. The above idea of paralleling quarter-wave lines will be brought up again when cavity resonators are developed.

While the modifications of reduced tube size and Lecher systems succeeded in raising the frequency that could be generated, a limit was nevertheless reached somewhere in the neighborhood of 600 megacycles. This represents the usable limit, while the

25 cm. mentioned before represents the experimental limit where little power output can be obtained. The reason for this is not in the Lecher tuning system but rather in something else, something which in ordinary tubes is completely disregarded, namely, the time of flight of an electron from the cathode to the plate, called transit time.

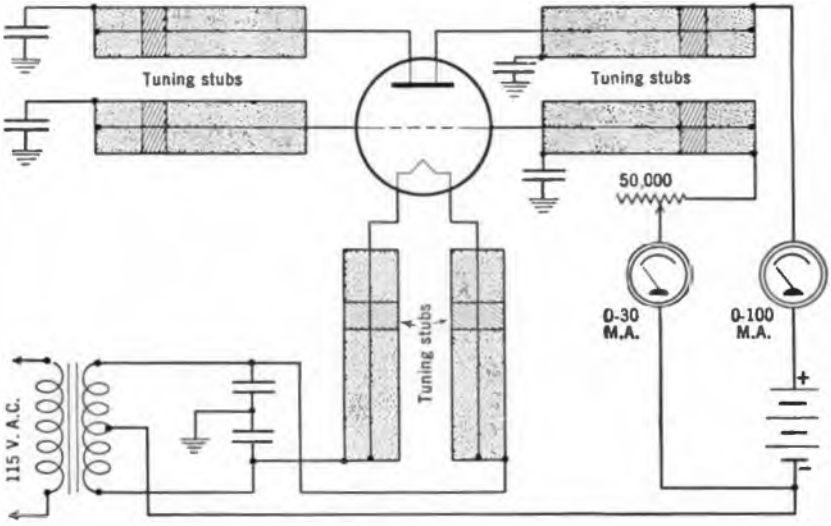


FIG. 1.6. A double-ended tube, with its two terminals for each grid and plate, will generate higher frequencies than conventional tubes.

Transit Time. At the low frequencies, the time it takes an electron to travel the interelectrode distance is negligible, because this time represents a very small portion of the interval necessary to complete one cycle of the grid or input a.c. wave. As the frequency of the input wave is increased, however, the time for one complete cycle of the a.c. wave becomes less and less, and soon the electron transit time becomes comparable to the period of the alternating wave. An example will make this clearer. Suppose that it takes an electron 0.000,000,001 sec. to travel from the filament to the plate. This may be ignored when the frequency of the input wave to the grid is, say, 1000 cycles per sec., for then one complete cycle requires 0.001 sec. This is a very long interval

compared to the electron transit time given just above. But suppose that the frequency of the a.c. wave is raised to 1000 megacycles. Now, one cycle is completed in 0.000,000,001 sec. or in the same time that it takes the electron to travel from filament to plate. When this happens, something occurs in the grid circuit that causes the grid-to-filament (or cathode, as the case may be) impedance to drop from its ordinary value (which is so high it can be considered as infinite). This affects the whole grid tuning circuit because the grid-to-cathode impedance is really in parallel with this tuner. Thus, if this high resistance is lowered in value, the same result is obtained as though a low resistance were placed in parallel with the resonant circuit. The *Q* is lowered, and with it goes the efficiency of the circuit.

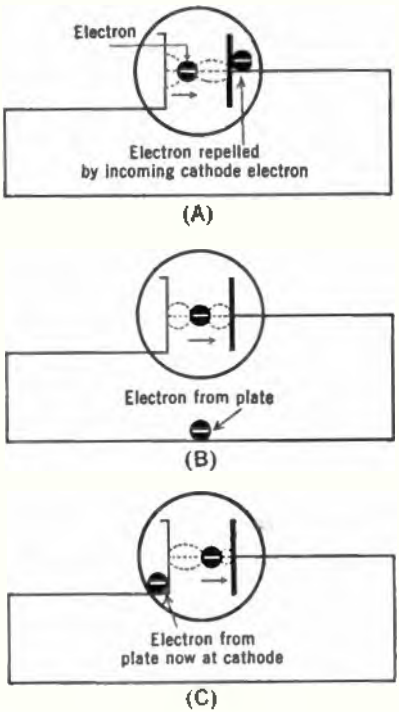


FIG. 1.7. Electron flow in a vacuum tube and its external circuit.

Electron Flow—Revised. In order to help visualize the entire process, some ideas on electron flow will be revised, or rather extended. Ordinarily, it is unusual to think of current flowing in either the plate or grid circuits until the electrons from the cathode hit these elements. Actually, however, this is the point that completes a cycle begun when the electron first left the cathode. The charge of the electron is negative, and at the moment it leaves the cathode the conditions shown in Fig. 1.7A prevail. Here it can be seen that on leaving the cathode the electron left an equal and opposite (positive) charge on the cathode. At the same time a very small positive charge is induced

on the plate. This is because the cathode electron is slightly closer to the plate and an electron on the plate has been repelled a small distance from the anode surface. As the cathode electron gets closer and closer to the plate, the positive charge of this anode will correspondingly increase, and the electron that was originally at the plate will be farther and farther repelled from the anode toward the cathode through the wire. See Fig. 1.7B. When the cathode electron finally reaches the plate, it neutralizes the positive charge there, and likewise the anode electron by this time has reached the cathode. The positive charge on the cathode is now also neutralized. The circuit is then in equilibrium and the current flow has ceased (see Fig. 1.7C). Multiply this electron by the billions that actually flow and there are the comparatively large plate currents observed in tubes. Some theorists may demand that it be explained that the electron that left the plate be shown as only going a small distance and pushing other electrons in the plate lead until one finally goes to the cathode and neutralizes the positive charge left there by the cathode electron that left for the plate. The plate electron has been described as going the full distance from plate to cathode. Be that as it may, the fact still remains that, when the cathode electron reaches the plate, the current stops. This is the important idea to keep in mind.

The above concepts are not new but have seldom been mentioned because, at the low frequencies, results obtained agree with the ordinary ideas of electron flow in tubes. With increase of frequency, however, the transit time effects as far as the grid is concerned require that this more general idea be used. Now the effect on the grid will be determined.

When an electron approaches the grid, there is induced in the grid a small charge, with the result that a small current starts to flow. This is derived from the displacement of the electrons in the grid wires caused by the oncoming electron. If the time it takes the a.c. grid voltage to change is slow compared to the speed of the electron, this induced current caused by the approaching electron will be equal and opposite to the induced current when

the electron passes the grid and goes on to the plate. The two currents — that due to the approaching electron and that due to the same electron as it recedes on the other side of the grid — will thus cancel. As the frequency of the a.c. grid voltage is increased, however, the grid voltage will change appreciably in the time the electron is approaching the grid as compared with the time it recedes on the other side of this element, and the induced currents due to this moving electron will not be equal in phase and thus will not cancel. This will give rise to a resultant current which, it has been found, causes losses in the grid circuit. This is known as the transit time effect in tubes and begins to act when the value of the a.c. grid voltage changes appreciably in the time it takes the electron to move through the tube. A specific example might make this much clearer. A wave having a frequency of 3000 megacycles per second, has one complete cycle occur in 1/3,000,000,000 of a second, or 0.000,000,000,3 sec. In ordinary tubes, the time of flight of an electron from the cathode to the plate is approximately 0.000,000,001 sec. Thus three complete cycles of a 3000-megacycle wave will occur while the electron moves from the cathode to the plate region. Any current induced in the grid by this electron as it approaches this element will occur at a different time in the a.c. grid voltage cycle than when the electron recedes from the grid on the other side. It is this effect that gives rise to losses in the grid circuit. The end result is a lowering of the usually high resistance between the grid and cathode. If the grid-to-cathode resistance drops, any tuned circuit in between these elements will be affected and the selectivity will drop.

Electron transit time action in a tube is a very complicated concept to grasp and, yet, it is so important that perhaps another impression, stated a little differently, might give rise to a mental picture of what takes place. Consider Fig. 1.8A, where the cathode and grid are replaced by a small condenser equal to the interelectrode capacitance. The a.c. generator inserted across this condenser is the input signal. Now, it is well known that, if the current responds instantly to voltage changes across this condenser, the

current will lead the voltage by 90°. The graph in the same figure, Fig. 1.8B, shows this. But suppose, for one reason or another, that the current cannot follow the voltage variations instantly. Then the current will not be 90° out of phase with the voltage but

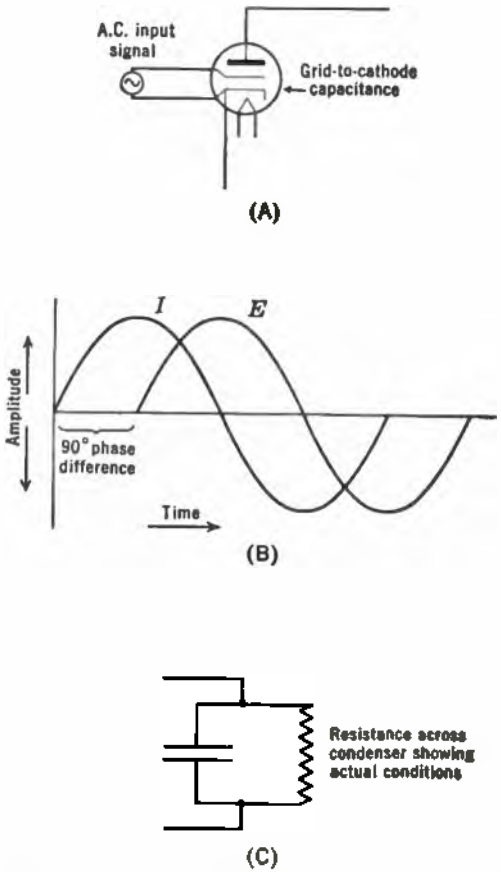


FIG. 1.8. (A) Replacing the grid and cathode elements by an equivalent capacitance. (B) Voltage and current phase relationship across a perfect condenser.

will have some other angle, usually less than 90°. This is equivalent to having a condenser and resistor combination, such as shown in Fig. 1.8C. For the first case (A) we can see that, if the current and voltage were 90° out of phase, the action would be

purely capacitive. But if the angle is something less than 90°, the equivalent circuit consists of a condenser with a resistor shunted across it.

To carry the illustration one step farther, replace the a.c. grid generator by a resonant circuit which can give rise to the same type of voltage. The resistance shown in Fig. 1.8C and mentioned just above will be in parallel with this tuned circuit and thus tend to lower its Q and selectivity and thus increase its losses. The tuning circuit may even stop functioning as such if the resistance should get small enough. It is due to this last mentioned effect that the transit time in a tube is important. If the conditions in the grid circuit (yes, and even the plate circuit) were not affected by the electron time of flight and the frequency of the grid voltage, then all this discussion would have been omitted. It is because of the general lowering in efficiency that such great pains have been taken to explain electron transit time in detail.

Mathematical formulas have been derived to give the exact relationship between conductance of the grid and the frequency. One such relationship is given here and is due to W. R. Ferris.[1]

$$G_g = Kg_m f^2 T^2$$

where G_g is the input conductance between the grid and cathode. (Conductance is the reciprocal of resistance.)
K is a constant of the tube.
g_m is the mutual conductance of the tube.
f is the frequency of the input signal.
T is the time it takes an electron to move from the cathode to the plate.

Inspection of the above equation reveals that the conductance rapidly increases as the frequency is raised due to the fact that the frequency term is squared.

Minimizing Transit Time Effects. Since the resistance between the grid and the cathode decreases with frequency, and since this resistance would shunt any tuned circuit placed between

these two elements, it can again be seen what an adverse effect this would have. One way to overcome the effect of transit time is to increase the electron velocity until the time it takes to reach the plate from the cathode again becomes negligible in comparison with any alternating voltage applied to the grid. A method of accomplishing this consists in placing higher voltages on the plate so as to attract the electron that much more and thus cause it to travel with greater speed. The electron, however, now that it has this added speed, will hit the plate harder and cause it to get hotter. Means must be devised to cope with this increased plate dissipation.

Another method suggested involves the moving of the electrodes closer together. This would again counteract transit time effects of electrons because now the electron has less distance to travel and any voltage placed on the plate will attract the electron more strongly. But there is a limit to the spacing of the electrodes without having each affect the other adversely. And with elements closer together, less heat can be dissipated at the plate, and the lower the amplification factor of the tube becomes. Thus an impasse has been reached and other ways must be found that will allow large amounts of power to be efficiently generated at the very high frequencies.

Barkhausen and Kurz Oscillator. One of the first attempts to get away from using the triode in its conventional manner was made by Barkhausen and Kurz in 1920. They discovered that, by placing a large positive voltage on the grid and either a zero or slightly negative voltage on the plate, oscillations would be produced when a tuned circuit was placed externally between the plate and grid. Over a short range of grid and plate voltages, it was found that the frequency of oscillation was independent of the Lecher wire system attached to the tube and was dependent only on the time of transit of the electrons within the tube that gave rise to these oscillations. Of course, greater output could be obtained when the frequency of oscillation of the electrons was the same as the resonant frequency of the Lecher wire system attached to the grid and plate. But the important point was that, even

if the resonant frequency of the Lecher system was changed, it would not affect the frequency of the electrons in the tube. Wavelengths generated by Barkhausen and Kurz varied from 43 to 200 cm. and were found related to the grid potential by the formula

$$\lambda^2 E_g = K$$

where E_g is the grid voltage
K is a constant of the circuit
λ is the wavelength generated.

In 1922, Gill and Morrell, working with an oscillator similar to the B-K oscillator, found that for the range from 200 to 500 cm. the generated frequencies were not independent of the external circuit but varied slightly with it. The formula for the frequencies of oscillation put forth by Gill and Morrell was

$$\frac{\lambda^2 I_g}{E_g^{1/2}} = K$$

which reduces to the B-K formula if the grid current, I_g, varies as the 3/2 power of E_g. The point to be noted is that in the Barkhausen-Kurz setup the frequency of oscillation does not change by changing the tuning circuit attached to the tube, while in the Gill-Morrell oscillator the circuit will not work correctly unless the electron oscillations in the tube are properly related to the period of the tuned circuit. It is quite obvious that the two oscillators are interrelated, and it has been found that one (G-M) will smoothly and gradually work into the other as the frequency is changed. There are many things that are still unknown about their operation and it may take some time before all the little discrepancies are cleared up to everyone's satisfaction. Electron transit time, which caused the conventional type of oscillator to fail at the high frequencies, is the operating principle of these two generators. Most of the oscillators to follow in the next two chapters will have arrangements whereby transit time is used to advantage rather than as a liability. It is the old case of using an opponent to advantage if he cannot be beaten.

How These Oscillators Function. The generation of oscillations in a positive grid oscillator (shown in Fig. 1.9) can be explained best if the motion of the electrons under two conditions is studied: (1) when there is a constant difference of potential between plate and grid, and (2) when the grid has an alternating voltage applied to it with respect to the cathode. The period of this alternating voltage should be close to the time it takes an electron to reach a point near the plate where the electrons reverse their direction.

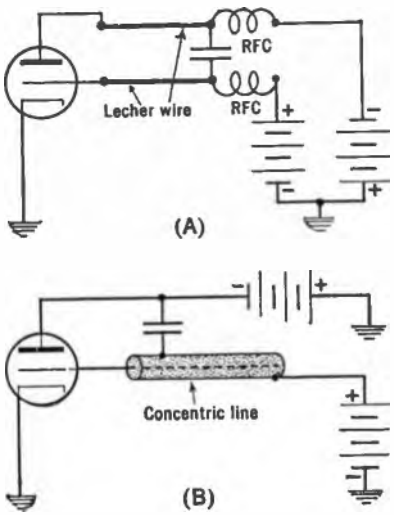


FIG. 1.9. The Barkhausen-Kurz oscillator with two different tuning arrangements. The inner and outer conductors of the concentric line are shorted together only at battery end.

Dealing with the first condition, the electron from the cathode is accelerated by the grid (which has a positive voltage on it) and, if it misses the grid wires, will continue on toward the plate. The electron, however, now is moving away from a highly positive potential and toward a slightly negative plate. Its velocity will therefore decrease until the electron stops somewhere near the plate. It will be attracted again by the positive grid potential and so be forced to speed toward it. If the grid is missed, the cathode will probably receive it. This action completes one cycle and the net change of energy of the electron is zero. The electron gained energy when it was speeded up by the positive voltage on the grid and it lost energy when it drew away from the grid and was slowed down. In the entire process there were two times when the electron was accelerated and two times when it was slowed down. The net result — no energy loss or gain. This point will be brought up again in the next few paragraphs.

The next condition calls for the superimposing of an alternating voltage on the direct potential of the grid. The period for a complete cycle of this alternating voltage is to be equal to the travel time of the electron from the cathode to the vicinity of the plate. In order to visualize the following discussion, refer to Fig. 1.10A, where the position of the electron is indicated at various times of the a.c. grid voltage.

Consider, first, time t_1, when the electron is leaving the cathode and the voltage on the grid is increasing. Because of the increased voltage, the electron will be accelerated more than it normally would have been without the superimposed a.c. voltage. At time t_2, the electron is just passing the plane of the grid and, since the grid is now going less positive than normal, due to the reversal of the a.c. voltage, the electron will travel on toward the plate with less deceleration and probably strike the anode. The electron, it can be seen, has gained more energy in this trip than the electron traveling in the same direction under condition 1 (see above), and this energy must have come from the alternating voltage on the grid. In case the electron does not gain enough *additional* energy to strike the plate, then at time t_4 it will return past the grid and at t_5 will strike the cathode with excess energy. Whether or not an electron strikes the plate, on its first trip under the above conditions, depends on the grid-to-plate distance and the alternating voltage on the grid. In any event, if the electron starts out at the time mentioned, it will absorb energy and so be

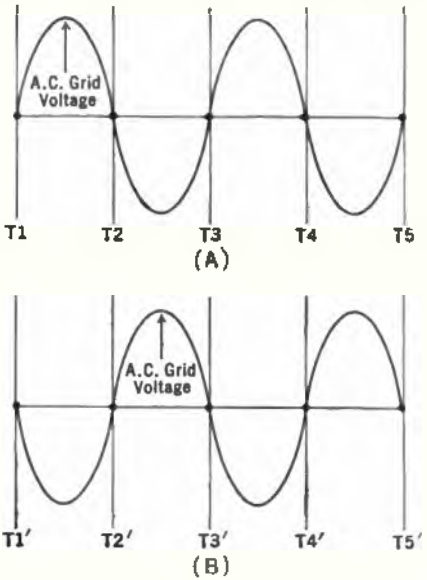


FIG. 1.10. These diagrams will aid in explaining the action of an electron in a vacuum tube with different voltages on the grid.

useless as far as the generation of oscillations is concerned. Electrons are needed which give up energy, not absorb it.

Referring to Fig. 1.10B, it can be seen that an electron leaving the cathode, when the grid a.c. voltage is zero and going negative, will be accelerated less than if no a.c. voltage were present. It should not be forgotten that although the a.c. voltage is going negative, there is a much larger d.c. potential on the grid (which is positive in value) and thus the resultant voltage on the grid is always positive. All the a.c. voltage does is to vary the positive potential on the grid above and below a certain fixed value. Returning to the discussion of the electron: at time t_2', the electron has reached the plane of the grid and is advancing toward the plate. As the grid now is going more positive than the normal d.c. value, the electron will have a greater deceleration and so will not approach as close to the plate as with constant grid voltage. On the return cycle, time t_4' will again find the electron at the grid and time t_5' will find it coming to a halt some distance from the cathode. The fact that the electron came to rest at a point farther from the plate than it normally would have with constant grid voltage indicates that it had lost some energy. The recipient of this energy was the source of alternating voltage on the grid.

Some readers may be a bit puzzled as to why the electron will hit the plate on one trip and come to a stop some distance from it on another trip. The answer lies in the fact that, when the electron is going away from a grid which is more positive than normal, this electron, being negative in nature and so attracted to a positive grid, will suffer a greater deceleration than it formerly did. Likewise, when it is going away from a grid that is less positive than normal, it will suffer less deceleration than formerly and so go farther before coming to a stop. In each case, the electron will be decelerated, but in different amounts.

Now return to the last electron considered, the one that started out from the cathode when the a.c. grid voltage was going negative. It was seen that this electron did not reach the plate or the cathode again because it gave up energy to the grid on each

trip. An electron loses energy, or gives it up (so to speak), when it is working against the grid voltage. The electron that has lost energy will continue to make trips back and forth between the cathode and plate, each time losing a little energy and so moving shorter distances from the grid. Fig. 1.11 shows the motion of an electron which is giving up energy to the grid as just described. It is possible for an electron to oscillate about the grid in this manner for a considerable time, but it is undesirable for two reasons:

1. As the distance that the electron travels decreases, less and less energy is delivered by the electron to the grid.

2. Eventually the electron may even start absorbing energy which would work against the above process. In order to guard against this tendency, the grid is so designed that it will, on the average, capture the electron by the time the phase has shifted sufficiently to cause absorption of energy.

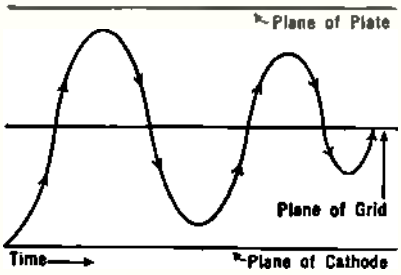


FIG. 1.11. The motion of electrons that sustain oscillations by giving up energy to the grid.

Although two separate and special cases have been employed to describe the action of the B-K oscillator, it can be shown that all electrons leaving the cathode when the alternating voltage on the grid is going positive will absorb energy and either strike the plate and be lost or return and strike the cathode and meet the same fate. On the other hand, all electrons leaving the cathode when the a.c. grid voltage is going negative will lose energy and so vibrate back and forth in the interelectrode space until they are caught by the grid. Since equal numbers of electrons leave the cathode when the grid is going positive as when it is going negative, and since the electrons that gain energy are taken out of action after one cycle, at most, while the other electrons (those that lose energy) remain in the space for several cycles, it can be seen that the alternating voltage source of the grid gains

more energy than it loses. This is important and differs from the first case described where there was only a constant voltage on the grid. Here, recall that there was no resultant gain or loss of energy. It is this gain by the alternating source of the grid that sustains oscillations.

What is the source of the energy that the electron gives up? The large positive voltage on the grid is accelerating the electrons and so imparting energy to them. When the electrons are slowed down they give back part of this energy to the a.c. grid voltage source. Hence the d.c. grid voltage is being changed, via the electron, into the a.c. grid voltage. Does this seem strange? It shouldn't, because this changing of d.c. to a.c. is true in all oscillators. The only difference occurs in the means whereby the changing process takes place, but the end result is the same.

The above point of energy transfer, via the electron, is so important that it deserves a more detailed explanation. This idea will arise again in the other oscillators and cavity resonators to follow, where energy will be obtained from the electrons that are moving through the tube or vibrating in the interelectrode space.

When an electron leaves the cathode it is speeded up by the large positive potential on one of the other elements, i.e., the grid or plate. Thus the electron is given energy which may be in addition to that it might have by virtue of the fact that it was emitted from the cathode with some small initial velocity. Thus the first phase of this process sees the electron receiving energy obtained from a d.c. source. The next phase in this process concerns the utilization of this energy for the continuance of the oscillations in the tuning circuit.

The electron, as it travels toward an element, either the plate or the grid, will cause the electrons in these elements to be displaced. This is due to the natural repellent effect electrons have on each other. If a large number of electrons are traveling toward the element, there will be a correspondingly large displacement of the element electrons. When the traveling electrons in the interelectrode space move away from an element, say the grid, the

electrons that were displaced will return to their previous positions.

But suppose that the traveling electrons pass this element at frequent intervals. Then the element electrons will be forced to move back and forth with the same frequency. This vibratory motion is our oscillation. If a tuning circuit is connected between the elements, we will have an oscillator.

The moving electrons must be kept continually in motion; otherwise the entire process will cease. The large d.c. potentials accomplish this. Hence, we say that the d.c. energy is converted to a.c. energy. The moving electrons lose some of the d.c. energy when they force the element electrons to move from their present positions. The loss in energy is evidenced by a reduction in the speed of the interelectrode electrons.

Carry these ideas over to the preceding explanation of the B-K oscillator. Here, when the electron leaves the cathode, it is accelerated by the large positive d.c. potential on the grid. On approaching the grid, the moving electron causes a displacement of the grid electrons. This effect decreases as the electron moves past the grid and speeds toward the plate. Near the plate, the electron is slowed down and finally stopped. Again the large positive d.c. potential on the grid attracts it and again the electron is accelerated as it travels toward the grid.

Approaching the grid, the moving electron causes another displacement of grid electrons. Then the electron passes through the grid wires and is brought to a stop near the cathode. By constantly passing the grid at equal intervals, this electron and all its companions will cause oscillations to be built up in any tuned circuit connected between the grid and plate. The entire process is dynamic, with the moving electrons losing members, gaining others, but on the whole keeping the net oscillations constant.

The reason for the negative voltage on the plate now becomes obvious. If it were positive, the electrons would flow to it and hence no vibration of electrons about the grid would occur. With the plate slightly negative, the electrons are forced back to the

grid and from this motion, oscillations occur. The frequency of the oscillations is, of course, dependent on how fast the electrons move back and forth. Thus, the B-K oscillator is said to depend upon the electron transit time.

In the oscillators to be described, the transit time magnetron uses a magnetic field to bring about interelectrode electronic vibrations, while the Klystron uses a long drift tube to produce oscillations a little differently.

Another point that can now be cleared up relates to the starting of oscillations in the B-K oscillator. In the previous explanation, an alternating voltage is assumed on the grid. Actually, however, when the circuit is first connected, no such a.c. voltage exists on the grid. How then does this process start? By closing the power switch, any slight disturbance will tend to set this oscillator in action, and the motion of the electrons in the tube due to this sudden disturbance helps build up the a.c. grid voltage. This method of starting an oscillator is just the same as in the Hartley or Colpitts oscillator. Here, too, the closing of the power supply switch gives rise to a disturbance in the circuit and this builds up the oscillations. After a very short time a state of equilibrium is reached and the oscillations become steady.

Efficiency of Operation. The efficiency of this device is low, with typical values of from 1 to 3 per cent, due to the fact that most of the electrons leaving the cathode are caught by the grid, therefore resulting in a low ratio of a.c. grid current to direct grid current. In addition, a tremendous amount of energy must be dissipated at the grid. As the frequency is raised, the elements must be brought closer together and higher voltages must be used. These are two conflicting considerations since higher voltages mean more heat dissipation which would ruin the closely spaced electrodes. Although the practical application of these oscillators is limited where very short wavelengths are concerned (30 cm. and less), they represent the first attempt to cope with the difficulties of transit time in ultra-high frequency generation. There are more efficient devices that are now being used and explanations of these will be given presently. In these devices, again, it

will be found that transit time will be the most important principle that is utilized.

Summary. Starting with the ordinary oscillator it was found that in order to reach higher frequencies it was necessary to decrease both the dimensions of the tube elements and those of the tuning circuits. However, as the resonant circuits became smaller, the Q or, in a sense, the selectivity, decreased to such an extent that other types of tuning apparatus were deemed necessary. The answer was found in Lecher wire systems and concentric cables, usually a quarter or a half wavelength long.

While one problem had been solved, there was still a new phenomenon that reared its head at the very high frequencies, namely, transit time. To understand this phenomenon it was necessary to revise the existing ideas on electron flow within a tube so as to see more clearly just why the transit time of flight of an electron became so important.

The transit-time trouble was found to be extremely hard to eliminate in any tube that was devised to counteract it, so the engineers turned toward putting it to advantage. One of the first attempts was in 1920 when Barkhausen and Kurz took the ordinary triode and reversed the polarity of voltages on the plate and grid electrodes. Oscillations of a very high frequency were generated and at certain points were found to be independent of the Lecher wire tuning system attached to the electrodes. The wavelengths ranged from 43 to 200 cm. and were found to be related to the grid potential by the formula $\lambda^2 E_o = K$, where K is a constant of the circuit.

The mode of oscillation depends on the vibrations of the electron about the grid of the tube. Applying an alternating potential on top of the positive d.c. voltage to the grid, it was found that those electrons that left the cathode when the a.c. voltage was going positive would usually absorb energy and hit the plate, whereas those electrons leaving when the a.c. grid voltage was going negative would be slowed down. These latter are the electrons that give rise to the high-frequency oscillations. However, even these energy-giving electrons cannot be left in the interelec-

trode space indefinitely, and means are usually provided for their removal.

The efficiency of this device is low, with values of about 1 to 3 per cent. Most of the electrons leaving the cathode are captured by the grid wires and thus result in a low ratio of a.c. grid current to direct grid current. Another point to remember is that 50 per cent of the electrons that leave the cathode are energy-absorbing electrons and so are of no use to us at all. These must be quickly removed. Due to these inherent deficiencies of the B-K process, this oscillator is not very widely used. Today it is possible to go to much higher frequencies with greater outputs and better efficiencies.

CHAPTER 2

THE MAGNETRON OSCILLATOR

An explanation of the first really usable ultra-high frequency oscillator

The Second U.H.F. Generator. In order to appreciate the mechanism of the magnetron, which is the next type of ultra-high frequency oscillator to be presented, several fundamental ideas and circuits will be described to show the connection between this new type of generator and those commonly mentioned in elementary radio texts. As every radioman knows, the electron in a tube travels to the plate of a tube when the potential applied to the plate is positive with respect to the cathode. Its path in this case usually is a straight line, the shortest distance between these two elements. The electron will be attracted to the plate whether it is at rest or in motion.

Electrons in Magnetic Fields. Now consider what happens to an electron when it is in motion and only a magnetic field is present. Imagine that an electron is projected into a magnetic field. If the motion of the electron carries it parallel to the lines of flux of the magnetic field, the magnetic field will exert no force whatsoever on the electron. This idea is not new since many people are aware of the fact that, when a wire carrying a current is placed in a magnetic field parallel to the lines of flux, no force is exerted on the wire. In the above case, only one electron is being considered, while for the wire there are many such electrons. The magnetic field will likewise exert no force on any electron that is stationary; the electron will remain at rest. However, if the electron motion carries it into the magnetic region at some angle to the lines of flux, the electron will be acted upon. Referring to Fig. 2.1, suppose that the electron enters a uniform magnetic field

at right angles to the field. Then the force acting on the moving electron will cause it to travel in a circular path. In the case chosen, it left the magnetic field before it could complete the circle but, if the entire space were under the influence of the magnetic field, the path of the electron would have been a true circle. It can be shown by simple considerations that the radius of the path is directly proportional to the speed of the particle. Furthermore, the period and angular velocity are independent of the speed or radius. This means that faster moving particles will travel in larger circles in the same time that a slower moving particle moves around its circle.

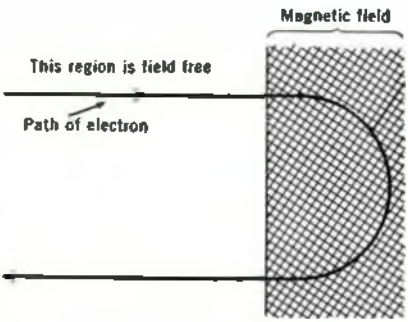


FIG. 2.1. The curved path of an electron when it enters a magnetic field.

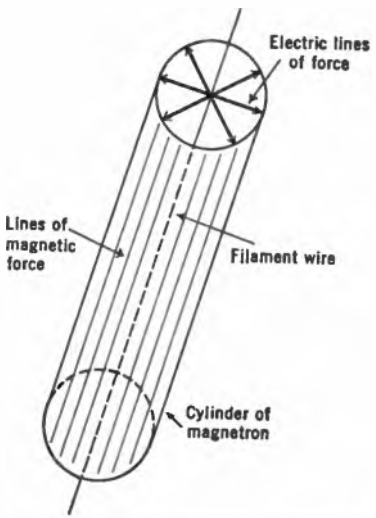


FIG. 2.2. Showing arrangement of magnetic and electric fields in a magnetron.

The magnetron combines both a magnetic and an electric field, each at right angles to the other. This may be accomplished as shown in Fig. 2.2, where the electric field is from the filament to the plate and the magnetic field is parallel to the filament wire and thus is at right angles to the electric field. When an electron is emitted by the filament, its first tendency is to travel straight for the plate. The minute it starts to move, however, the magnetic field also begins to act and so the combined effect of these two fields causes the electron to describe one or more circles while mov-

ing forward. It is now said to be traveling in a helical path which can best be visualized if one imagines the electron as moving in a circle and at the same time also traveling forward. The best mechanical device to describe this type of motion is a screw being put into a wooden board. The motion of the screw is circular, yet it is always moving forward. See Fig. 2.3 for a pictorial representation of this motion.

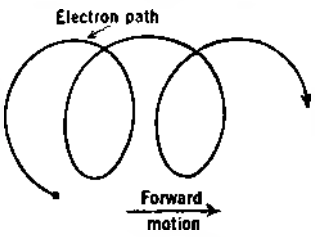


FIG. 2.3. Helical path of an electron under the influence of a magnetic field while moving in the forward direction.

In the magnetron the electron is emitted by the filament and is attracted to the plate. However, due to the presence of the magnetic field, it describes the path pictured in Fig. 2.4. This is the same motion as shown in Fig. 2.3, but it is now applied to the magnetron tube. Fig. 2.4 shows the end view of the cylindrical anode of Fig. 2.2. As the value of the magnetic field increases, the electron describes circular paths with decreasing radii, until finally the electron no longer hits the plate and the plate current stops. It is at this point in the phenomenon

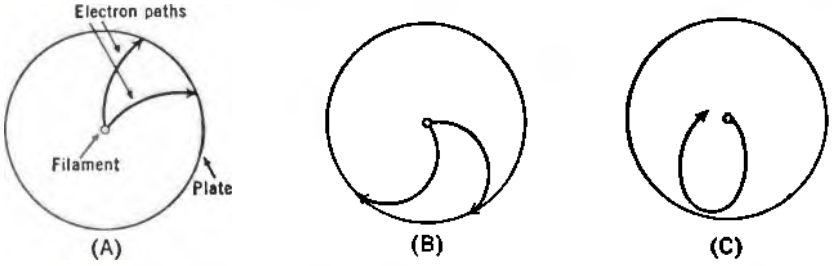


FIG. 2.4. Electron path in a magnetron tube, from cathode to plate. (A) With weak magnetic field causing circular electron path. (B) Increased magnetic field. (C) Strong magnetic field preventing electron from reaching plate.

that the magnetron becomes useful, as will be described in the following paragraphs.

The Two Magnetrons. Having shown what a magnetic field of force can do to an electron, a review of the principles and cir-

cuits of the magnetron ultra-high frequency oscillator may now be given. As first used, it consisted of a cylindrical plate structure and a filament set at the center of this cylinder (see Fig. 2.4), with a uniform magnetic field directed along the filament length. The magnetron most widely in use today differs only in having the plate structure split lengthwise in two sections, see Fig. 2.5, and appropriately called the " split-anode magnetron." The field set up by the magnetic coil is usually constant in value, although alternating magnetic fields have also been tried. The best results, however, are obtained with the steady field.

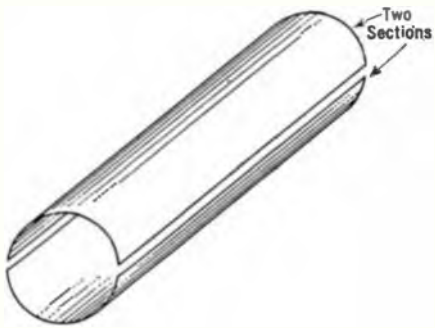


FIG. 2.5. The cylindrical plate is split into two sections for the "split-anode magnetron."

The magnetron, used as an oscillator or generator of high frequencies, can be divided into two types — one is the type known as the negative-resistance magnetron oscillator, and the other is the transit-time oscillator. The negative-resistance magnetron, or dynatron magnetron as it is sometimes called, actually operates for a portion of its characteristic curve as a negative resistance and, hence, as will be shown, will sustain oscillations. Its action is somewhat analogous to the tetrode in which the plate current actually decreases with small increases of its plate voltage. The frequency generated by the negative-resistance magnetron oscillator is less than that of the second type, the transit-time oscillator, in which the frequency of oscillation is determined by the vibration of the electrons between the filament and plate segments.

THE NEGATIVE–RESISTANCE MAGNETRONS

The basic ideas of negative-resistance magnetrons can best be understood if a discussion of negative resistance and its application to oscillators is first described. Everyone knows what a

positive resistance looks like. It is the ordinary type of resistor that can be bought for a few cents in any radio store. This resistor impedes the flow of current in any circuit in which it is used. Power is expended or lost in such a resistance, and the current and voltage are in phase across such a device. If the current increases, so does the voltage. This is what is meant when we say that the current and voltage are in phase.

Now, if a resistance could be found in which the voltage decreased, as the current increased, the two would be 180° out of phase, and this device would exhibit properties opposite to the

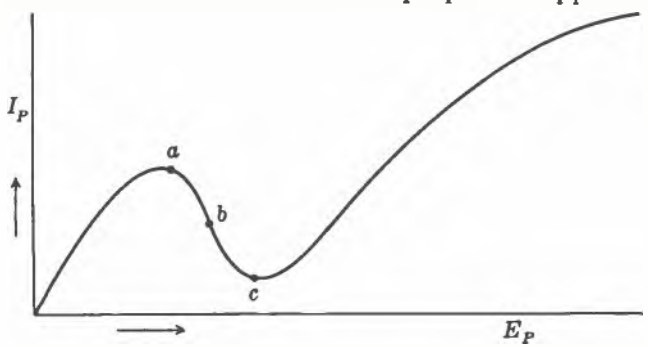


FIG. 2.6. Characteristic curve for a tetrode tube showing portion of curve (a–c) where the plate current decreases for slight increases of E_P.

ordinary resistance and so could be called a negative resistance. It would give power instead of absorbing it and, if combined with an ordinary resistance, it could annul or cancel part or all of this positive resistance, depending upon which were greater. Sounds strange and impossible, and yet such a device is known and can be quite easily obtained by means of a tube.

The tube that can be used to obtain this negative resistance at low frequencies is a screen grid tube such as the 24 tube. A diagram of the characteristic curve for this tube is given in Fig. 2.6. It is only over a portion of this curve that the negative-resistance effect can be obtained, the region lying between points a and c, with b as their midpoint. In order to demonstrate the operation of the tube, Fig. 2.7 shows the hookup needed. An important point to note in this diagram is the fact that the plate voltage is

less than the screen potential, whereas for most usual arrangements the plate voltage is higher. The biasing voltage on the control grid is fixed and does not vary throughout the entire operation.

From the characteristic curve it will be noticed that at point *a* an increase in plate voltage will not result in a higher plate current, but an actual decrease. This can easily be demonstrated by adjusting the plate voltage until the tube is operating at point *a* and then increasing the applied voltage. The plate current drops.

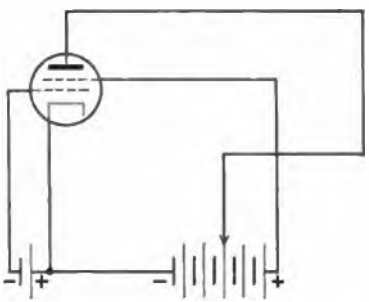


FIG. 2.7. Schematic diagram of circuit hookup to obtain negative-resistance effect in plate circuit.

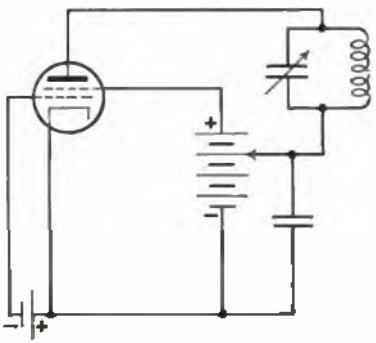


FIG. 2.8. Circuit diagram of a Dynatron oscillator.

Thus, as far as the plate circuit is concerned, a negative-resistance effect has been obtained for the current and voltage between points *a* and *c*. Here these two quantities are 180° out of phase with each other. When operating the tube in this region it is best to keep it at point *b*, since here the greatest stability is obtained.

To apply this principle to an oscillator, it must first be recalled that any oscillating circuit, such as a coil and condenser, has a certain amount of resistance. If some of this resistance could be neutralized, say by a negative resistance, then it would be quite easy to cause currents to oscillate in such a circuit with little damping effect. This principle is applied to the Dynatron oscillator shown in Fig. 2.8. The circuit is the same as that shown in

Fig. 2.7 except that now the plate lead has been opened and the oscillating circuit inserted. The negative resistance in the plate circuit neutralizes some of the positive or actual resistance in the resonant circuit and oscillations take place. The frequency of the oscillations are, as usual, determined by the coil and condenser values. It is this practical application of negative resistance that is used for one of the magnetron oscillators that will now be explained.

Action in the Negative-Resistance Magnetron. The basic idea of the negative-resistance magnetron oscillator was brought out by Habann in 1924, and the arrangement shown in Fig. 2.9

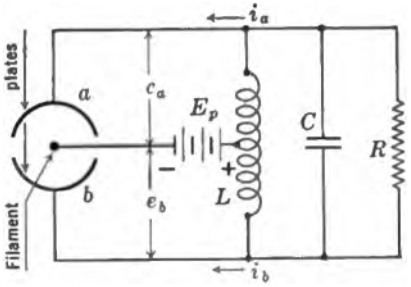


FIG. 2.9. Circuit for a split-anode magnetron oscillator. This diagram closely resembles a push-pull amplifier.

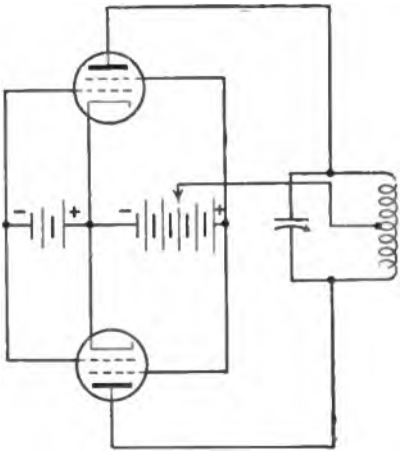


FIG. 2.10. Push-pull Dynatron oscillator. Note similarity between this diagram and Fig. 2.9.

was developed by Manns in 1927. The circuit is really a push-pull oscillator as can be seen by comparison with the low frequency push-pull Dynatron oscillator shown in Fig. 2.10. Each tuned circuit is placed between a plate segment and the filament. Although the tuned circuit shown consists of a coil, condenser, and the inherent resistance of the circuit, actually a Lecher wire system is used since ordinary coils and condensers are nearly useless at the very high frequencies. The coil and condenser arrange-

ment, however, is much easier to understand since radiomen have been brought up on it. The actual circuit with the Lecher wire arrangement is given in Fig. 2.11.

In order to start oscillations, the magnetic field is increased to slightly above the point where the electrons just graze the plate. The first value at which the electrons miss the plate is known as the critical value. The formula usually given for the critical magnetic field is:[2]

$$H_C = \frac{6.72}{R_A}\sqrt{E_p}$$

where R_a is the plate radius in centimeters,
E_p is the plate potential in volts.

Once oscillations have been started, the value of the magnetic field is not critical and may be varied over a considerable range. The oscillation frequency is essentially determined by the natural resonant frequency of the external tuning system. The explanation of why this circuit should oscillate can be found by referring to the volt-ampere, or static, characteristic curves of the split-anode magnetron shown in Fig. 2.12. These curves can be obtained experimentally by increasing the potential of plate A by small amount and, at the same time, decreasing the potential of plate B by a corresponding amount. This is what happens in any push-pull oscillator or amplifier and so it applies here too. In this way conditions similar to actual operations are obtained. When the currents to the plate segments (in this case I_a and I_b) are measured, it is found that more current flows to plate B than to plate A, even though plate B is at a lower potential than plate A. Furthermore, as this difference between plate potentials

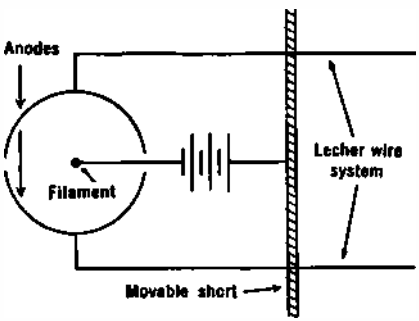


FIG. 2.11. Magnetron oscillator with Lecher wire tuning system instead of actual coils and condensers.

$(E_a - E_b)$ is increased, up to a certain point, the difference between I_a and I_b increases. Since the current increases for the plate of lower voltage a negative-resistance effect is obtained.

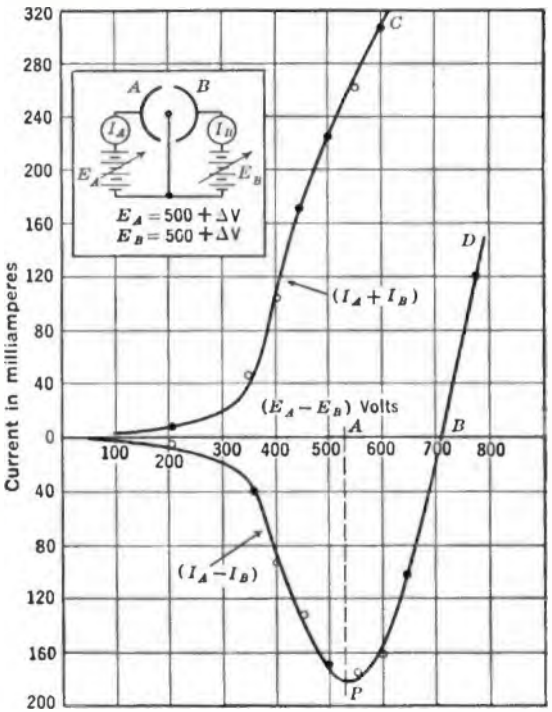


FIG. 2.12. Characteristic curves of a magnetron illustrating how negative resistance is obtained.

This can be plainly seen when the current $I_a - I_b$ is plotted against $E_a - E_b$, and curves like those of Fig. 2.12 are obtained. The portion of the curve shown as OP represents this negative resistance. It is this effect which is responsible for the oscillations, just as in the foregoing description of the Dynatron oscillator. Fig. 2.13 shows the instantaneous potentials on the two plate segments during one complete cycle of oscillation.

The reason for the peculiar behavior of the electrons in going to the plate of lower potential, instead of to the other plate, can

only be found when the fields due to the electric and magnetic forces in the tube are analyzed. It is due mostly to the magnetic field present, since removal of this field would force all the electrons to go to the plate of higher potential.

It must not be thought that the negative resistance is due to one plate alone. While it is true that the characteristic curves shown in Fig. 2.12 state that the voltage on one plate is increasing while on the second plate it is decreasing, remember that this is true only during one half cycle. During the next half of the cycle, the reverse is taking place and so now the other plate is exhibiting this negative-resistance characteristic. This situation is the same as in the push-pull Dynatron.

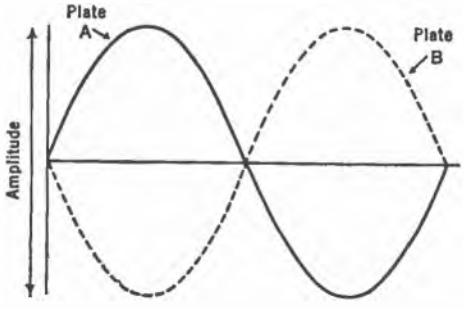


FIG. 2.13. Alternating voltage variation of the potentials on the magnetron plates.

To visualize what the electrons actually do in the tube, refer to diagrams and a photograph published by Kilgore,[3] which are shown in Fig. 2.14. In Fig. 2.14A, the potentials of the two plates are equal and the magnetic field is beyond the critical value. Here the electrons describe circular paths, and none reach either one of the plates. In Fig. 2.14B and C the voltage of one plate has been increased while that on the other has been decreased. Now the electrons describe more or less complicated paths, finally ending up at the plate of lowest potential. This point would not have been immediately obvious and is only due to the strong magnetic field. In Fig. 2.14D a special gas was introduced which did not have any effect on the path of the electrons but, nevertheless, made their paths visible. The photograph clearly substantiates the previous line of reasoning.

Operating Efficiency. In order to obtain good efficiency with this type of oscillator, the electron transit time should be kept to $\frac{1}{10}$ or less of the oscillator frequency. This is also true for ordinary

negative-grid triode oscillators. The importance of these negative-resistance oscillators is due to their large power output and high efficiency even at fairly short wavelengths. It has been

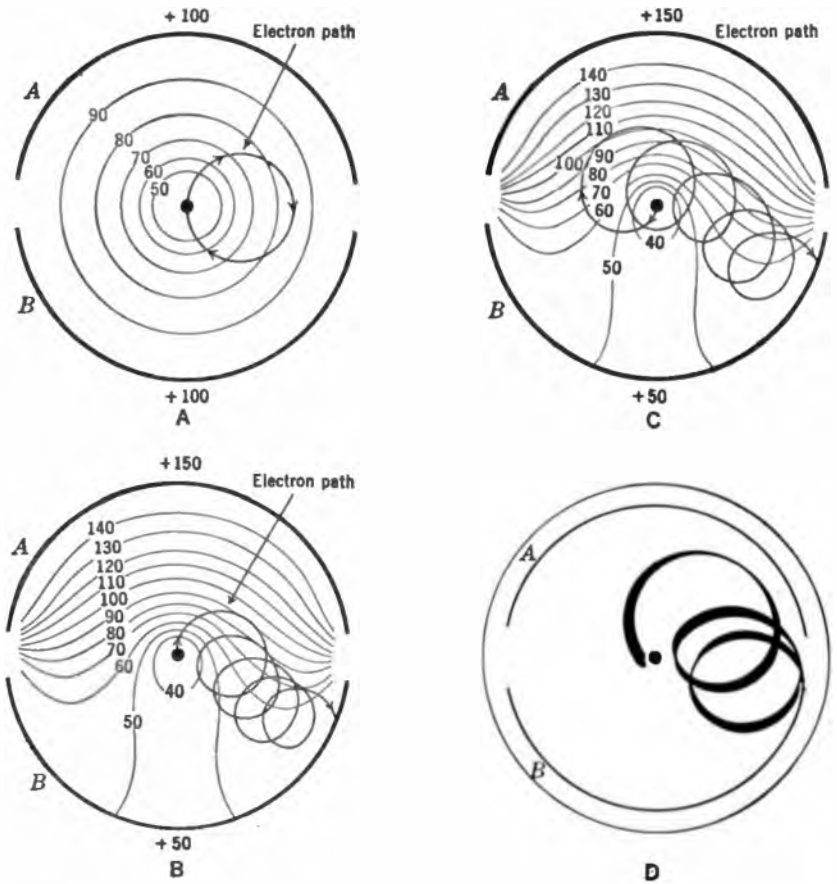


FIG. 2.14. Electron paths in a magnetron of two plates illustrating why electrons reach the plate of lower potential. At (*D*) is a drawing made from an actual photograph showing the same motion.

reported that, by using air-cooled tubes, power outputs of the order of 50 watts or so have been attained at a wavelength of 50 cm. and, when water cooling of the tube was employed, several

hundred watts of power output were obtained with an efficiency of between 40 and 55 per cent. These values represent the best results that have been attained to date.

As the frequency of oscillation desired is raised, the same troubles that one meets with the ordinary negative-grid oscillators are encountered here. The main fault is again due to transit time. This can be overcome in part by either high plate voltage or small plate diameters. Included also, of course, is the much stronger magnetic field that must be used. As plate size is decreased, the allowable plate dissipation must be lowered and, consequently, decreased power output. These limitations again turned the engineers to using this device operated on the transit-time principle or, to put it differently, to utilize transit-time effects to advantage.

TRANSIT-TIME MAGNETRONS

The other class of magnetrons that is used for the very short wavelengths operates at frequencies that are determined by the vibrations of the electrons in the interelectrode space between the filament and plate. Hence the name, transit-time magnetron. In order to reach these high frequencies, Lecher wire tuning systems are used for the resonant circuit. The oscillation of the moving electrons depends upon (1) the distance between the elements, and (2) the strength of the electric and magnetic fields. Any of these can be varied and so change the frequency. However, it has been found most effective to change the magnetic field. This is easily accomplished by either increasing or decreasing the current through the magnetic coils. Usually associated with each piece of equipment is a current-regulated power supply to insure a steady current once its value has been fixed by the operating conditions. The same type of tube can be used here as with the negative-resistance oscillator. And, as has already been explained, plate voltage will also have an effect on the radial motion of the electrons and must similarly be quite carefully set at the correct value. In addition to all the above, it was found that for optimum results the magnetic field should be slightly inclined

to the tube axis. The inclination should never be more than a few degrees.

The mechanism of the transit-time magnetron is in many respects similar to the positive-grid oscillator previously described. However, in order to keep this discussion complete, the theory will be given here separately. In the operation of all magnetrons, the magnetic field is adjusted so that the electrons just graze the plate, as shown in Fig. 2.4. The negative-resistance magnetron uses this effect one way, the transit-time magnetron another way, as will now be shown. Suppose a small alternating voltage is superimposed onto the direct voltage already applied to the plate. This will cause the plate voltage to vary above and below its normal or d.c. value. Any electron leaving the cathode when the alternating voltage is going positive will be given a greater acceleration (subject to a larger positive force) than if the a.c. voltage had been absent. Hence, this electron will probably hit the plate with excess energy absorbed from the alternating voltage source. Since this electron absorbed energy it is imperative that it be removed as soon as possible. The object, as stated before, is for the alternating source of power to absorb energy, not lose it. This is the only way to sustain oscillations.

An electron which leaves the cathode one half cycle later, when the a.c. voltage causes the d.c. voltage to decrease, will not approach as close to the plate as when the plate voltage was constant. Consequently, energy must have been given up by the electron to the source of alternating voltage because, without the superimposed a.c. fluctuations, the electron ordinarily would have had enough energy at this point to keep on going. On its return trip the electron will not quite reach the cathode because now the plate is going more positive and the attraction for the electron is greater than it ordinarily would be. The electron can continue to circle around with decreasing amplitude and to deliver energy to the source of alternating voltage, until it comes to rest at a point between the anode and the cathode, or is removed from the interelectrode space because of its sidewise motion. Keep in mind throughout this entire discussion that the reason the

electron moves to the plate and then swings back to the cathode is due entirely to the fact that there is a magnetic field present which tends to force the electron to move in a circular path. Without the magnetic field none of this action would be possible. Fig. 2.15A shows the motion of an electron that absorbs energy from the a.c. source. Fig. 2.15B shows the path of an electron giving up energy, causing oscillation.

Returning now to the mechanism of the tube, it can be seen that all electrons leaving the cathode, when the plate voltage is increasing, will absorb energy and probably be captured by the plate, whereas those electrons that are emitted by the cathode when the plate voltage is decreasing will not reach the plate, but rather will oscillate in the manner previously described. As the same number of electrons are emitted when the alternating plate voltage is going positive as when it is going in the opposite direction, and as the energy-absorbing electrons are removed as soon as they strike the plate, while the energy-losing electrons oscillate for several cycles, the source of alternating voltage gains energy. This condition is precisely the same as in the positive grid of the Barkhausen-Kurz generator.

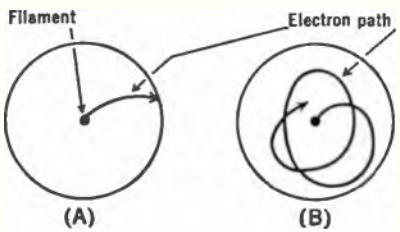


FIG. 2.15. The path of two different electrons in a magnetron when (A) the electron absorbs energy and (B) when it gives up energy.

At first thought, it might be desirable to keep the energy-losing electrons in the interelectrode space as long as possible. However, it has been found that this is not effective. The reason is due to the changing phase relation of the induced voltage. With each successive cycle of oscillation, the electron has less and less energy available and so the distance through which it moves decreases and the transit time becomes less. This condition is equivalent to an advance in phase of the electron oscillation relative to the alternating plate voltage. If this shift continues, not only will less and less energy be given up by the

electron, but it may even start absorbing energy. It has been found that the magnetic field must be slightly tilted (not more than a few degrees) with respect to the tube axis in order to obtain the best results. The tilting action imparts a velocity to the electrons parallel to the filament and, after several cycles in the interelectrode space, the electrons strike the plate and are removed. The angle is quite critical and must be changed with plate voltage. The electrons should be allowed to remain in the interelectrode space until all the energy that can be absorbed has been absorbed. Then they should be quickly removed; hence, the tilting action. One unfavorable result of this tilting is the heating of the glass envelope of the tube itself which, if allowed to continue too long, will melt the glass, with the eventual destruction of the seal.

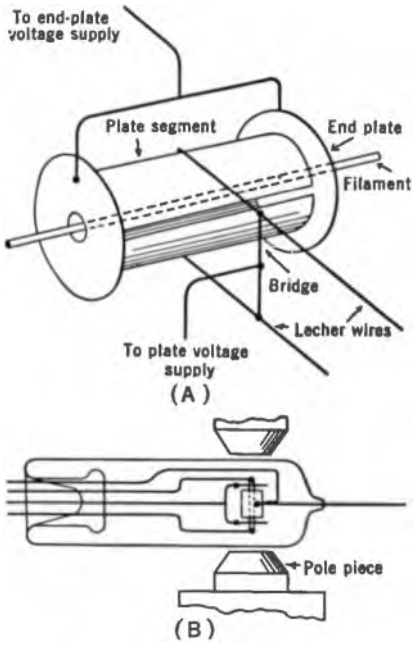


FIG. 2.16. (A) Physical construction of an end plate magnetron. (B) The arrangement of the pole pieces in applying the necessary magnetic field.

Linder[4] has advanced a solution of the tilting process that is used most widely with magnetrons. He found that by putting in end plates, shown in Fig. 2.16, and applying a small positive voltage to them, the tube no longer has to be tilted in order to remove the electrons. It has also been found that the efficiency has increased due to these end plates.

The theory of the mechanism of transit-time magnetrons holds essentially whether the anode is a one-piece cylindrical plate, a two-segment plate, or even a four-segment plate. It is true that there are slight differences in the electron motion in these various

types of magnetrons, but these need not be of concern because, as mentioned, the explanation is essentially the same.

Difficulties Encountered in Practical Operations. For optimum output, the transit-time magnetron requires a smaller filament emission than the negative-resistance tube and is far more critical to changes in this emission, a situation probably due to the effect these changes have on the space charge in the tube itself. It is customary to interpose a filament-current stabilizer between the source of power and the filament. Another source of trouble is in keeping the magnetic field strength constant, since the formula previously given for wavelength is directly concerned with this factor. The difficulty may be avoided by using electromagnets at high voltage and low current. The schematic diagram for a power supply given in the chapter on the Klystron tube may also be used here.

The useful frequency range of any one magnetron tube is approximately 2 to 1. In order to cover a wide range of frequencies many tubes of varying sizes must be used. It was found that the optimum anode diameter should be $\frac{1}{20}$ the wavelength desired. Of interest to engineers and circuit designers is the fact that the transit-time magnetron can be operated with a much lower load impedance than the negative-resistance magnetron and hence can be operated with tuning elements of a lower Q. One explanation of this observation states that it is due to the fact that a small amount of potential on the plate segments can cause large circulating currents which in the ordinary tube would require a much greater voltage. Oscillations have been known to exist in tuning circuits with an impedance as low as 100 ohms, which is very helpful in attaining the high frequencies.

In searching the literature on the magnetron tube, the statement is often found that a smaller magnetic field is needed in the transit-time tube than in the negative-resistance version. That this is true can be seen from the following line of reasoning. For the negative-resistance type magnetron the efficiency begins to fall off when the transit time of the electrons is greater than about $\frac{1}{10}$ of the period of oscillation. The magnetic field must therefore

be correspondingly greater in order for the electron to complete the cycle that much sooner. In the transit-time oscillator, the allowable electron transit time is ½ of the oscillation period. Hence, not so great a magnetic field of force is required, the difference being in the ratio of about 5 to 1. This results in a considerable saving, since the difficulty of constructing a suitable electromagnet increases rapidly as the number of required uniform flux lines increases.

The subject of modulation of the magnetron has not been fully settled to the satisfaction of all concerned. It might be stated that the method probably adopted will tend more toward frequency modulation than amplitude modulation.

Before closing the chapter on magnetrons, one more comment should be made regarding the effect known as " filament bombardment." This effect has been noted by many experimenters and contributes much toward limiting the usefulness of this type of oscillator. When attempts are made to obtain large amounts of power, it is found that the filament of the tube begins to heat up and eventually disintegrates. The reason has been attributed to the bombardment of the filament by electrons or ions containing considerable amounts of excess energy. One means used to overcome this condition lies in shielding the filament to some extent, while another method proposed is to attempt a filament-stabilization process. As stated, this defect shows up when the power requirements are raised, but it is believed that newer types of magnetrons have already eliminated this difficulty.

Summary. In summarizing the contents of this chapter it should be pointed out that, in the negative-resistance magnetron oscillator, the action comes from the negative slope of the static-characteristic curve which gives rise to the necessary negative resistance. Once the electrons leave the filament they travel onto one of the split-anode segments and do not oscillate in the inter-electrode space. Oscillations do not result from the travel time of the electrons, but from the fact that the electrons have a tendency to hit the plate of lower potential, a condition brought about by the presence of a strong magnetic field.

In the transit-time magnetron, on the other hand, the oscillations are dependent on the time of flight of the electrons, and the Lecher wire tuning system is resonated at that frequency for large output power. Here the electrons that are important remain in the interelectrode space for a relatively long time, circling around between the filament and plate. The path is usually complicated and any diagrams are at best only approximations. Fig. 2.15 is such a diagram. It is difficult to say where the magnetron oscillator ceases the negative-resistance operating principle and switches over to the transit-time idea. The transition is not entirely distinct and the only generalization that can be made is to say that for very high frequencies the transit-time oscillator should be used, whereas for longer wavelengths, the negative-resistance magnetron is more efficient. The circuits remain the same in either case, the only change occurring in the operating conditions.

CHAPTER 3

THE KLYSTRON

Wherein transit time is used to our advantage and a high efficiency oscillator is evolved

The radio field in the 1930's, so far as the ultra-high frequencies were concerned, contained only one good ultra-high oscillator, the magnetron, described in the preceding chapter. The efficiency of this device was fair, but it still left a great deal to be desired. The rather low efficiency was due to the fact that not all electrons operating in the tube gave up energy to the alternating voltage source. There was a large number of electrons that absorbed energy and so had to be removed as soon as possible. It is to be remembered that oscillations can only occur when the power delivered to the oscillating circuit is greater than the power dissipated. And, finally, it will be recalled that only 50 per cent of the electrons leaving the cathode give up energy to the resonant circuit. The other 50 per cent represent a loss in efficiency since these electrons absorb energy and have to be removed quickly.

With these limitations in mind, tubes based on a new principle, called velocity modulation, were developed. Of these, there is one that was developed by the Varian Brothers in 1939 that has been quite widely used. Either it or modifications of it will probably be the most popular of this group and hence will be described. The transit-time principle is very much in evidence in this tube, but the fundamental process of absorbing energy from the electrons is entirely new. The name given to this new tube is the Klystron.

Cavity Resonators. Before continuing with a description of the Klystron tube, it is necessary to digress for a moment and consider the tuned circuit used with this device. It is unlike any-

thing that has been encountered before and was developed principally to meet the peculiar conditions inherent in the Klystron. An entire chapter will be devoted to this tuning device, but an introduction is necessary here to cover the Klystron tube completely. In previous tubes described, a Lecher wire system

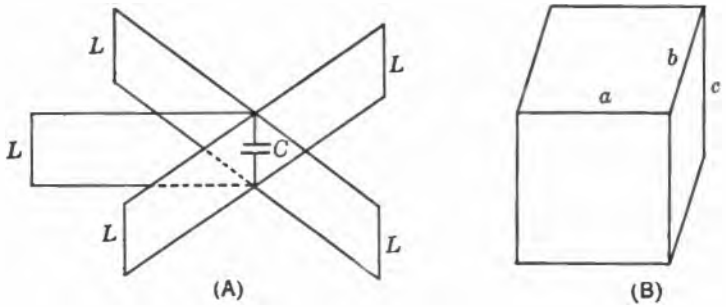


FIG. 3.1. A shows how we can obtain a cavity resonator by adding Lecher wire systems in parallel. B shows a rectangular resonator; all the sides are solid conductors – inside is hollow.

proved very successful as a tuning unit. Also, if the 8012 tube was placed at the center of the Lecher wire system, greater efficiency and a higher frequency could be obtained. If this idea is carried to the limit, then by adding more and more Lecher wire systems in parallel, a hollow container such as shown in Fig. 3.1 would be evolved. This container has a resonant frequency determined by its dimensions. Furthermore, the electric and magnetic fields are confined to the inside of this chamber and so prevent loss caused by energy radiation. Due to this shielding, body capacitance has no effect on the resonant frequency, a difficulty that amateurs still clearly remember about the old 5-meter sets. Several types of cavity resonator shapes are shown in Fig. 3.2. The one great disadvantage of these tuning elements is their lack of flexibility when it comes to changing frequency. The frequency may be varied slightly by distorting the shape of the resonator, but the amount by which the frequency can be altered in this way is very small.

In order to excite the resonators, an electron beam is sent

through holes provided at the center of the resonator, as shown in Fig. 3.2A. The center is usually constricted so as to provide a very short path for the electrons inside the resonator. In order to remove energy from the resonator, a small antenna is inserted at a point where the electric field is greatest. For this purpose a small opening is made on one side of the resonator. It is also possible to insert a small loop to absorb energy, but this loop should be placed at a point of high magnetic intensity. In the

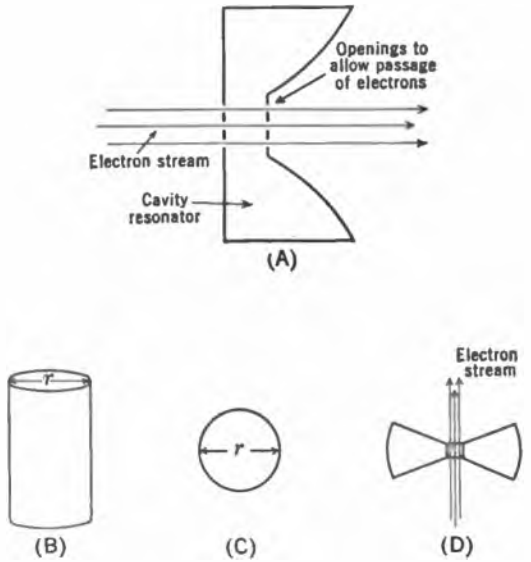
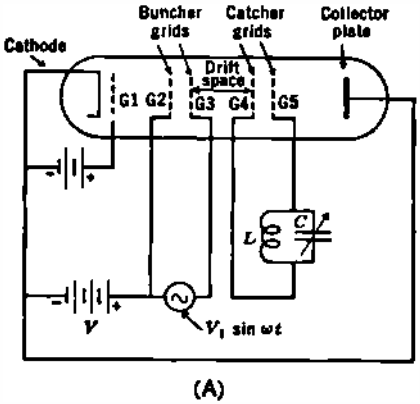


FIG. 3.2. Various shapes of cavity resonators — cross-sectional views.

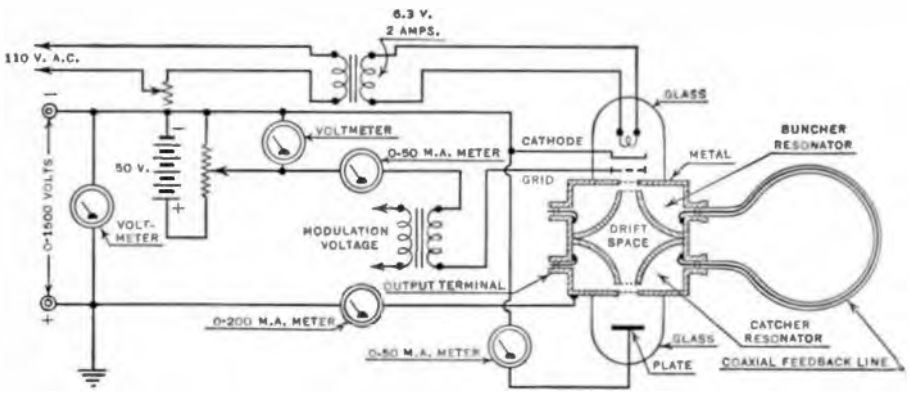
Klystron it will be found that these resonators are made part of the tube itself, one way by which good efficiency can be obtained. This arrangement, then, is the tuning unit used with the Klystron.

Action in a Klystron. In order to understand how a Klystron tube operates, study the set-up shown in Fig. 3.3A. A stream of electrons produced by the cathode and focused into a sharply defined beam with the aid of the first grid G_1 is accelerated toward G_2 and G_3 by the high positive voltage V (usually well over 1000 volts). The grids G_2 and G_3 are placed very close together so

that the electrons spend only a relatively short time between them. Superimposed on the large d.c. voltage V is a small alternating voltage $v_1 \sin \omega t$. If an electron reaches the grids



(A)



(B)

FIG. 3.3. (A) The relative placement of the electrodes to produce velocity modulation. (B) A complete Klystron oscillator circuit showing the form of a modern Klystron tube and the attached components.

when the alternating voltage between them is zero, it will pass on unaffected. However, any electron reaching the grids when the alternating voltage $v_1 \sin \omega t$ is going positive will be accelerated

faster than if this a.c. voltage were not present. Similarly, when the alternating voltage is going negative, the electron will be subjected to less positive voltage V than usual, and so its velocity will be less. In this way, the electrons will tend to have different velocities, depending on the time in the alternating cycle at which they passed through the two grids. The grids G_2 and G_3 are technically referred to as the buncher grids. The reason for the name will become obvious when, as we will see later on, the action of these grids is to cause the electrons to form in bunches along their path down the tube. Since the alternating voltage $v_1 \sin \omega t$ is very small in comparison with the accelerating voltage V, the electronic velocities are not greatly changed as the electrons pass through the buncher. It can be said that just as many electrons enter the buncher grids as leave and that, since the electrons spend only a very short time between the grids, there will be no resultant current induced in the grids by these electrons. Thus the transit-time effect of the electrons on the grids is negligible, a fact that is pointed out here in order to show that the Klystron tube does not have the weakness of the conventional tubes mentioned in Chapter 1.

So far the electron stream has been but slightly affected, and the effects are not sharp or clear enough to be of much use. However, with enough time, this velocity-modulated beam can be converted into a density-modulated beam, in which form energy can be extracted from it. There are several conversion methods, but the most effective is the drift-tube process. By this method, the electrons are allowed to move down a relatively long tube in which there are no electric or magnetic fields. The electrons that were accelerated in going through the buncher grids will eventually catch up with those electrons that are ahead and were previously decelerated. Thus, bunches of electrons will form as the electrons continue on their path in the long tube. At the center of these bunches will be found, of course, those electrons that passed G_2 and G_3 with their velocities unchanged. Another way of visualizing this bunching process is to consider three electrons: the first, which passed through the bunching grids when the alternating

voltage $v_1 \sin \omega t$ was still negative (around 340° of its cycle); the second that passed through when $v_1 \sin \omega t$ was zero; and the third when $v_1 \sin \omega t$ was going positive. The first electron was slowed down, the second passed through with its velocity unchanged, and the third electron had its velocity increased. At some point in the drift tube these three electrons will meet to form a bunch. So the electron stream that started out velocity-modulated ended up as a density-modulated beam. A drift tube, however, should not be too long. If the electrons keep on traveling, the sharp bunches will soon disappear or tend to merge, in which case the beam is no longer usable. The dispersing of the beam is due mostly to the repelling action of the negatively charged electrons upon each other.

Two other grids, G_4 and G_5, placed at a sufficient distance from G_2 and G_3 and connected to a tuned resonant circuit (in the Klystron, a cavity resonator), will be set into oscillation if the bunches of the electrons pass through at intervals equal to the natural period of the circuit. These grids are called the catcher grids since it is their express purpose to intercept and absorb the alternating energy that was imparted to the electron stream as it passed through the buncher electrodes. Energy will be delivered to the catcher grids only if the electron bunches pass through when the field between these grids is a retarding field. The energy transfer should occur in one direction, i.e., from the electron beam to the catcher grids, or, when the electron beam is decelerated.

Recall again the action of an electron giving energy to a resonant circuit and reread the explanation given in Chapter 1 and summarized here. Since an electron has a certain mass, energy must be expended on it to make it move, i.e., give it a velocity. The energy of the electron in motion is known as kinetic energy. If any attempt (by the tube elements) is made to slow the electron, the device used for the slowing action will receive some of this electron's kinetic energy. The energy that is transferred represents the difference between that which the electron had when moving at a fast speed over that which it had when moving at the

slowed down rate. Hence the reason for the retarding field at the catcher grids.

There are several requirements which the catcher grids must satisfy in order to work most efficiently:

1. The grids should not be placed too far from grids G_2 and G_3. There is a certain spacing that gives the best results in bunching the electrons.

2. The resonant circuit of the catcher must be tuned to the correct frequency.

3. The strength and phase of the oscillations in these grids should be such as to absorb from the electron bunches as much of the high-frequency energy as possible. After the beam has given up its high-frequency power to the catcher, it expends its direct current power on a collecting electrode which then returns these electrons to the cathode, thus completing the circuit. Throughout the entire process the total current remains constant and is referred to as the conduction current.

Summary of Process. The action in a Klystron may be summarized as follows:

1. The electron stream must first be velocity-modulated.

2. The velocity-modulated beam must then be converted into a density-modulated beam by means of the field-free drift tube.

3. Finally, the high-frequency power must be absorbed from this beam by means of a suitable circuit. In actual Klystrons, the " circuit " consists of a cavity resonator because of its great efficiency and high Q. However, even an ordinary Lecher wire system would work.

Thus the Klystron tube used all the electrons to transfer power from the buncher to the catcher grids. The first set of grids changed the velocity of the electrons and the drift space allowed them to form into bunches. The bunches of electrons, as they passed the catcher grids, delivered energy to these grids and so gave rise to oscillating currents in the cavity resonator attached to them. If the bunches of electrons passed the catcher grids at the right time, the oscillations would be strong, and large electric and magnetic fields could be formed inside the resonator. By

means of a loop inserted through a small hole on one side of the resonator, energy could be transferred to various other points. Although not specifically referred to, it has been assumed in our discussion that the resonant frequency of the cavity resonator is such that the groups of electrons passing it will give rise to oscillating currents. The entire situation may very easily be compared to a class C amplifier where the current flows in small sharp pulses. Each one of these sharp pulses keeps the currents oscillating in the plate tank circuit, but these oscillations could not occur unless the frequency of the sharp current pulses were properly related to the frequency of the tuned circuit.

Applegate Diagram. A diagram that has been drawn up in connection with the bunching action in the Klystron is shown in Fig. 3.4. This is known as an Applegate diagram and shows, by means of lines, the paths of the various electrons in their travel from buncher to catcher grids. The lines all start at the top of the diagram, at the buncher grids, and all the electrons are assumed to enter these grids with the same velocity. In passing through the buncher grids the electrons are acted upon by the alternating voltage shown at the top of diagram. The horizontal axis, the one that goes across the page, refers of course to time. Starting at the extreme left-hand side of this diagram, it can be seen that the first electrons passing through the buncher grids, when the a.c. grid voltage is zero, will pass on unaffected. The flight of these electrons can be pictured by following the first few lines as they proceed through the drift space until they reach the catcher grids. The lines are tilted to the right because the electrons arrive at the catcher grids at a later time. By following all the lines, it can be seen that most of the electrons leaving, when the a.c. voltage is passing from negative to positive, will tend to form bunches by the time they reach the catcher. From the diagram it can also be observed that there is only one bunch of electrons per cycle and so the frequency of the wave is not changed.

It is interesting to note from this diagram that, for best results, the catcher grids must be placed at a certain distance from the bunching grids. This distance depends upon the large d.c. acceler-

ating voltage and upon the a.c. voltage on grids G_2 and G_3. Usually the distance of the two sets of grids from each other is fixed and so the voltages must be varied. This factor will soon be considered in much greater detail. It is mentioned here in order to point out that the voltages at which this tube is operated are quite critical and must be obtained from regulated power supplies.

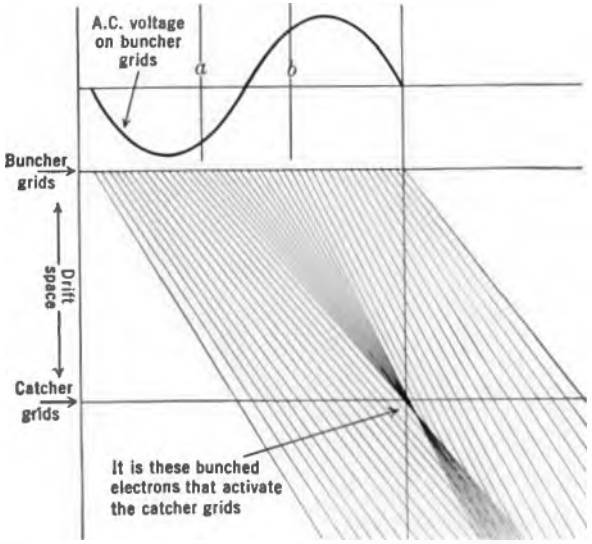


FIG. 3.4. Applegate diagram of Klystron showing (by means of lines) the paths of the electrons from bunches to catcher grids.

Complete Klystron Oscillator. In order to convert the previously explained Klystron tube into an oscillator, the alternating voltage generator connected between grids G_2 and G_3 of Fig. 3.3A is removed. In its place, the grids are connected to a cavity resonator having the same frequency (resonant) as that of the catcher resonator, as in Fig. 3.3B. Then, by means of a coaxial transmission line, energy is fed back from the output to the input circuits and the unit is made to oscillate. Fig. 3.3 shows a typical set-up, complete with all the voltage values and apparatus necessary. Inspection of this diagram shows that high voltages are used throughout. This is necessary in order to have the elec-

trons travel the relatively long distance from the cathode to the plate in a short time and also to impart sufficient energy to the electrons for large output power. Another precaution that must be observed is the filament voltage. The full 6.3 volts should not

Courtesy Sperry Gyroscope Co., Inc.

FIG. 3.5. Sperry Type 2K30/410-R Klystron.

be suddenly impressed on the filament; rather, the process should be extended over a period of from 4 to 5 minutes. The same thing should be done when the oscillator is turned off.

Phase Relations in Klystron Oscillator. As is usual in all oscillators, energy must be fed back from the plate to the grid in correct phase. Due to the inherent characteristics of the tube itself, there is a considerable phase shift, arising mostly from the long time that the electron spends in the drift tube space. Signals that act at the buncher grids show up some time later at the catcher grids. In amplifiers this fact is of little significance but, in designing oscillators, phase shift is very important and must be

taken into account. In order to see where the various phase shifts arise, it might be instructive to study the action from the input grids to the output circuit and then through the coaxial cable back to the buncher grids again.

The first phase shift that is encountered in going from the buncher grids to the catcher grids is that due to the drift tube. The time that an electron will spend in traveling this relatively long distance will depend upon the following factors:

1. The distance between the centers of the buncher and catcher grids.

2. The speed with which the electron enters the first set of grids.

3. Finally, the change in velocity brought about by the a.c. voltage on the input grids, G_2 and G_3.

The time that the electron spends in going this distance will usually be equal to several cycles of the a.c. voltage on the buncher grids. This is the reason that the drift tube space is said to be relatively long. Actually it may be quite small but, in comparison to the time it takes the a.c. voltage to complete one cycle, it is long. These relative values must be constantly kept in mind, otherwise erroneous ideas may be formed. Mathematically it can be shown that the actual phase shift introduced by the drift space is proportional to each of the following quantities:

$$f, \quad D, \quad V$$

where f is the frequency of input wave,
D is the length of drift space,
V is the direct accelerating voltage.

As an example, if the frequency f is 1000 megacycles, V is 1600 volts and D is equal to 3.6 cm., it can be shown that the phase shift will be 3π radians. Since π radians, from elementary trigonometry, are equal to 180°, then 3π radians equal 540°.

The next large phase shift occurs in the coaxial cable connecting the two cavity resonators and can be denoted by θ_2. It might be assumed that this is the entire phase shift, but this is not the

case. There is an additional phase shift of 90° due to the relationship of the buncher to the catcher. It has been mentioned that the electrons would tend to reach the catcher in groups. From the Applegate diagram it can be seen that these bunches are made up of electrons that passed through the buncher when it was going from negative to positive, say that section of the a.c. wave between lines *a* and *b* (Fig. 3.4). Now these two lines are equally placed with respect to the point where the sine waves cut the zero line. Thus it can be said that the center of the bunches arriving at the catcher consist of those electrons that passed through the buncher grids when the a.c. voltage was zero. But when these bunches are passing the catcher, the catcher grids must have maximum retarding voltage in order to get the most energy out of passing electrons. Thus maximum voltage at the catcher is associated with zero a.c. voltage at the buncher. This leads to a 90° phase shift because there is 90° difference between zero and maximum on a sine wave.

Now, all these phase shifts must equal $2\pi n$, where n is any whole number. With this condition, the energy fed back will be in correct phase to aid the oscillations and hence keep the process in operation. This is called positive feedback, in contrast to negative feedback. For the latter, the energy is returned 180° out of phase. Of all the above factors that determine the total phase shift, there is only one that is easily variable — the voltage. Thus, to get this unit to oscillate at any given frequency, all that is needed is to determine the correct voltage. This may either be computed mathematically or obtained experimentally by trying various voltages until the correct one is found. The latter may be accomplished quite readily by the simple expedient of allowing the d.c. voltages to vary slightly about some mean value. In practice this may be done in most power supplies by temporarily removing one or more filter condensers. The a.c. ripple will usually provide enough variation so that the operating point is quickly reached. The above idea that the voltage determines the correct operating point for an oscillator is something that has not been encountered before to the same extent.

Regulated Power Supply. In order to maintain a constant voltage for application to the Klystron tube (once its value has been determined), a regulated power supply of the type shown in Fig. 3.6 is widely used. For those readers who have never seen a unit of this type, the following brief explanation is given. This circuit is essentially an inverse feedback amplifier or, as it is sometimes referred to, a degenerative amplifier. The tube T_3 and resistor R_5 form this amplifier. It is the function of these com-

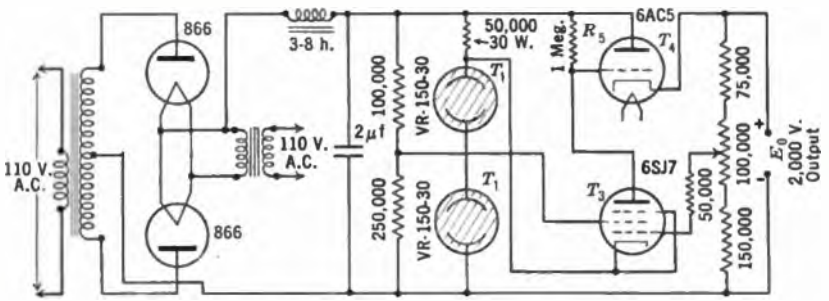


FIG. 3.6. High voltage power supply, compensated for changes of input voltage and load current by using a two-stage inverse-feedback circuit.

ponents to amplify the voltage applied to the grid of T_3. Assume that the circuit is stable and then, for one reason or another, there is a tendency for the output voltage E_0 to increase. The increase of E_0 will decrease the negative voltage applied to the control grid of T_3, and the plate current will increase. Since this plate current must flow through resistor R_5, a greatly amplified voltage will appear across this resistor and drive the grid of T_4 more negative. The load current will thus decrease and E_0 will be brought back to its normal value. When E_0 decreases, the action is the reverse of that just described, and E_0 is automatically brought back to its proper value.

The function of the voltage regulating tube T_1 is to keep the cathode of T_3 at a constant potential with reference to the negative side of the line. T_1 may be one or more tubes of the VR-150-30 variety depending upon how much voltage is needed for the bias. The 100,000-ohm and the 250,000-ohm resistors in

series are used to provide the correct screen potential for T_3. The 50,000-ohm, 30-watt resistor in series with the two voltage regulating tubes is employed to limit the current through these tubes. This type of stabilizing unit is used extensively in installations where good regulation is desired. Although the current carrying capacity is rather small, appreciable power can be obtained by using several tubes in place of the one tube T_4.

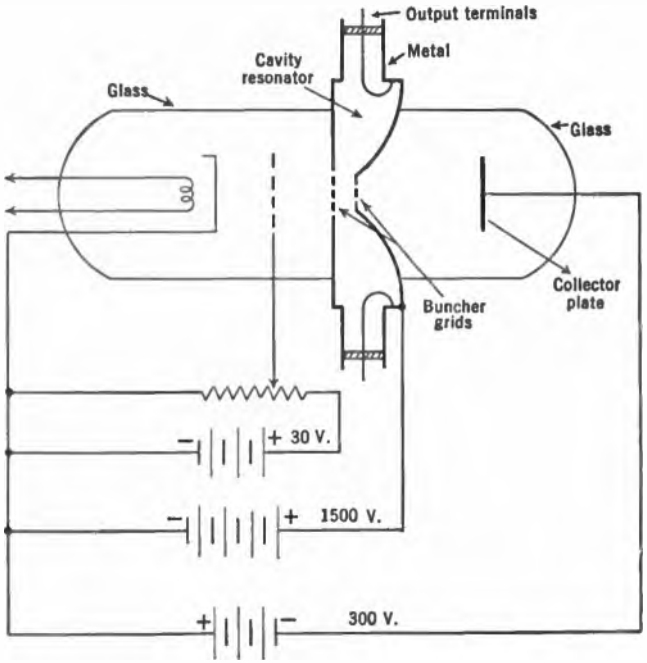


FIG. 3.7. A reflex Klystron oscillator with single resonator as buncher and catcher.

The Reflex Klystron Oscillator. Another form of the Klystron oscillator is shown in Fig. 3.7. Here, a single cavity resonator is used to perform both the functions of a catcher and of a buncher. The electrons are still velocity-modulated by the buncher grids and then travel toward the plate. Since the plate is at a negative potential, the electrons are slowed down and finally are sent back in the direction from which they came. If, now, they again pass the cavity resonator at the right time, they will deliver energy

to it and oscillations will take place. This unit is called a reflex Klystron oscillator.

Modulation. In order to modulate the Klystron type of oscillator, a modulating voltage may be inserted onto the grid, G_1, located between the cathode and the buncher. However, due to the variation of frequency with accelerating voltage, it is difficult to obtain linear amplitude modulation. Only very small percentages of amplitude modulation have proved successful in tests carried out in the preparation of this book. Possibly some sort of frequency modulation might prove more successful.

The Klystron as an Amplifier. The one great advantage of the Klystron tube over most other tubes used at the high frequencies is its ability to function as an amplifier. By driving the buncher from some external source of power of correct frequency, and by having the electron beam great enough to deliver more energy to the catcher than is needed to drive the buncher, it is possible to use the Klystron tube as an amplifier. For power amplification an efficiency of 15 per cent can be obtained and for voltage amplification a gain of 20 at 1000 megacycles has been observed. Webster has shown that the theoretical efficiency of the Klystron is 58 per cent. Up to the present time, however, there have been no published applications using the Klystron as an amplifier.

Design Suggestions for Future Klystron Oscillators. The discussion of ultra-high frequency generators is concluded with the end of this chapter. The principles upon which the magnetron and the Klystron function should be used as groundwork for many future generators in the range of ultra-high frequencies. No doubt many other applications and modifications will evolve. With a good understanding of the basic principles involved, little trouble should be experienced in making the transition from the old to the new. It might even be advisable to stop and see just what might be done to these oscillators to improve them. The weak points of the Klystron will be used as illustrations, since the faults of the magnetron have been previously pointed out in Chapter 2.

First is the matter of stability. An oscillator, to be of any use, must be able to stay put, so to speak, at the point where it is adjusted. If the slightest voltage variation upsets this equilibrium, ways must be found to eliminate this fault. In the case of the Klystron, it has just been shown that the voltage regulation is critical. Oscillations will occur only at certain definite voltages and at no other values. It would help, if some future modification allowed for some voltage variation, say 5 to 10 per cent. There would then be a much better chance of "staying put," and this would give a unit that is much more flexible.

Secondly, an oscillator is desired that is capable of operating over as wide a band of frequencies as possible. As it now stands, the frequency limits of the Klystron are very narrow; narrow, that is, in the sense that once the cavity resonators are set, the oscillator is more or less confined to that one frequency. It is true that by means of mechanical pressure, slight physical deformations can be made which result in a very small percentage frequency change. But such changes are small. Furthermore, it must be remembered that each time the frequency changes, the correct voltage that will cause oscillation must again be found. The whole process is tedious and quite lengthy. Here again there is need for modifications that will make the above procedure smooth and quick. No doubt these changes will be forthcoming in the future.

The last point that could be corrected or improved would involve the compactness of the apparatus itself. The entire equipment requires too many parts; some might be done away with or, failing that, their size might be reduced. U.h.f. radio will probably be more useful if made more portable. Admittedly this is a minor point, but improvement can be made.

Summary. The Klystron tube operates on a principle entirely different from any of the preceding tubes used for producing ultra-high frequency oscillations. The principle of its operation has sometimes been named velocity modulation, but this is really only part of the process. The remainder is concerned with changing the velocity-modulated beam into a density-modulated beam

and then of drawing out or absorbing the alternating energy inherent in this beam of electrons. The action all starts when the electrons are subjected to an a.c. voltage as they pass through the first pair of grids (called the buncher grids). At this point, one of three things can happen to the electrons. They may either be slowed down, speeded up, or unaffected, depending upon the time in the a.c. cycle when they passed through the grids. The electrons that are speeded up tend to catch up with the electrons that were slowed down and these two groups form bunches. All this occurs in the relatively long drift tube space.

The bunches, as they pass through the catcher grids, set up oscillations, and energy is transferred from the electron beam to the catcher cavity resonator. If the resonant frequency of the cavity resonator is equal to the frequency at which the bunches arrive at its grids, strong oscillations will be produced in the output circuit. Feeding some of this energy back to the input will cause the Klystron to act as an oscillator capable of generating wavelengths as low as 1 cm. The efficiency of this device may be made as high as 25 to 35 per cent.

CHAPTER 4

TRANSMISSION LINES AT THE U.H.F.'S

Showing how these lines can be used as inductances, capacitances, and tuned circuits, at the ultra-high frequencies

In the first three chapters of this book the development of the ultra-high frequency oscillator was traced, and the various modifications were explained. It will be recalled that Lecher wire tuning systems were found to be more useful at the ultra-highs than actual coils and condensers. Even in filament circuits a half-wave line was introduced to keep the ultra-high frequency currents from the filament power supply. But these lines, which are in reality sections of transmission lines, were put in without explanation of their basic operating principles. It is the express purpose of this chapter to clarify that portion of the preceding material. In order best to accomplish this end, the operating principles of transmission lines at the low frequencies will be given. Then the ultra-high frequency transition will be more understandable.

Although transmission lines at low frequencies find their most extensive use in radio as coupling systems between the transmitter and its antenna, at ultra-high frequencies their use becomes more universal. Not only are they used as mentioned before, but they also find application as interstage coupling devices, tuning circuits, inductances, capacitances, and even as high- or low-pass filters. And all this comes about by the simple expedient of varying the length and end construction of any ordinary transmission line.

Components of Transmission Lines. It would perhaps be best to begin with the ordinary inductances, capacitances, and

resistances — the types that can be bought in any radio shop. These units are said to be lumped because each is complete and distinct from the other. The resistance that is inherent in the wires of an inductance is, for this discussion, disregarded.

Now, if two wires are mounted close together, using one to conduct the current away from an a.c. generator and the other to conduct the current back, it would be found that what the generator " sees " is not just two wires with some resistance in them, but rather a complex impedance that, when broken down will reveal inductance and capacitance also. These last two are not visible upon examination of the wires, yet they exist, none the less. Of these two quantities, the presence of the capacitance can perhaps be more easily understood when the definition of capacitance is recalled, namely, two conductors separated by a dielectric. In this case, the dielectric is composed of the air between the conductors and the insulation, if any, wrapped around the conducting wires. Since condenser action is not perfect, it is shown schematically by placing a large resistance across the condenser, thus denoting leakage.

The presence of inductive reactance, however, is not so easily explained. It is well to go back to the concept of magnetic lines of force and attack the problem from that angle. Inductance may be defined as proportional to the total number of magnetic lines of force that encircle the wire, due to some standard amount of current, say 1 ampere. By starting with this idea, the inductance per unit length of any wire or system of wires can be developed. It will be found that the inductance of a pair of parallel conductors is given by

$$L = 1.482 \log_{10} \frac{d - r}{r} \times 10^{-3} \text{ henries/mile},$$

where d is the distance between the centers of the two wires,
r is the radius of one wire.

It does not make any difference what units d and r are measured in so long as they are both the same. For a concentric cable

where one conductor is inside of another one, the inductance is given by

$$L = 0.741 \log_{10} \frac{b}{a} \text{ mh./mile}$$

where b is the radius of outer conductor,
a is the radius of inner conductor.

To represent the above inductance, capacitance, and resistance in schematic form for analytical purposes, a diagram such as shown in Fig. 4.1 is often used in engineering books on transmission lines. Although the various components are shown separate

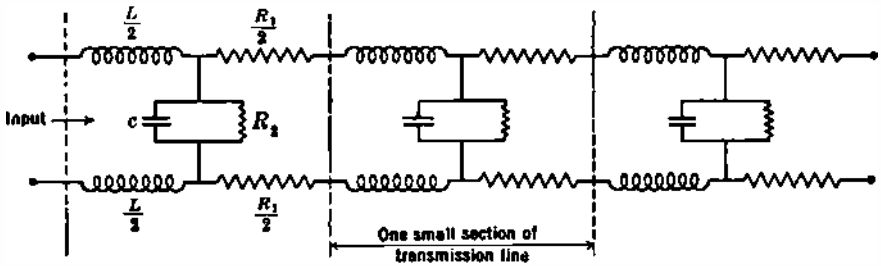


FIG. 4.1. The components of a transmission line. Although shown as separate, they are actually distributed.

and distinct from each other, they are really distributed evenly along the line. It is only because of our inability to show these components as they really are that we are forced to resort to this method.

It is advisable to point out at this time that any expression given for resistance and inductance should take into account the fact that both these quantities vary with frequency. The resistance must be corrected because, as the frequency increases, the current tends more and more to flow near the surface of the conductor and thus increase the effective resistance. This, of course, is the well known skin effect and is a function proportional to frequency. As for the inductance, it is found to change slightly because most of the current at high frequencies flows at the surface of the conductor, giving fewer internal flux-linkages. The percentage change of the inductance is far less than the corre-

sponding change in resistance with change in frequency. Any differences in the capacitances at the various frequencies are usually ignored since the error introduced this way is negligible.

Impedance Matching. Since the transmission line has these various impedances distributed along its length, there must be some definite value of impedance that a generator, or other electrical circuit, will see as it " looks " into one end of the line. Of course, it is assumed that there is some sort of circuit or " load," connected to the opposite end so as to form a complete path. If the circuit or impedance at the end of the transmission line is correctly matched to the transmission line, all the power from the line will be absorbed by the load and none will be reflected. But if there is a mismatch of impedances, some power will be reflected, some absorbed. Naturally, this case does not represent the maximum power transfer and hence results in a loss in efficiency. Also, the reflected energy results in standing waves of voltage and current along the transmission line — a condition that may or may not be desirable, depending upon the requirements of the particular case. For the ultra-high frequencies it is desirable.

The subject of reflection in transmission lines may prove puzzling to some readers who have not heard of this expression. Energy, and by this is meant voltage and current, will vary smoothly down a line if no change in the path constants occur. But suppose that the end of the line is reached and the load does not match the impedance of the line. In this case the energy will be broken up into two parts, one part being dissipated in the load and the other sent back down the line whence it came. It is entirely analogous to the situation of light waves hitting a piece of glass. Some of the light rays penetrate the glass and some are reflected. A new condition or medium was hit and this called for a redistribution of energy. Some energy was absorbed, some sent back. Every radio man knows that the output transformer in a radio set must be matched to the voice coil of the speaker, or else distortion results. This distortion might be thought of as being due to the rearrangement of the energy (current and voltage) in the circuit due to this mismatching. At broadcast fre-

quencies, when transmission lines are used, care is usually taken to see that all systems match. But at the ultra-high frequencies there are many occasions where the reflected wave is knowingly set up because certain effects are desired. These will be discussed shortly in more detail.

Characteristic Impedance. Associated with the transmission line are three other constants that need to be defined in order to make this discussion complete. The first of these additional constants is called the characteristic impedance of the transmission line and is denoted by the symbol Z_o. To understand this constant, imagine a transmission line which is infinitely long and which has a generator connected to one end. The impedance that this generator will see is termed the characteristic or surge impedance and will have one definite value dependent, of course, on the resistance, inductance, and capacitance per unit length of the particular cable. Cutting off a portion of the end of the line will have no effect on the input impedance, since there will still remain an infinite line. Any finite number subtracted from infinity leaves infinity unchanged. The equation for the characteristic impedance is quite simply developed in any communication engineering text and turns out to be

$$Z_o = \sqrt{Z_1 Z_2}$$

where Z_1 is the series impedance, as made up of the resistance and inductance, per unit length of line, as shown in Fig. 4.2.

Z_2 is the shunt impedance across the line, as represented by the condenser and resistor R_2 in parallel.

It should be remembered in all this discussion that a uniform transmission line is assumed so that any section may be taken as representative of the whole line.

Now follow the current and voltage values as they proceed down the line, and try to see the reason for the various changes. The type of measuring instrument used is of little consequence at this point but will be taken up in detail in a subsequent chapter. In Fig. 4.3, let the current that leaves the generator be labeled I. As this current is sent down the line, part of it (I_1) is diverted

through the leakage resistance and condenser. Thus, the current left (I_2) is equal to I minus the amount diverted (I_1). Farther on, the same process takes place again, resulting in the smaller

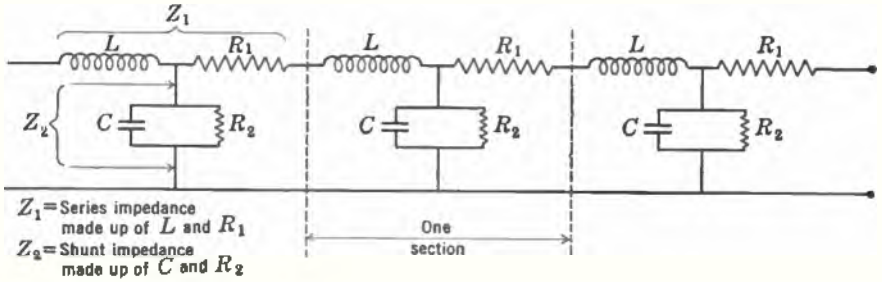


FIG. 4.2. The transmission line components L and R_1 are placed, for convenience, in one leg of the line instead of placing equal half portions in each leg as shown in Fig. 4.1.

current I_4. It can be seen that as this continues the resultant current will be eventually reduced to zero. In an actual cable, where the inductance, capacitance, and resistance are uniformly

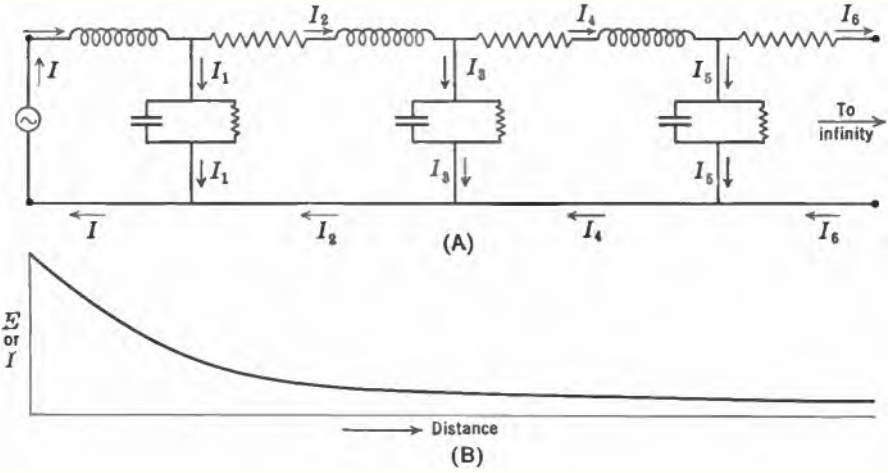


FIG. 4.3. Diminution of current with distance from generator on a transmission line.

distributed along the wires, rather than lumped as in Fig. 4.3A, there is no point at which the process is broken. Hence the diminution of current is gradual, as pictured in Fig. 4.3B.

Investigating the voltage between the wires as one progresses along the line, the same sort of diminishing effect is found, due, of course, to the various voltage losses which occur across the series resistances. These are pointed out in Fig. 4.4A. In Fig. 4.4B, it can be seen that the voltage across the line at any point will be less and less the farther it is from the source.

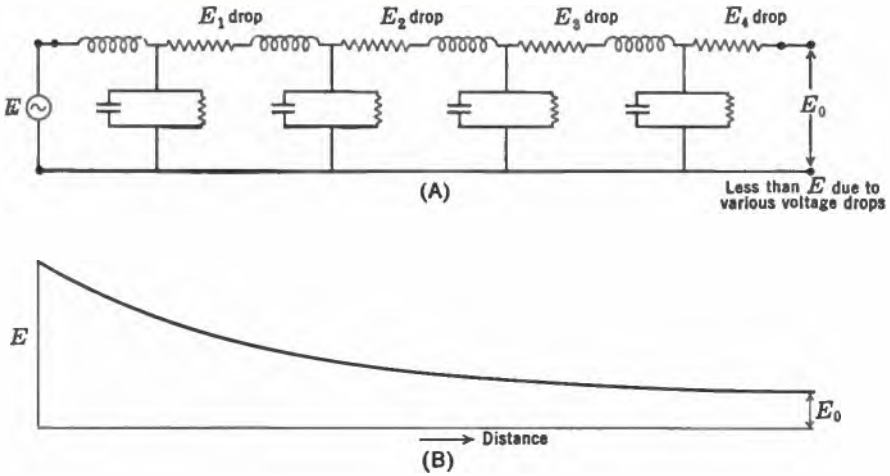


FIG. 4.4. Due to the voltage drops in the various series resistors, the voltage across the line diminishes as the distance from source increases.

Now, although all of the preceding discussion may be very enlightening, some readers may wonder as to its practicability, since infinitely long lines are not encountered in actual practice. However, the same effect as an infinite line may be obtained by taking any length of line and terminating it in an impedance equal to its characteristic impedance. By doing this, an infinite line is simulated in so far as any generator connected to the input of the circuit is concerned. All power transmitted down the line will be absorbed without reflection by the load, just as it is likewise absorbed eventually with an infinite transmission line. Fig. 4.5 shows this process pictorially and may be of further help in visualizing the action. Again, the case of matched and mismatched loads can be applied. By terminating a transmission

line in an impedance equal to its characteristic impedance, a matched condition is achieved, and the current cannot distinguish between this line and an infinite line. Then, no reflection occurs. The load absorbs all the energy as soon as it becomes available. When an impedance other than the characteristic impedance is used, the matching is not complete and reflection occurs. The importance of the characteristic impedance Z_o may now be appreciated.

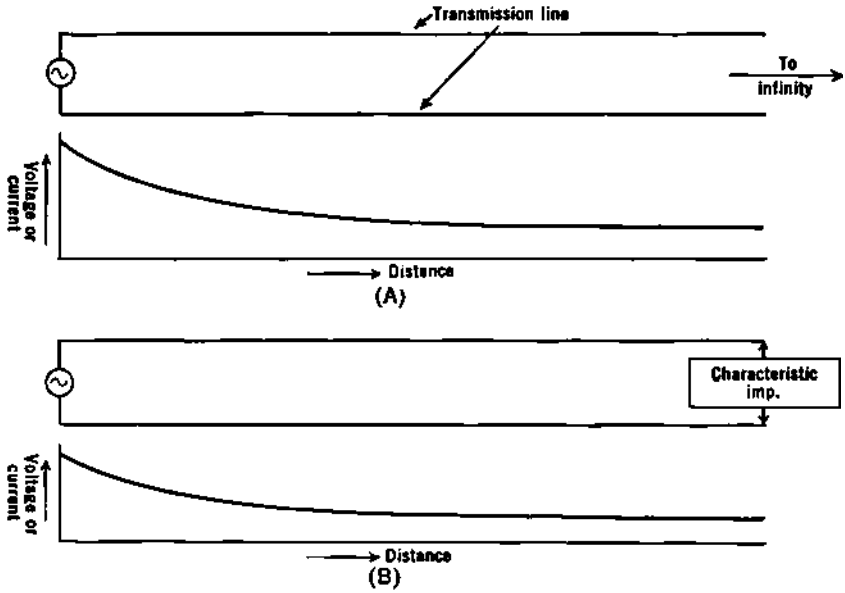


FIG. 4.5. On a line terminated in its characteristic impedance, the voltage and current values decrease gradually.

When a transmission line is to be used for communication purposes, the idea just developed with regard to terminating the line in Z_o is usually observed as closely as possible, since then little reflection will take place and the distortion will be kept to a minimum. However, if the characteristic of the line and its load should suddenly change, as for example by terminating the line in an impedance other than Z_o, the current and voltage distribution will no longer be gradual but will, in general, tend to vary sharply

from point to point. While this may prove unwise when intelligence is to be transmitted, it may be very desirable if the transmission line is to be used for some other purpose — perhaps as a tuned resonant circuit or even as an inductance or as a capacitance. This will be taken up very shortly and is mentioned here only to point out that at the ultra-high frequencies, transmission lines are not always used as such but may assume various forms. In order to achieve these other applications resort must be made to different terminations of the transmission lines. But do not forget that at any one frequency, if the transmission line is terminated in the appropriate Z_o (that is, appropriate for that frequency, since Z_o changes with frequency), all power transmitted will be absorbed and none will be reflected.

Attenuation and Phase Distortion. Distortion due to mismatching the transmission line at its load is just one of several considerations. There are two other types of distortion that may arise in a line in a manner quite independent of the load. One type is called attenuation and is due to the resistance inherent in the wires of the line itself. The other introduces a phase shift between various frequencies that are sent down the line. That this should be so is fairly obvious when it is recalled that a line has inductive and capacitive properties and that these reactances vary with frequency changes. The symbol used for attenuation is α and for phase shift, β. Usually both of these are expressed as having such-and-such a value per foot or per mile or other unit of length of the transmission line.

There are two widely used methods that are employed to minimize the attenuation and phase distortion. One method extensively used in telephone lines is to insert networks at various points in the line that tend to accentuate the frequencies that have been the most attenuated by the transmission line. In other words, these units have properties that are just the opposite of the properties of the line. They are sometimes referred to as compensators. The second method is usually more expensive and involves the use of specially constructed cable. This cable is built so that the attenuation constant is not dependent on fre-

quency, and the phase constant is made linearly proportional to frequency. With these two changes it can be shown that very little distortion of the signal will result.

Modification for the U.H.F. Line. At the ultra-high frequencies there are some simplifications that take place in the formula for characteristic impedance which has been given. The simplification is derived from the fact that at these very high frequencies the inductive reactance is usually much larger than the resistance in series with it (the line resistance), and the parallel conductance of the line is much less than the shunt capacitance. With these simplifications, the characteristic impedance Z_o reduces to

$$Z_o = \sqrt{\frac{L}{C}}$$

where L is the inductance per unit length of line, in henries, and C is the capacitance per unit length, in farads. The impedance acts as a pure resistance because the equation contains no reactive components (imaginary terms). As only the ultra-high frequencies are of interest here, this case will be discussed.

The three constants, namely Z_o, α and β, are not separate and distinct from the really fundamental units of the line, resistance, inductance, and capacitance. They were introduced to simplify the equations for the various properties of the line and can always be expressed in terms of the really fundamental units previously mentioned.

Use of Lines at the Higher Frequencies. We have now discussed the conventional low-frequency theory of transmission lines. It was seen that, when the transmission line was loaded with an impedance equal in value to the characteristic impedance, all the power sent down the line was absorbed and maximum power transfer was accomplished. This is very desirable, for example, in feeding energy from a transmitter to an antenna located some distance away. The transmission line that is the connecting link between the two units is carefully designed so that all input and output impedances are matched as nearly as possible.

There are many cases, however, where it is not desirable to terminate a transmission line in its characteristic impedance, but rather to go to the extreme conditions of placing a direct short on the end of the line or else leaving it entirely open. The line is now mismatched. We shall consider these conditions to see why they are important. In order to simplify the discussion and at the same time take the conditions usually encountered in practice, only transmission lines that are one quarter and one half wavelengths long will be presented.

Quarter-Wave Line. It will be recalled that the voltage and current values along a matched transmission line vary smoothly from one end of this line to the other. To determine the voltage and current distribution for a mismatched condition, let an alternating-current generator be connected to one end of a line which is one quarter of a wavelength long and let the other end be left open or, what is equivalent, terminated in an infinite impedance. Any signal sent out from the generator will hit this open end and bounce back, so to speak, somewhat as a sound wave hitting a stone wall is sent back or reflected. The reflected wave will travel away from the open end of the line and naturally cause interference with the waves coming from the generator. The result is a standing wave which will have values that vary from point to point. All zero values are called nodes and all maximum points are referred to as loops. Quite logically, the current at the open end of the line will be zero since here the line is open and no current can pass from one line to the other. The standing waves of current and voltage are due to the combined effect of the wave going toward the open end and of the reflected wave. Where they cancel each other we have the nodal points and where they reinforce each other the result is a maximum or loop. In between, the reinforcement or cancellation is only partial and so intermediate values are obtained.

Fig. 4.6 illustrates the relative phase of the reflected and incident waves when the far end of the line is open circuited. The current is used as the illustration. Here, both going and returning waves are shown separately. On the other hand, any

meter used to measure the value of current (or voltage) on the line will read the resultant of these two. Since it is the resultant that is important, only this value will be used hereafter on all diagrams depicting standing waves on transmission lines. Keep in mind, however, that the incident and reflected waves combine to form this resultant or standing wave.

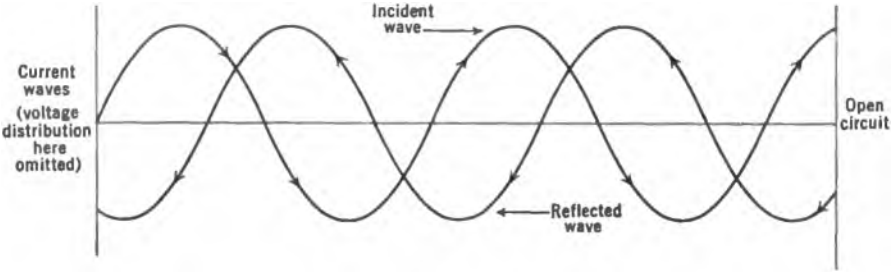


FIG. 4.6. A pictorial diagram showing what happens when a current wave on a transmission line hits an open circuit. These two waves combine to give a resultant.

Now, whereas the current is zero at an open end, the voltage is a maximum — a result of the voltage drop across the infinite impedance. The voltage distribution will also vary in a sinusoidal manner, but the voltage maxima will always occur where the current is zero. Hence the waves, as depicted pictorially, will be displaced as in Fig. 4.7. Using the ratio of voltage to current to define the impedance ($E = IZ$), it is possible to calculate the impedance at any point along the transmission line. At the current nodal points, the value of the impedance will be very high since

$$\frac{\text{large voltage}}{\text{small current}} = \text{large impedance}$$

At the current maxima points, the opposite situation will occur and the impedance will be low. In general, then, the value of the impedance along the line will tend to vary as the voltage, being high when the voltage is maximum and low when the voltage is a minimum.

It should be emphasized again that the only time standing

waves will occur on a transmission line will be for a mismatch between terminating impedance and line impedance. It is only under this condition that any energy will be reflected since, with a perfect match, all the energy is absorbed as soon as it reaches the load. Incidentally, it might be mentioned at this time that when amateurs wish to determine the amount of mismatch on

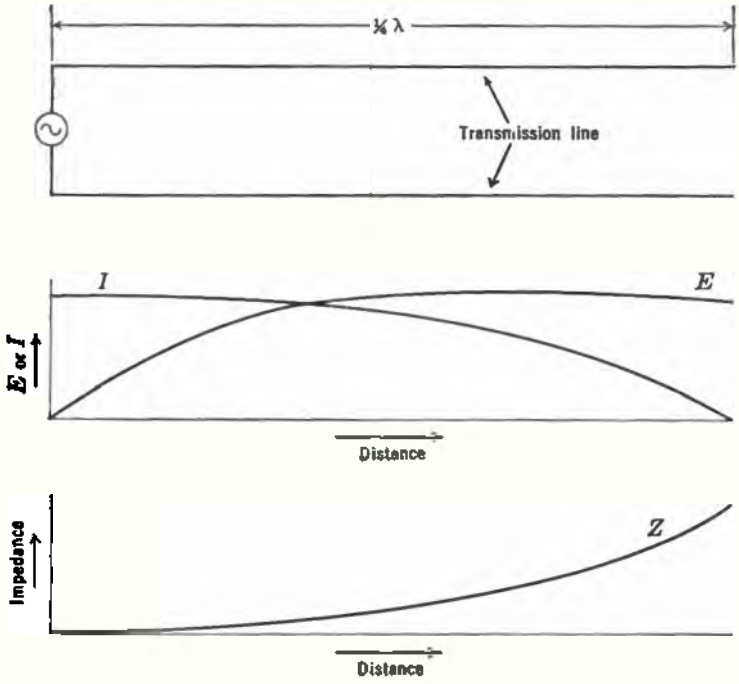


FIG. 4.7. Voltage, current, and impedance values on a ¼ λ transmission line that is open circuited.

lines which they are trying to match up perfectly, they use the ratio of maximum to minimum voltage as found on the line. Suppose this ratio turns out to be 6 to 1. This means that the voltage as measured at the voltage maximum points is six times greater than the minimum voltage. The minimum voltage, theoretically, should be zero, but actually it is found to have some small value. This ratio, then, will indicate that the load impedance is either

six times too large or one sixth as large as the line impedance. To determine which of the above holds true, it is necessary to measure the voltage across the load, a maximum voltage indicating the load is six times greater than the line impedance and a minimum indicating that the load impedance is too low. This generalization becomes more inaccurate as the reactive component of the load becomes greater.

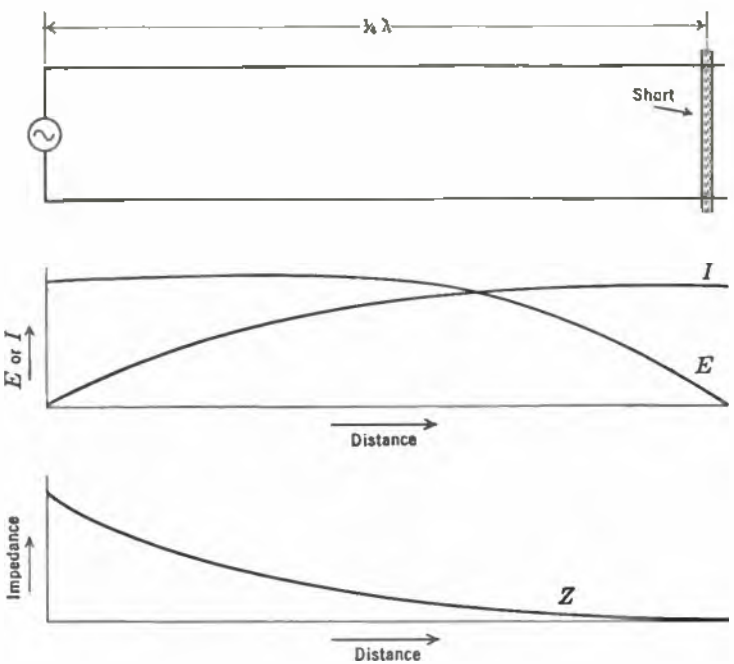


FIG. 4.8. Current, voltage, and impedance values on a short-circuited, ¼ λ transmission line.

If, instead of having the end of a quarter-wave transmission line open, it is now short circuited, the voltage and current standing waves formed along the line will be much different from the previous open case and, similarly, the impedance at various points will be changed. Fig. 4.8 shows the relationships in detail. By comparing Figs. 4.7 and 4.8 these differences will be immediately discernible.

Practical Applications of the Quarter-Wave Line. Now that the general differences between an open-ended and a short-circuited transmission line have been noted, it is time to bring out the practical applications that may be derived from these two conditions. To employ a transmission line as a resonant circuit, use is made of the standing waves to produce, at the input, either a high impedance or a low impedance, depending upon whether a parallel or a series resonant effect is desired. To show this graphically, start with Fig. 4.8, in which a quarter-wave section has been shorted at one end. By drawing the voltage and current distribution along the line and, from them, the impedance values, it can be seen that, when one end is shorted, the input will present a high resistive impedance to anything connected to it. The reason the impedance is resistive is due to the fact that the line is tuned and the quarter-wave line is now equivalent to a parallel tuned circuit. And, as is true with ordinary coil and condenser resonant circuits, this high value is obtained at only one frequency, the high impedance falling off on either side of the resonant frequency.

For the open-ended, quarter-wave line, the conditions depicted in Fig. 4.7 are obtained. With the far end open, the input end of the line will appear as a very small impedance and, since this quarter-wave line is adjusted for one frequency, its impedance will be resistive in nature just as is a series tuned circuit. From the preceding explanation, it can be seen that the quarter-wave transmission line makes a load appear opposite to what it is as, for example, short circuiting one end will make the input end show a high impedance. This is called load inversion or impedance transformation and is one of the most important properties of these quarter-wave lines.

For a practical example, refer to Chapter 1 and locate the high-frequency oscillators that used quarter-wave Lecher wire systems for tuning purposes. A Lecher wire system is nothing more than a two-wire transmission line and its action has been explained. Naturally, as the frequency is changed, the length of the Lecher wire tuning system has to be altered, since a quarter-

wave line at one frequency has a different length at some other frequency. For a wavelength of 5 meters the line length would be 5 divided by 4 or 1¼ meters. In practice, the actual length is not changed by cutting off any excess wire but rather by moving the shorting bar back and forth until the resonant point is found.

Phase Shifting with Quarter-Wave Lines. A second property of a quarter-wave line involves its use in circuits where a 90° phase shift is desired, such as in the Doherty modulating system. First, it can be shown that if Z_{in} is the impedance looking into a quarter-wave line, and Z_{out} is the impedance connected across the opposite end of the line, these two quantities are related by the formula

$$Z_{in} = \frac{Z_0^2}{Z_{out}}$$

where Z_o is the characteristic impedance of the line and is, of course, a constant. Since, in the ultra-high frequency case, Z_o is usually a resistance, the above formula means that Z_{in} times Z_{out} must result in a resistance and this can only be true if the two impedances are conjugates of each other. By conjugates is meant that if Z_{in} is $R + jX_L$ (a resistance plus an inductance) then Z_{out} will have to be $R - jX_C$ (a resistance minus a capacitance). These two quantities multiplied together will give a pure resistance. Another way of saying the same thing is to state that the quarter-wave line produces a 90° phase shift. To understand this refer to Fig. 4.9. Line A will represent the vector quantity $R + jX_L$. This will be Z_{in}. Line B will represent $R - jX_C$ and

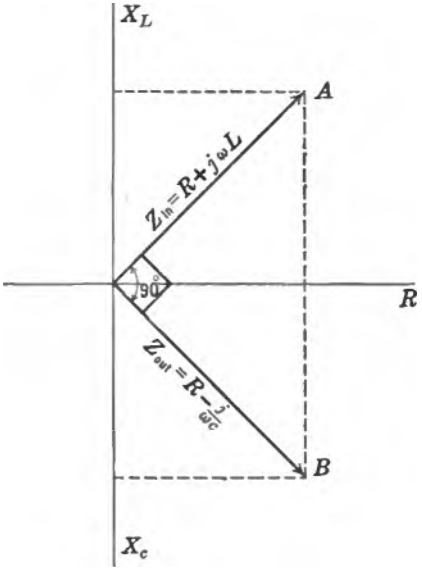


FIG. 4.9. Graph showing how a ¼ λ line changes the angle between the output and input.

will be Z_{out}. The angle between these two lines is 90° (as shown). Hence, if a network is needed that will produce a 90° phase shift, a quarter-wave line will do nicely. Thus, quarter-wave lines not only change the magnitude of an impedance, they also change the character of the impedance.

Impedance Matching. Another use to which the above inverting property can be put involves the practical application encountered by amateurs when they want to match a 600-ohm line from the transmitter output to a 72-ohm dipole antenna. Using the above formula and substituting for Z_{in} 600 ohms and for Z_{out} 72 ohms, the following result is obtained

$$Z_o = \sqrt{600 \times 72}$$

or

$$Z_o = 208 \text{ ohms.}$$

Thus it is necessary to obtain a line whose characteristic impedance is 208 ohms and which, when placed between the 600- and 72-ohm impedances, will effect a match.

To summarize the three important properties of a quarter-wave line:

1. The line can act as a resonant circuit, either series or parallel tuned.
2. Due to its inverting properties, a 90° phase shift is obtained.
3. Also due to its property of inversion, it can match two impedances that differ widely in value.

Half-Wave Lines. The next step in the development is concerned with the half-wave transmission line, its properties and uses and wherein it differs from the quarter-wave line. No trouble should be encountered if the half-wave transmission line is simply regarded as two quarter-wave lines joined together. Since it was found in the foregoing sections that one quarter-wave line inverts the load, two joined together will only bring it back to its original form by another inversion. By referring to Fig. 4.10, it may be noted that the same conditions prevail at both ends of the line, thus demonstrating the statement just made.

A half-wave line short circuited at one end will appear as a very low impedance at the other end and likewise, if the one end is left open, the other end will appear as a high impedance. Thus there is a 1 to 1 impedance transformer action here. Such a unit finds practical application in circuits where it is necessary to isolate the

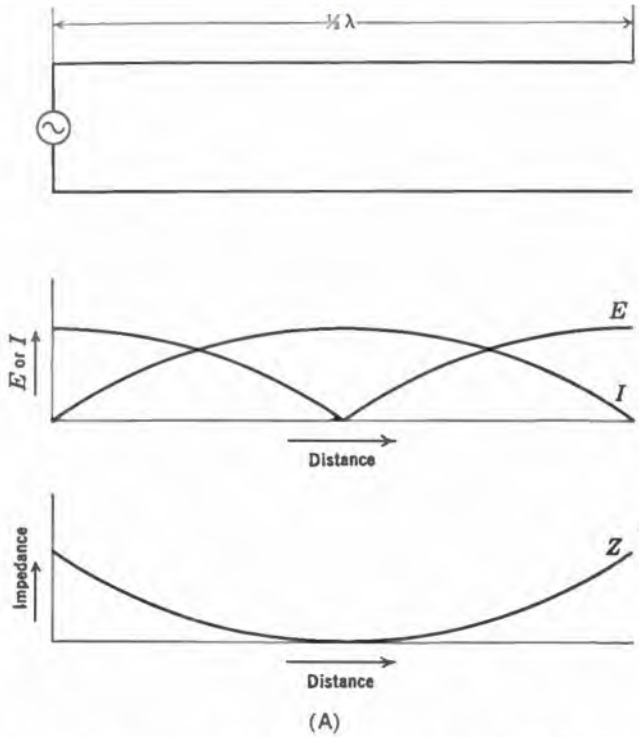
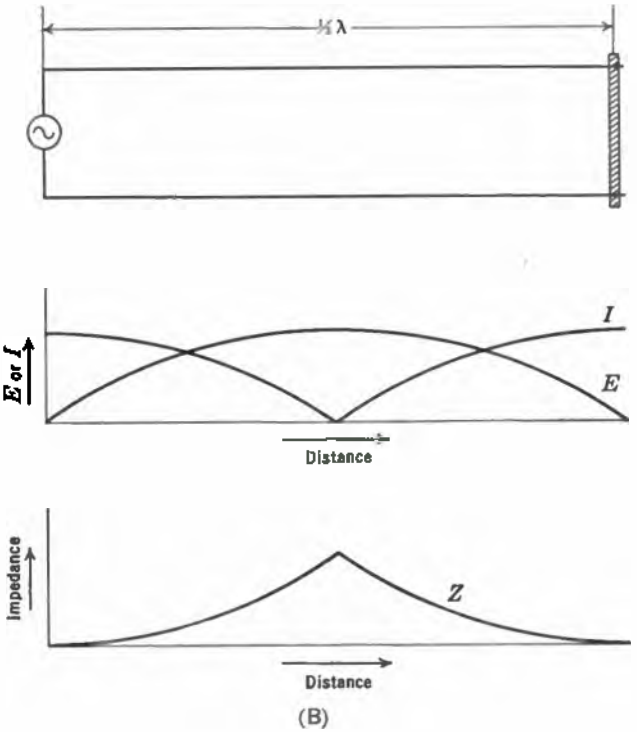


FIG. 4.10A. Voltage, current, and impedance values along an open-circuited, ½ λ line.

radio frequencies from the power supply but still allow passage of the direct current. For a specific example, refer to Chapter 1, where the high-frequency currents in the filament circuit are kept from the filament power supply by means of these half-wave lines. The isolating effect is obtained with a short-circuited, half-wave line. In this way, the point where the direct current

enters the high-frequency system from the power supply acts as a short circuit only for the ultra-highs and they can be considered as being by-passed just as with by-pass condensers.



(B)

FIG. 4.10B. Voltage, current, and impedance values along a short-circuited, ½ λ line.

Lengths Other Than One Quarter or One Half. Up to this point it will be noticed that the entire emphasis has been on either quarter-wave or half-wave lengths of transmission lines. The next possible question that might be posed would inquire as to the effects obtained when lengths of line between a quarter-wave and a half-wave length are used, or when lengths less than a quarter wave are used. In order to tie these in with the material that has been just presented, it would probably be best to delve a bit into the mathematical equations associated with transmis-

sion lines. It is quite logical that the input impedance of a transmission line will vary as the length of the line is changed since, by this process, more or less resistance, inductance, and capacitance are brought into the circuit. Of interest is the relationship between the input impedance Z_{in} and the length of line used. For a short-circuited line the relationship turns out to be

$$Z_{in} = Z_o \tan \theta$$

where Z_o is, as usual, the characteristic impedance of the line. As mentioned previously Z_o is a constant. The angle θ requires a little more explanation since it might not at first be obvious that the length of a transmission line can be measured either in feet or in degrees, the two being equivalent. To demonstrate this, consider a line that is 2 feet long and at a certain frequency has one complete standing wave on it. A complete wave has 360° and so, for this one particular frequency, 360° is equivalent to 2 feet of this line. One foot would allow only one half wave to be put on the line, again at this frequency, and now there is only 180°. The words " at one frequency " are important because, as pointed out previously, the amount of line needed to resonate at one frequency is different than that required at any other frequency. It is analogous to an ordinary tuning coil and condenser which will resonate at one

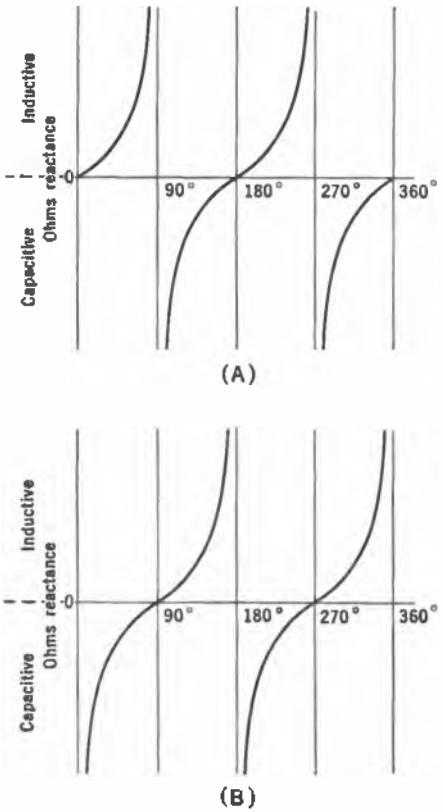


FIG. 4.11. (A) Input impedance for short-circuited line. (B) Input impedance for open-circuited line.

frequency but which must be changed before they will respond to any other wave, except a harmonic.

The plot of the above equation, in Fig. 4.11A, shows that when the length of line used is less than one-quarter wavelength, or 90°, the value of Z_{in} is positive (above the zero line) and so shows inductive properties. For values of θ between 90° and 180° the value of Z_{in} is negative, and so a capacitive effect is obtained. The remaining lengths will give rise to either inductances or capacitances, depending upon whether Z_{in} is positive or negative. At exactly 90°, 180°, 270°, and 360° the line is neither capacitive or inductive but acts as either a series or parallel tuned circuit, depending upon the angular length involved. If this seems strange, remember that in ordinary coil and condenser tuned circuits, whenever you tune above the resonant frequency, the combination has a capacitive reactance whereas, if the circuit is tuned below resonance, an inductive effect is obtained. And since a transmission line is a tuned circuit, it is only natural to expect the same behavior.

For the open-circuited line, the equation differs from the above and is given by

$$Z_{in} = Z_o \cot \theta.$$

The same interpretation should be given to this set of curves as in the previous case. See Fig. 4.11B.

Transformer Action in the Transmission Lines. Not only can a transmission line act as an inductance, capacitance, or resonant circuit, but it is also used as either a step-up or step-down transformer, depending again upon the type of load across the end of the line. For example, take a line that is an odd number of quarter wavelengths long and draw the voltage and current distribution all along the line. This has been done in Fig. 4.12A on a short-circuited, quarter-wave line. Now, if at some place like PP', a resistive load is connected across the line, it will be subjected to a voltage that is less than the voltage across the input of the line. And, as is true in all step-down transformers if the voltage is decreased, the current will be increased in like propor-

tion. Fig. 4.12B shows the similarity of the transmission line to the ordinary tuned circuit, where the same effect may occur. For voltage step-up purposes an open-ended line can be used if it is also any number of odd quarter wavelengths long and is connected as shown in Fig. 4.13. The load that is shunted across

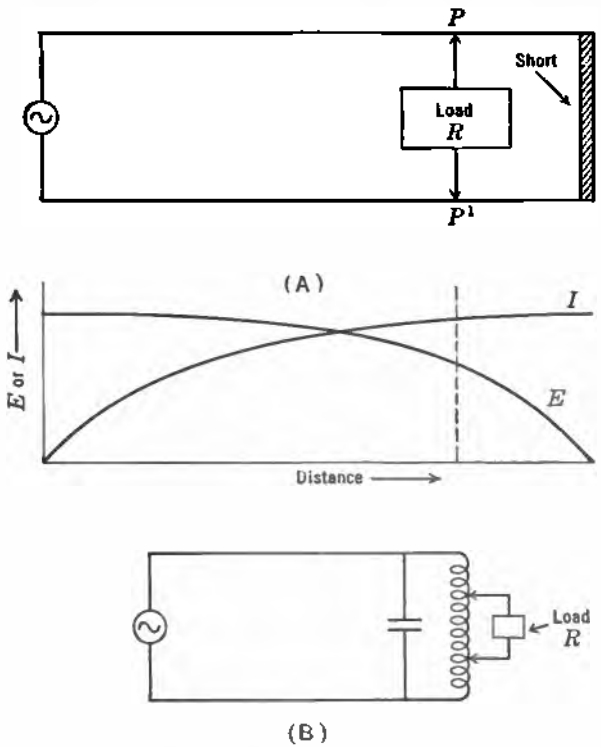


FIG. 4.12. Showing similarity between transmission line and step-down transformer.

the line should not contain too much reactance nor should its impedance be too low, as either case will disrupt the conditions as shown in the preceding diagrams and result in voltage and current distributions that will vary with each case.

Matching Stubs. Now that the fundamental ideas of transmission lines have been discussed, it is possible to investigate some applications where more than one property comes into use. An

excellent example would be the quarter- or half-wave matching stub that is so widely used by amateurs on short-wave antennas. To facilitate the explanation, refer to Fig. 4.14 where a transmission line and a matching stub are shown connected together at points AA'. The load at the end of the transmission line, Z_L,

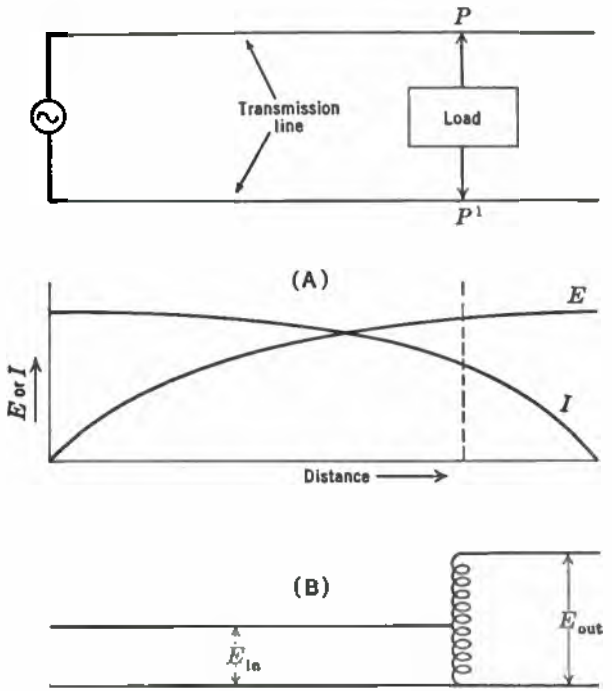


FIG. 4.13. (A) Using a transmission line as a voltage step-up transformer. (B) Low-frequency analogy.

may be an impedance; in amateur work it is usually the antenna. The point of this entire arrangement is to help couple the transmission line to the load and to have both impedances match as closely as possible for maximum power transfer. In most applications at the high frequencies, the characteristic impedance of the transmission line can be represented by a pure resistance while that of the load may not be. It is the purpose of the matching stub to cancel any reactive component in Z_L and thus have it

appear as a resistance to any generator feeding the transmission line. If the distance from the load to the points AA' is properly adjusted, the resistive component of the load will match Z_o and so leave its reactive component to be neutralized. The matching stub is now introduced to eliminate this unwanted load reactance.

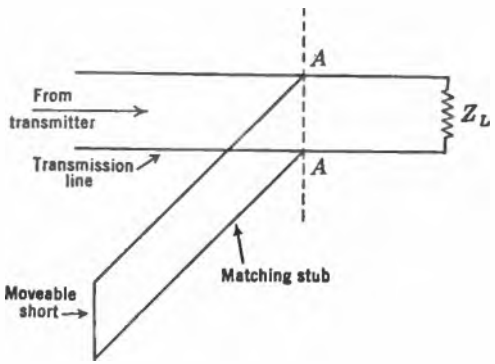


FIG. 4.14. Method of using a quarter-wave stub to neutralize any reactance in the load Z_L.

To adjust the length of the matching stub, short circuit one end (which is also near one quarter wavelength) and move the shorting bar until standing waves, which are present on the transmission line due to the unneutralized reactance of the load Z_L, disappear. When this happens, the reactive component of Z_L has been neutralized and the load presents an impedance matching that of the transmission line. Sometimes amateurs use a half-wave matching stub where the end is not short circuited but left open. The wire is cut off until the line is adjusted. Those readers who desire to use this arrangement on the amateur frequencies should consult the " Radio Antenna Handbook " published by the American Radio Relay League.

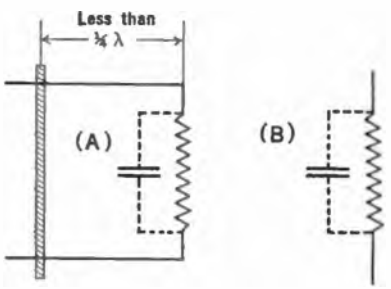


FIG. 4.15. Neutralizing the capacitance inherent in a resistor by means of a short length of transmission line.

Another variation of the above principle is illustrated in Fig. 4.15 where the capacitance that usually develops across a resistance at the ultra-high frequencies is neutralized by adding a short amount of transmission line, long enough to present an inductive reactance at the resistor and thus neutralize its capacitive reactance. The same idea is frequently employed to remove some of the effects of interelectrode capacitance, an ever present trouble at the ultra-highs.

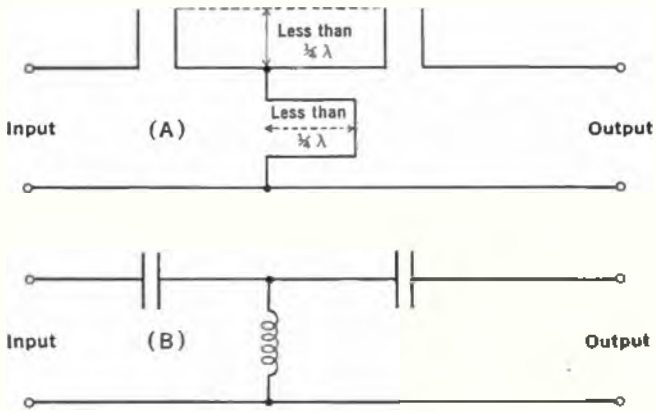


FIG. 4.16. (A) An ultra-high frequency high-pass filter. (B) Low-frequency equivalent.

Perhaps it has occurred to some readers that, as resonant lines can take the place of coils and condensers, such systems may be used to act as high-pass, low-pass, and band-pass filters. Such use has already been suggested. A typical circuit may look like that of Fig. 4.16A, where a high-pass filter is shown. The corresponding, conventional, low-frequency circuit is shown in Fig. 4.16B. Comparison of the two will aid in bringing out again the equivalent circuits developed at the beginning of this section. Many other arrangements will probably occur to the reader and it would help considerably if these were put on paper and practiced. It might be mentioned that all of the preceding material will also apply to concentric cable, as well as the older parallel wire systems.

In conclusion, a consolidated summary of the various properties of the transmission line is given in pictorial form. See Fig. 4.17.

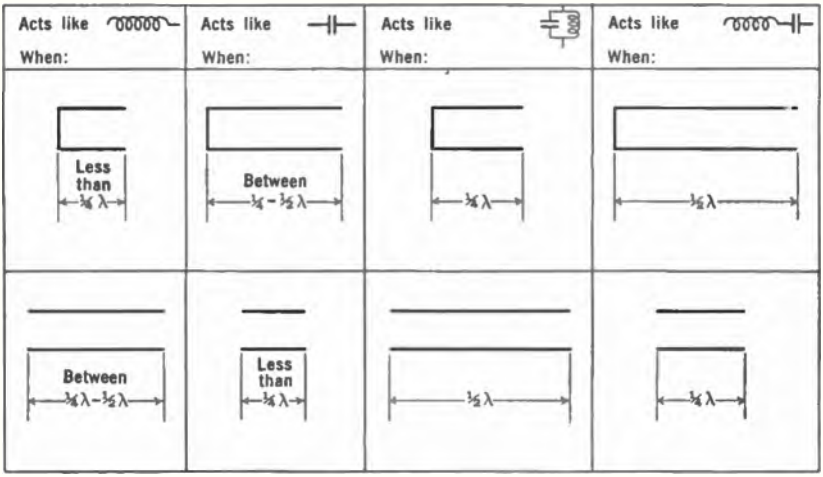


FIG. 4.17. Summary of properties of transmission lines ½ λ or less.

Summary. The transmission line consists of distributed inductance, capacitance, and resistance. So long as the line is terminated in its characteristic impedance, any energy transmitted down the line will be completely absorbed at the load end. This situation represents the ideal case when intelligence is to be sent from one point to another by lines.

For ultra-high frequency use, however, it is many times more desirable to utilize small sections of transmission lines as tuning units or filters. That a transmission line can be used as a resonant circuit arises directly from the fact that it is composed of the above mentioned inductances, resistances, and capacitances. Quarter-wave lines, short circuited at one end, present a high impedance at the other end and so act as a parallel tuned circuit. Like the ordinary tuned circuit, the impedance is high at only one frequency, falling off on either side of this frequency. This quarter-wave line replaces the parallel tuned circuit in high-

frequency generators because it is easier to handle and presents a high Q.

For quarter-wave lines that are open circuited, a series resonant effect is obtained. Half-wave lines are 1 to 1 transformers since the same conditions occur at each end of the line. They are used to keep the u.h.f. currents from d.c. power supplies. Combinations of these various units will give rise to high- and low-pass filters, quite analagous to the low-frequency circuits. By using proper lengths of lines, inductive or capacitive effects can be obtained and these may be used in the same manner as our ordinary condensers and coils.

If the foregoing facts are kept in mind, no trouble should be encountered whenever transmission lines are used in u.h.f. circuits.

CHAPTER 5

WAVE GUIDES

Wherein empty pipes become power conductors at the U.H.F.

The Old Way. If a radioman were asked just how he would send power from a transmitter to an antenna, the answer would most likely be — by the use of a transmission line. And this would be correct for most frequencies, at least until the ultra-highs were reached. And even at these frequencies transmission lines could be used. However, a better way has recently been devised, a method that introduces far fewer losses than even well constructed concentric cables. So it is only natural that this system should replace transmission lines at frequencies from 1500 megacycles up. Below this frequency, transmission lines will hold preference because the size of the conductors needed in the new system becomes too large. The name given to these new conductors is wave guides. It is the purpose of this chapter to illustrate their mechanism and resulting properties and to show how they can be used to guide power from the oscillator to its antenna. The important point to remember is that these wave guides take the place of transmission lines at the ultra-high frequencies. Here is another example of how low-frequency apparatus had to be modified to deal with the extremely short wavelengths.

The New Method. Returning to the usual method, one end of the transmission line is coupled to the output of the transmitter (or oscillator) and the other end is connected to the antenna itself. As the frequency increases, the size of the antenna decreases until at 1500 megacycles the length of a half-wave dipole amounts to one half of the wavelength in centimeters, or 10 cm. (½ of 20).

This is about 4 in. and not very long. Now suppose that, instead of using a long transmission line with which to connect the energy from the oscillator to the antenna, a very short length of line is used and the antenna is placed close to the set itself. Immediately some one would object, stating that for best transmission it is more desirable to place the radiator well in the clear and not close to any other objects. To overcome this objection, suppose that this small dipole antenna is left near the set, but enclosed in a rectangular pipe, the other end of which is placed well in the clear, in order that the waves coming out will not be close to other objects. Fig. 5.1 depicts the situation. In this case, the

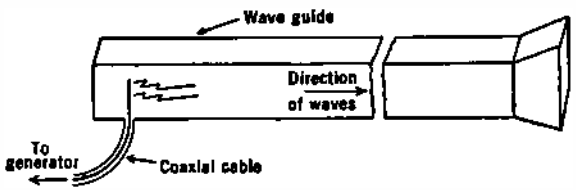


FIG. 5.1. Rectangular wave guide being used to transmit energy from generator to open space.

rectangular-shaped pipe acts as a guide for the radio waves given off by the small 4-in. dipole antenna. The waves travel down the guide and out into space when the end of the guide is reached.

The questions that might appear in a reader's mind would be: (1) would not the wave guide absorb a lot of this energy and (2), if this system is so good, why isn't it used for lower frequencies. To answer the first question, it will be shown that, when the waves are sent down the guide under certain conditions, much less absorption of energy takes place than would occur in an equivalent length of transmission line. And for the second question, it will likewise be shown that the frequency of the waves that can be sent down a guide without too much attenuation or loss depends upon the size of the guide itself. For the lower frequencies, wave guides could be used but they would be much too large and not very easily moved. However, they could be used.

To summarize this new method, then, a transmission line or

coaxial cable is used for only an extremely short length, just enough to bring energy from the transmitter to the small antenna placed in a wave guide close by. Once the energy reaches the small dipole antenna, it is conducted by the wave guide out into the clear where it is transmitted to some distant point to be intercepted by an appropriate receiver.

Now, in order to discuss the properties of a wave guide and their relationship to the waves that are propagated down these guides, it would perhaps be best to begin with the characteristics of the waves themselves and then tie these in with the necessary conditions for transmission of waves down the guides.

Electromagnetic Waves — How They Travel. Electromagnetic energy that is radiated by an antenna is the same whether the antenna is confined in a wave guide or unconfined as in the case of an antenna placed in the clear. A difference is apparent, however, in the direction in which the energy will travel. In the case of an open antenna, energy is propagated in all directions, whereas in a wave guide it is forced to remain within certain limits, determined by the guide's walls. Thus anything that is said regarding the general nature of these waves will apply equally to both cases.

The waves as they leave the antenna are referred to as electromagnetic waves. They have been set up by the oscillating or varying currents in the antenna itself. The set of rules or laws that these electromagnetic waves follow is referred to as Maxwell's equations. An idea of what these equations mean will be given in this chapter. For our purpose, it will be sufficient to use the general conclusions that are derived from Maxwell's equations. Electromagnetic waves or fields that are set up by the antenna are of the same general character as those sent out at the broadcast frequencies. Thus any general knowledge of wave propagation over the surface of the earth applies equally well here, with the two exceptions that a wave guide confines these waves and that the frequency is very high.

When dealing with an electromagnetic field there are several fundamental ideas that should be kept in mind. It might be

well briefly to consider the basic concepts at this point. Whenever energy is sent out by an antenna, it will be found that this energy consists of two forms. One is the electric field and the other is the magnetic field, and each is dependent on the other. It is impossible to have one present without the other, a relationship that is clearly brought out by Maxwell's equations. Besides their dependency on each other, these two fields are generally at right angles to each other.

Everyone is probably familiar with the fact that when currents flow through a wire, they set up a magnetic field around this wire. The magnetic field varies directly with the current. Another fact well known is that varying magnetic fields, when they cut across a conductor, will likewise induce an electromotive force (e.m.f.) in this wire. These two statements may be generalized to obtain the following:

1. Changing magnetic fields give rise to varying e.m.f.'s or, what is the same thing, varying electric fields or forces.

2. And changing electric fields or forces will likewise give rise to varying magnetic fields, and both will exist together.

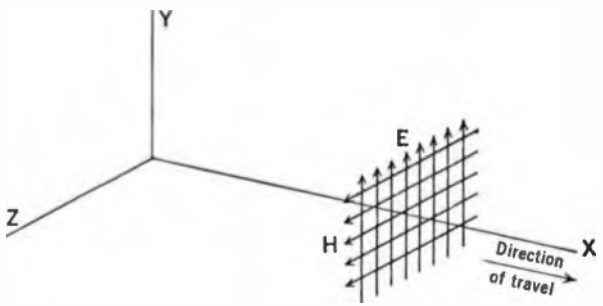


FIG. 5.2. A plane electromagnetic wave (shown at one instant).

Every electromagnetic wave may have electric and magnetic fields with components in either the X, Y, or Z directions. (The conventional rectangular system of axes is employed.) If this system of axes is not well known, refer to Fig. 5.2 where the above facts are shown. The fields referred to throughout this chapter are always changing in magnitude. Remember that it is only a

varying magnetic field that will produce an electric field and only a changing electric field that will give rise to a magnetic field.

It is from these variations that an insight into the wave propagation can be obtained. It would be best to start at the antenna where the high-frequency currents are rapidly oscillating back and forth along the transmitting wires. The moving currents set up rapidly changing fields that travel away from the antenna with the velocity of light (approximately 3×10^{10} cm. per sec.). Focusing attention on the magnetic wave for an instant: as this wave travels outward, it will, because it is changing in magnitude, induce an electric field in the surrounding space. But this electric field will not only appear in the same place where the magnetic field is, but it will also appear a little beyond; and this electric field just produced will, as it travels outward, produce a magnetic field — again slightly removed from the agent that produced it. It is by having these fields sustain each other, and at the same time move outward, that the mechanism of wave propagation may be pictured in a general way.

In space, the energy of the wave is divided equally between the electric and magnetic components. The balance of energy is disrupted only when the wave enters some new medium, such as some of the Kennelly-Heaviside layers found about 90 to 250 miles above the earth. Since no new energy is usually added to the wave, the only way that the readjustment can be brought about is for some of the energy already in the wave to be rejected, a process that we call reflection. Thus this process can be compared to light being reflected from a mirror, or sound waves producing an echo. In each case a new medium was in the path of the energy and in each case a readjustment was necessary.

Polarization of Radio Waves. In dealing with electromagnetic waves, many times it is easier to consider the electric field as being in only one direction, say the Y axis of Fig. 5.2, and the magnetic field as being in some other direction. By confining these fields to special directions and not allowing them to be in every direction, a process known as polarization has taken place. In general, any field that is restricted to certain directions is said

to be polarized. The reason this is mentioned at this time is due to the fact that in radio almost every radio wave suffers some sort of polarizing action as it travels through space. In the consideration of wave travel down wave guides, the direction of the electric and magnetic fields or, in other words, the polarization of the wave, is very important.

The plane wave shown in Fig. 5.2 is an example of a polarized wave, since the electric field is only in the vertical direction; everywhere else it is zero. In radio, the direction of polarization is always taken to be the same as that of the electric vector, although this is arbitrary and is only by definition. Vertical antennas send out vertically polarized waves and if these waves do not change their sense of polarization as they travel, they may best be received or picked up by a vertical antenna. At broadcast and at moderately high frequencies, the sense of polarization of the electromagnetic wave is not very important because the atmosphere is used for transmission of signals. At the very high frequencies, however, it will soon be shown that when using wave guides the type of polarization has an important bearing on the behavior of the wave as it travels down the guide.

All this has been by way of an introduction to the behavior of electromagnetic waves and has been presented here for two reasons: first, because most elementary radio courses only deal hazily with the fundamental concepts of radio waves in general and, secondly, because a basic knowledge of their properties must be at hand before any intelligent discussion of wave guides can be undertaken. And there are still more fundamental ideas to come which will be combined with the facts just given to form as complete a picture as possible without the extensive use of higher mathematics.

The Wave Guide. The case of an antenna sending out waves in free space unhampered by any man-made obstructions and free to move in any direction has just been mentioned. Let the antenna now be confined by an enclosure constructed, for example, of copper, and let this enclosure be rectangular in shape. Fig. 5.3 shows such a device and also orients it in space with regard to the

ordinary three-dimensional coordinate axes. The waves leaving the antenna will not be free to move wherever they desire, but will be confined or directed by the walls of the wave guide. It should be noted particularly that the electromagnetic wave does not travel through the conducting walls of the wave guide, but rather through the dielectric material contained by the walls. In most instances the dielectric is air, but it does not necessarily have to be so. The dielectric may be any non-conducting material. But it is usually air for the very simple reason that the losses are lower.

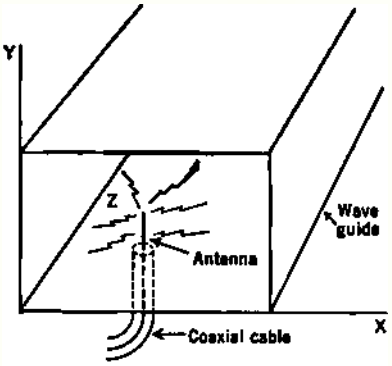


FIG. 5.3. A small ½ λ antenna in a guide.

Now, in order that the wave be propagated down the wave guide, several precautions must be observed. First, the electric lines of force that comprise part of the electromagnetic wave (see Fig. 5.2) must intersect the walls of the guide only at right angles. If these lines of force cut across the confining walls at any other angle, a current will flow and the original field will be distorted. That this is so can be understood from the following line of reasoning. An electric field will exert a force on any electrons that happen to be in its vicinity. If the electric field cuts across a conductor at any angle other than a right angle, the electrons in the conductor will move under this force. Moving electrons constitute a current and this will give rise to a magnetic field. But, if the ends of the electric field lines hit the conductor at right angles, the electrons in this conductor will only move a very small distance, equal to the thickness of the walls (which should be very small). Soon they come to rest, unable to leave the confines of the enclosure and continue into space. In this case, the electric field will not be distorted to any noticeable degree.

The second condition that must be obtained is to have the magnetic lines of force at right angles to the electric lines of force.

This was previously mentioned when discussing Maxwell's equations. Correlating all these facts will give us Fig. 5.4, in which the simplest type of electromagnetic wave, the so-called $TE_{0,1}$, is

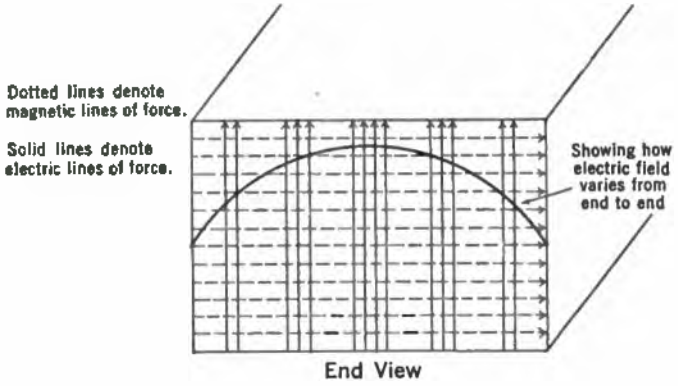


FIG. 5.4. Distribution of electric and magnetic fields of the $TE_{0,1}$ type wave.

pictured. The method of labeling the waves will be considered presently. The view shown is that which would be seen if one were to stand squarely in front of the wave guide and look down

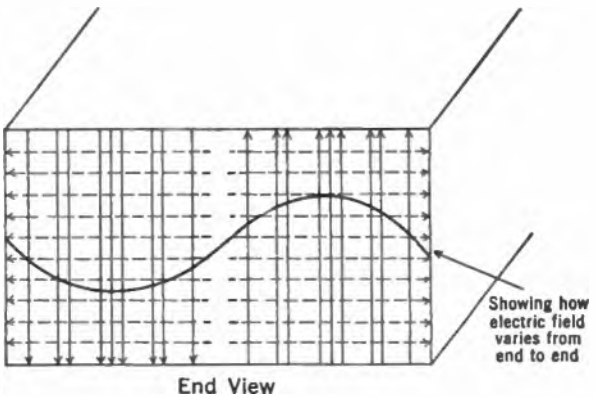


FIG. 5.5. Distribution of electric and magnetic fields of the $TE_{0,2}$ type wave. (Same notation as Fig. 5.4.)

its length. For the next higher order type of wave, refer to Fig. 5.5, in which the distribution of the various fields are shown. The wave here is referred to as the $TE_{0,2}$ type.

Depicting Electric and Magnetic Forces by Arrowed Lines. It is necessary for the reader to become familiar with the method of illustrating electromagnetic waves by means of their electric and magnetic field components. The reason for this will become increasingly obvious when, as this chapter progresses, it will be shown that the differences between various waves depend only on the differences between the electric and magnetic field arrange-

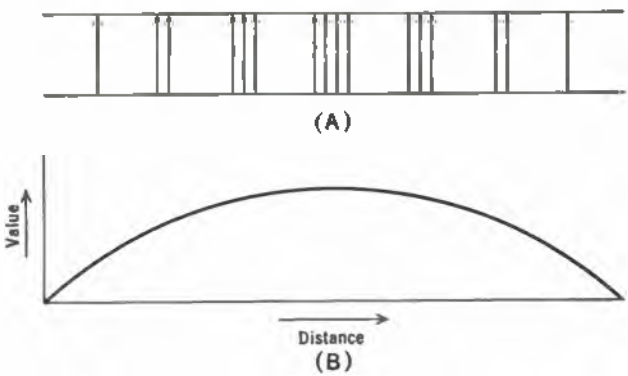


FIG. 5.6. (A) Using arrowed lines to show strength and direction of arrowed fields.
(B) Equivalent half-wave curve.

ments. It has been the custom in many electrical textbooks to show electric and magnetic fields by means of lines, these lines being called either lines of flux or, what is the same thing, lines of force. Arrows are usually placed on the ends of these lines to indicate the direction in which the forces act. To indicate that a certain region has a greater electric or magnetic force it has been the custom to put more lines of force per inch here than in other regions where the force is not quite as great. Fig. 5.6A shows what is meant. In part of this diagram there is shown a region in which, for this example, an electric force is acting. The force may be due to a group of electrons or any other charged body that is capable of exerting an electrical force. In the middle of the diagram the field is supposed to be concentrated or strongest, and this is shown by the group of four lines bunched together. Moving away from the center on either side the bunching of the arrows

grows less, indicating that the strength of the field is decreasing. At the extreme right and left margins there are no lines, indicating that here the strength of the electrical field has diminished to zero. If the diminution in electric field strength from the center was accomplished gradually, a diagram such as part B of Fig. 5.6 could be used. This is equivalent to one half of an a.c. wave, where the same conditions exist, such as a maximum value in the middle and zero at either end.

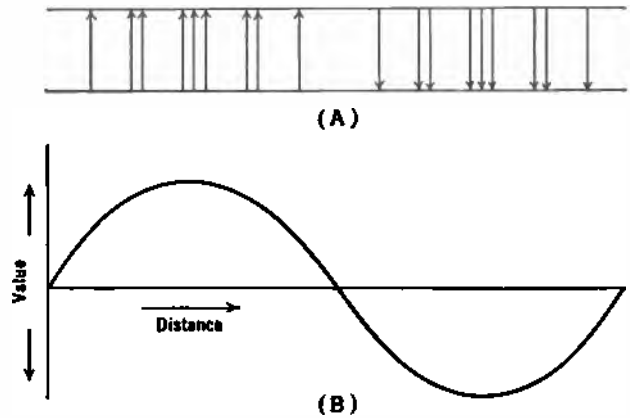


FIG. 5.7. Same as Fig. 5.6 except for full wave variation in field strength.

Fig. 5.7 shows the type of electric field distribution obtained when there is a full wave variation from one side of a wave guide opening to the other side. For the first half wave, the field is shown going in one direction, while in the other half wave the electric field is naturally going in the opposite direction, just as the current reverses in going from one half cycle to another in ordinary a.c. circuits. The process is continued for one more illustration in Fig. 5.8 where a variation in electric field strength equal to a wavelength and a half is shown. Following this line of reasoning, any number of cycles can be drawn.

Since, in any illustration of an electromagnetic wave, there are both magnetic and electric lines of force, the magnetic lines are shown as dashed in order to distinguish them from the electric

lines of force. The important point to remember is that, the greater the number of lines shown at a point, the greater the field strength at that point.

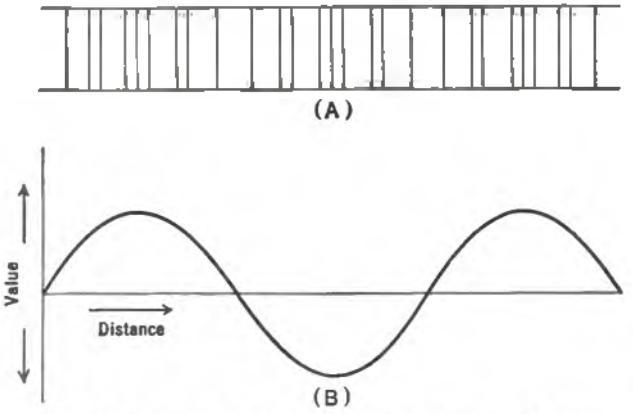


FIG. 5.8. Full wave and a half variation in electric field strength.

Wave Travel in Guides. In sending an electromagnetic wave down a wave guide, it might at first be thought that the energy travels in a straight line through the interior or dielectric of the guide. If this were true, and if the dielectric used consisted of air, there should be no difference between the travel time of the wave measured inside or outside the rectangular guide. But it has been found that the measured time inside the wave guide is always greater than the corresponding time taken in free space. This difference could only be explained if the path inside were longer than the outside distance and this could only be true if the interior waves followed a zig-zag path, reflecting back and forth off the walls as the wave moved down the tube. Fig. 5.9A shows how the reflections take place in order to bring about this increase of distance. The action could have been predicted to a certain degree, since the energy waves that leave any of our land-based antennas usually have a series of reflections from the Kennelly-Heaviside layers and the surface of the earth as they travel along. For rectangular wave guides, the velocity of each wave is found to depend on the following factors:

1. It is dependent on the frequency. It varies inversely as the wavelength, although in no case exceeding that of light.

2. It is dependent also on the height of the guide. (Dimension X in Fig. 5.9.)

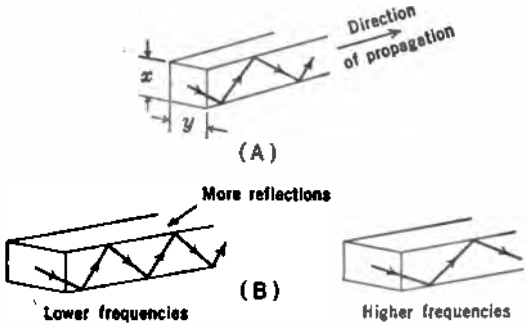


FIG. 5.9. Reflections of an electromagnetic wave as it travels down a wave guide.

Cut-Off Frequency. If the walls of the guide were perfect conductors, the energy of the wave would not decrease as it bounced back and forth along the length of the guide. However, the actual materials used, such as copper or brass, introduce losses that increase as the number of reflections increase and these increase as the wavelength gets longer. Referring to Fig. 5.9B, it can be seen that the longer wavelengths bounce off the walls more often as they are propagated down the guide. Each one of the reflections introduces losses, and so these wavelengths are attenuated much faster. In fact, as the frequency is lowered, a point is soon reached where the wave just bounces back and forth across the wave guide and is not transmitted at all. This point is known as the cut-off frequency of the guide and represents the lower frequency limit that can be transmitted by this particular guide. Here we have one of the important properties of a wave guide, namely, that it acts as a high-pass filter. All waves of a certain frequency will be transmitted more or less freely, whereas waves with frequencies lower than the cut-off frequency will have a large attenuation and will not be transmitted at all, or at least not very far.

To illustrate, take the case of a small rectangular wave guide where the dimension X (Fig. 5.3) is 10 cm. For the $TE_{0,1}$ wave pictured in Fig. 4, there must be a half-wave variation in electric field from one side of the guide to the other. In this case the distance is 10 cm. If the frequency of the wave being sent into the guide is low, a half wave will occupy more space than the allowable 10 cm., a situation that will not be tolerable under the rules laid down by Maxwell's equations. Thus the problem is to find the frequency (of the $TE_{0,1}$ type) which will give a half wavelength of 10 cm. so as to enable this wave to fit into the dimension X of the wave guide.

The formula that can be used is

$$\text{frequency} = \frac{\text{velocity of light}}{\text{wavelength } (\lambda)}$$

which, when solved for the frequency in the example above, gives an answer of

$$f = \frac{30{,}000{,}000{,}000}{2 \times 10}$$
$$= 1500 \text{ megacycles.}$$

Any wave having a frequency less than this would have a correspondingly longer wavelength (since λ is inversely proportional to frequency), and it would then be impossible to fit a half wavelength into 10 cm. This wave cannot be propagated down the guide. But a wave having a higher frequency will have a shorter wavelength, in which case it might be possible to fit one or more half waves across the mouth of the guide and still satisfy our previously stated conditions of zero electric field intensity at the sides. These waves will be propagated down the guide with very little attenuation. Thus it can be seen that a wave guide gives rise to a filtering action. The initial conditions must, at all times, be completely satisfied.

Now take the same rectangular wave guide and decrease the previous X dimension of 10 cm. to 5 cm. Using the above

formula, we obtain

$$f = \frac{3 \times 10^{10}}{2 \times 5}$$

$$3000 \text{ megacycles.}$$

The cut-off frequency has now gone up and no wave having a frequency less than 3000 megacycles will be propagated down the guide. Here is a peculiarity connected only with wave guides since there is nothing comparable to it in transmission lines.

The foregoing illustration has dealt with a relatively simple type of wave which allowed the direct usage of the ordinary formula showing the relationship between frequency and wavelength. However, as the type of waves used becomes more complicated, where more than one dimension is brought into use, such as X and Y, special formulas must be resorted to. At the present time this does not concern us since the more complex waves would involve frequencies that are too high to be generated easily. However, whether the waves are in the X direction alone (such as the $TE_{0,1}$ wave), or also in the Y direction, the rules laid down by Maxwell's equations must still be followed. In the previous discussion, the electric fields in the X and Y directions are meant; the Z axis contains the magnetic field at right angles to the electric field or fields. Refer to the figures of the TE wave if this idea is difficult to understand.

Transverse and Longitudinal Waves. Up to this point, the types of waves have been merely mentioned without taking time to explain the system used in obtaining this peculiar notation. It might be best to start at the beginning and explain the difference between a transverse and longitudinal wave. This will lead more directly to the desired results. To demonstrate longitudinal wave motion, take an ordinary tuning fork and hit it sharply, causing it to vibrate. This motion will force any air particles near the fork to vibrate, and the resultant waves will travel outward in all directions. The vibrations of the air particles will be in the same direction as the propagation of the sound waves, thus causing their paths to be parallel. This action is the defining

characteristic of a longitudinal wave. It should, of course, be remembered that the air particles vibrate about a mean point while only the energy is expanding outward. It is somewhat analagous to a situation where a straight row of billiard balls is tapped at one end. The energy is transmitted forward by the balls while the balls themselves remain relatively fixed in position.

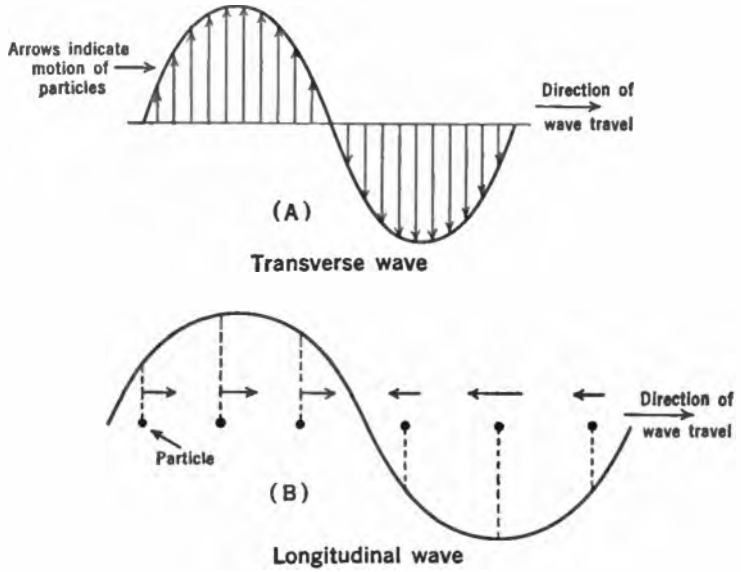


FIG. 5.10. (A) Motion of particles is up and down.
(B) Motion of particles is back and forth.

For the case of a transverse wave, consider the motion obtained by means of a rope, one end of which is secured and the other end moved up and down by hand. Here the rope particles move up and down while the wave itself is propagated forward. The two motions, that of the rope and that of the wave, are at right angles to each other. This combination of motions defines a transverse wave. Fig. 5.10 emphasizes pictorially the difference between the two types of waves.

***TE* Waves.** Transferring attention from sound waves and rope waves to electromagnetic vibrations, it will now be shown

how the preceding information can be used to label the various types of waves in rectangular guides. As mentioned before, every electromagnetic wave may have electric and magnetic fields with components along any of the three conventional directions, namely, X, Y, and Z (see Fig. 5.3). Suppose a wave had a magnetic field along the Z axis (Fig. 5.3) but no electric field in this direction. The electric components could be either in the X or the Y direction or both. Since the direction of *propagation* of the wave is in the Z direction, it can readily be seen that, as far as the magnetic field is concerned, since it has a component in the Z direction, this is a longitudinal wave. As described at the beginning of this section, the necessary conditions are satisfied. However, since the electric field has no components in the Z direction, but only in the X or Y directions, it can be called a transverse wave. This also follows from the definition of transverse waves. Thus, there is a choice. With regard to the magnetic field, it is a longitudinal wave, whereas with respect to the electric field, it is a transverse wave. The custom has been to label the wave after the transverse component. Hence in this case the notation would be TE, designating a transverse electric wave. Another name sometimes used is "H wave."

TM **Waves.** To reverse the situation, let there be an electric field component along the direction of propagation, the Z axis, but no component of the magnetic field in this direction. The latter will be confined to either the X direction, the Y direction, or both. (It has not been mentioned recently, but it may be recalled, that the electric and magnetic fields must be at right angles to each other. There could not possibly be a magnetic field and its corresponding electric field in the same direction.) In this case there would be a longitudinal wave with respect to the electric field and a transverse vibration with respect to the magnetic field. Retaining the usual notation of labeling the wave after the transverse component, we get, for this wave the name TM wave. The TM is an abbreviation for transverse magnetic. Another name that is sometimes used is "E wave." Having explained the two most important notations, a third and

final type of wave, in reality a plane wave, might be mentioned. See Fig. 5.2. Here the electric and magnetic fields are at right angles to each other and at the same time are perpendicular to the direction of propagation, truly a transverse wave, for there are no longitudinal components. The name given to this type of wave is a combination of the preceding two types and is *TEM*.

So far the nomenclature has been general, indicating only the class of waves in operation. There is still lacking a specific notation that will indicate the particular wave in any group. In order to bring this out more clearly Fig. 5.11 shows three types of

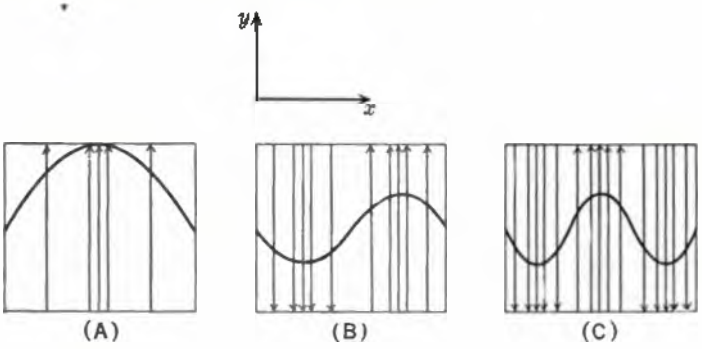


FIG. 5.11. The electric field of the $TE_{0,m}$ type of wave.
(A) $m = 1$, so we have $TE_{0,1}$.
(B) $m = 2$, so we have $TE_{0,2}$.
(C) $m = 3$, so we have $TE_{0,3}$.

waves that are in the *TE* class. The simplest wave in this group is shown in Fig. 5.11A, where it can be seen that the dimension X is just wide enough to accommodate a half wave of the electric field. In Fig. 5.11B, the frequency of the wave has been raised, so that now the same dimension X can accommodate a full wave. In Fig. 5.11C the frequency has again been increased and now there are three half waves. This process can be continued as long as higher frequencies can be generated. Each wave is differentiated from the other by a small subscript m. Thus the lowest frequency wave (Fig. 5.11A) would be represented by setting m equal to 1, Fig. 5.11B would have an m of 2, Fig. 5.11C

would have an m of 3, etc. The nomenclature is now more specific than it was and takes the form TE_m.

But there is no reason why the dimension in the Y direction should be ignored as it is perfectly possible to have an electric field in that direction also. In this case, use the subscript n and call each half wave of variation in this direction 1. Thus the final notation is $TE_{n,m}$. Specifically, the wave in Fig. 5.11A is $TE_{0,1}$, that in Fig. 5.11B is $TE_{0,2}$ etc. Zero is used for n because there is no component of the electric field in the Y dimension in the case shown; but let it not be forgotten that waves like $TE_{1,1}$, $TE_{2,1}$, etc., are perfectly possible. They are not frequently mentioned because of the extremely high frequencies involved, which for the present at least are not easily generated. Whatever has been said about the subscripts n and m is equally applicable to the TM type of waves.

An Explanation. Pause here for a moment and consider in a little greater detail the foregoing designations for different types of waves. The reader may well become confused by the great variety to be encountered. The situation can be made clearer if the fact is constantly kept in mind that the wave is always confined by the walls of the guide and so must obey Maxwell's equations. For an ordinary antenna placed out in the clear the limitations are less rigid, and so the above system of labeling different types of waves is unnecessary. In the clear, all waves can be transmitted. Now let us consider the situation in a wave guide. The lowest frequency wave that can travel down the guide (the cut-off frequency) is the one that can fit in a half wave across the mouth of the guide, say dimension X. If a higher frequency wave is desired, then two half waves or more must be fitted into the same space. To distinguish these latter waves from the previous simple half-wave distribution, they are given correspondingly higher numbers. Thus the reason for waves such as $TE_{0,2}$, $TE_{0,3}$, etc., becomes apparent. Each represents a higher frequency and each is still in agreement with Maxwell's equations. All these waves (of the $TE_{0,m}$ group) have the same basic characteristics, and they are governed by the same formulas.

The $TE_{0,m}$ waves have the electric field transverse, or at right angles to, the direction of propagation of the waves. There may be reasons, however, for generating waves with the electric field longitudinal or in the direction of propagation. These waves are labeled *TM* waves. One reason for using a *TM* wave in place of a *TE* wave of like frequency may be the question of attenuation. Perhaps one type of wave will have lower losses than the other. Or it may be that different signals are to be sent simultaneously, in which case both a *TE* and a *TM* wave could be transmitted and separately received, without any interference with each other. The situation may, in a sense, be compared to the use of certain combinations of antennas over others at the low frequencies.

For extremely high frequencies, complicated arrays may be employed, giving rise to waves like the $TE_{2,3}$ in which we have a full-wave electric field distribution along the *Y* axis and three half waves along the *X* axis (using the previous notation). The cut-off frequency would now depend upon both dimensions whereas, for the $TE_{0,m}$ case, only the *X* axis was the determining factor. The electric and magnetic fields become extremely complex, and the frequencies are so very high that we have not been able to generate them in any appreciable quantity. It may be some time before there will be any practical use for the complex waves.

It is interesting to note in connection with the generation of these fields that a straight antenna may be used to set up the electric field, which will in turn give rise to the accompanying magnetic fields. It is also possible to place a loop of wire properly in the guide so that the circulating currents in this conductor will set up the necessary magnetic fields. These will, in turn, likewise produce the electric forces. Either method may be employed, but practically, the single straight antenna is used because it involves fewer difficulties. The fields that are generated by any one antenna system do not belong exclusively to one type of wave pattern; they may give rise to other electric and magnetic field distributions. Such fields are, however, less intense and so may not prove bothersome. For most cases they may be ignored.

Excitation of Wave Guides — The *TE* Type of Wave. The discussion of wave guides has progressed far enough so that it is now possible to investigate the various methods which may be used to excite the guides and obtain the different types of waves. Excitation was only vaguely referred to previously when it was stated that an antenna was placed in the wave guide and the electromagnetic waves radiated from it. Each general pattern

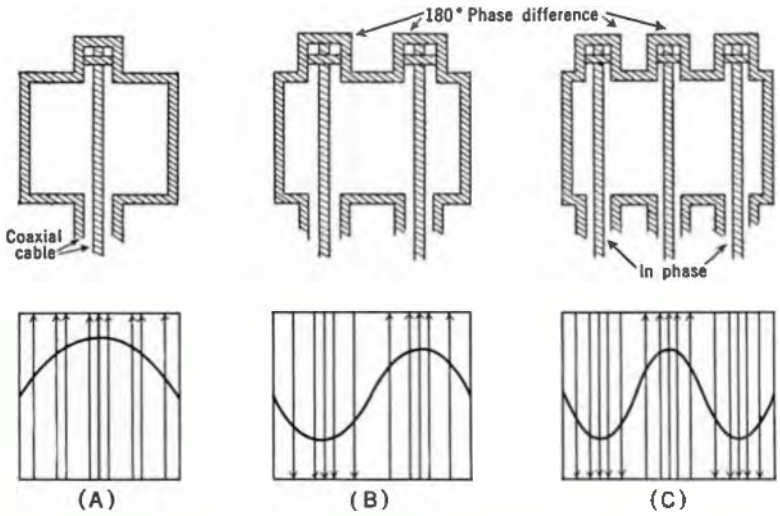


FIG. 5.12. Methods of generating the $TE_{0,1}$ (A), the $TE_{0,2}$ (B), and the $TE_{0,3}$ (E) types of waves.

of wave (both *TE* and *TM*) and each specific wave require a different set-up to produce the desired field patterns. However, the similiarity between the various waves of the *TE* group will be readily evident and the same can be said for the *TM* group. The generator will be omitted in each case and only the placement of the concentric cable in the wave guide will be shown. It will be assumed that the connection of the cable to the U.H.F. generator is understood. The simplest wave of the *TE* group, namely $TE_{0,1}$ can be generated by the type of arrangement shown in Fig. 5.12A. The electric field pattern is given underneath so a mental connection between the two can be formed. The $TE_{0,2}$

wave can be produced by the arrangement shown in Fig. 5.12B. Since, in the latter case, the exciting rods are 180° out of phase, the electric fields are in opposite directions on the left and on the right halves of the guide. The closer the lines of force, the stronger the electric field. As stated before, the intensity at the sides of the wave guide must be zero, and this condition is adhered to here. The case of the $TE_{0,3}$ wave is also shown in Fig. 5.12C. Note that in all cases just described it is the central or inner conductor of the coaxial cable that extends into the wave guide and acts as the antenna.

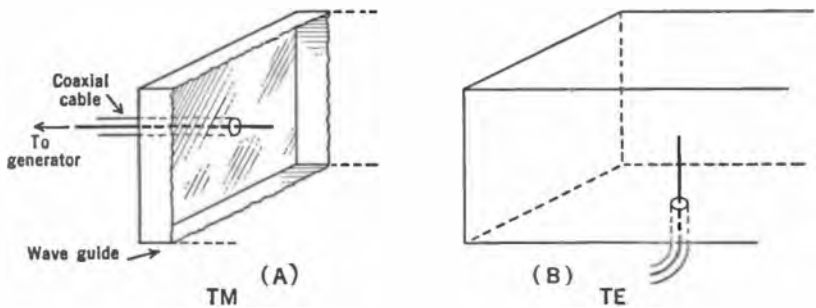


FIG. 5.13. A comparison of the general methods of exciting *TE* and *TM* types of waves.

***TM* Wave Excitation.** In order to produce fields of the *TM* type of wave, a set-up such as that shown in Fig. 5.13A is used. This is similar to the case of *TE* excitation except for the position of the antenna in the wave guide. For comparison, both are shown together in Fig. 5.13B. In general, for the $TE_{n,m}$ type of wave, the antenna will be at right angles to the direction of propagation, while in the $TM_{n,m}$ case, the antennas are parallel to the direction of propagation. This is logical since, as previously pointed out, the *TE* wave has no electric field in the forward direction while the *TM* wave has. It is only natural, then, that the placement of the radiators should also differ by 90°.

Usually, provision is made whereby the length of the antenna in the wave guide can be adjusted to give best results. This is an important adjustment since the measured intensity of the trans-

mitted wave may vary a great deal between the best condition and the worst. Another adjustment that is important is to get the proper distance between the antenna and the near end of the wave guide. This end wall acts as a reflector or backboard. Correct alignment of the antenna can also greatly improve the transmitting efficiency. Whereas, only transmission of the wave has been discussed so far, every radio man knows that a good transmitting antenna is also an excellent receiving antenna. Thus the adjustments just mentioned are important for the receiving antenna as well; the results work both ways. In the case of reception, however, the coaxial cable is connected to a crystal detector which rectifies the received current and makes it available for further use. Incidentally, these antennas, when used for reception, are very selective and will pick out one type of wave almost to the exclusion of all others. By having several types of transmitting antennas placed in various positions and by mounting receiving rods in similar positions at the other end of the guide, the guide may be used for multiple channel transmission. This is a feature that will probably be put to good use in days to come.

Cylindrical Wave Guides. The only wave guide discussed so far has been the rectangular wave guide. This kind of guide was chosen first because it is easier to visualize the actions of the electromagnetic wave in it than in guides having other cross sections. However, there are other forms of guides, most of them of little practical importance, except the cylindrical wave guides. As with the rectangular forms, it is possible to use the two notations ($TE_{n,m}$ and $TM_{n,m}$) to label the various types of waves that can be propagated down the length of a cylindrical tube. Propagation is only possible in these, as in rectangular guides, provided that the frequencies used are not too low for a particular diameter of the guide. The same conditions pertaining to electric and magnetic fields still hold true even if the shape of these fields has changed. The change is due to the difference in shape of the guide.

Usually, when dealing with cylindrical wave guides, the geometrical notation is changed from that used with rectangular

guides. Examine Fig. 5.14. Note that the axis in the direction of propagation is left the same, the changes occurring where we previously used the X and Y axis. Instead of these, we now deal with the radius r and the angle θ. This allows us to specify distances from the central axis and also to indicate the amount of rotation, through the complete 360° about the axis. Electric or magnetic fields which have components along any of these directions are differentiated from each other by subscripts. Thus E_z means an electric field component in the Z direction, E_r pertains to the radial direction, and E_θ tells of any field in the circular path. The same, of course, holds for the magnetic field components.

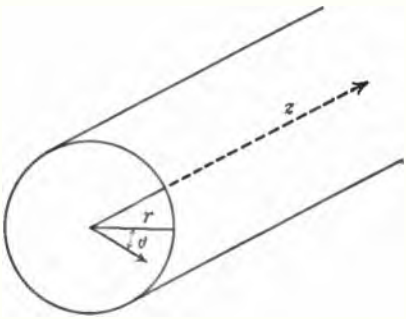


FIG. 5.14. The directions in a cylindrical wave guide, using cylindrical coordinates.

For excitation of, say the $TM_{0,1}$ wave in cylindrical guides, it is possible to use the set-up pictured in Fig. 5.15A. The resulting field pattern is also shown (Fig. 5.15B), the view being an end one.

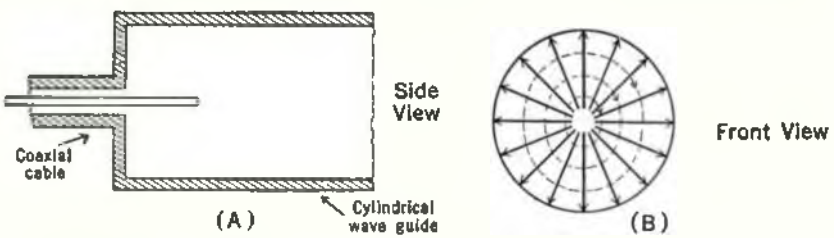


FIG. 5.15. (A) Method of generating $TM_{0,1}$ wave in cylindrical wave guide. (B) Arrangement of electric and magnetic lines of force.

The $TE_{1,1}$ type depicted in Fig. 5.16B has only transverse components of the electric field. Its method of excitation is shown in Fig. 5.16A. The critical or cut-off frequency is higher for the $TM_{0,1}$ wave than it is for the $TE_{1,1}$ wave, identical diameters being used in each case. This can be seen easily when the form-

ulas are given:

$$f_{\text{cut-off}} = \frac{1.15 \times 10^{10}}{r \text{ (cm.)}} \text{ cycles/sec.}$$

for the $TM_{0,1}$ wave
and

$$f_{\text{cut-off}} = \frac{8.79 \times 10^{9}}{r \text{ (cm.)}} \text{ cycles/sec.}$$

for the $TE_{1,1}$ wave. Note that due to the inverse relationship of the frequency to the radius, a higher cut-off frequency goes hand in hand with a smaller opening of the cylinder. Remember

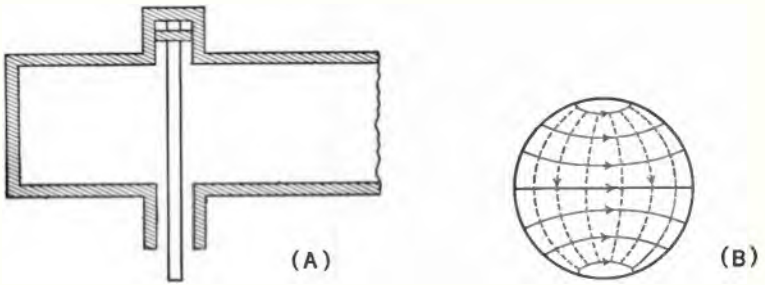


FIG. 5.16. (A) Method of excitation of the $TE_{1,1}$ wave.
(B) Arrangement of lines of force for $TE_{1,1}$ wave.

again that this frequency represents the lowest frequency (or longest wavelength) that will be propagated down the tube without too much attenuation. Practically, it is possible to transmit below this frequency, but the range is rather limited. Theoretically, all higher wavelengths can be easily transmitted without much attenuation (using actual conductors). In practice, however, the upper frequencies begin to show excessive attenuation because of increased skin effect. In between these two extremes a frequency can usually be found that will have least attenuation. Whether or not it is used will be determined by other factors that may assume greater importance than attenuation.

Uses of Wave Guides. It is now possible, since the action of a wave guide has been described in detail, to study some of the

uses that have been evolved for them. First and foremost is the important application of carrying ultra-high frequency power from one circuit to another. The wave guide is essentially a transmission line which has been modified for the very high frequencies. Its extensive use is due to its low loss or small attenuation. The losses occur because the walls of the guide are not perfect conductors. Some readers may resent the fact that it has been mentioned several times that imperfect conducting walls will result in power losses although no explanation has been advanced which would prove or explain this. The explanation can be found quite readily, but only if the investigator is equipped with a good knowledge of electric and magnetic fields and also the differential equations developed by Maxwell. As only an elementary knowledge of mathematics has been assumed throughout this text, all complicated expressions have been omitted.

Another use for the wave guide is possible, that of a high-pass filter. This property is concerned with the action of a guide that allows only frequencies above a certain minimum value, the cut-off frequency, to pass and to reject all lower values. This effect does not necessarily have to take place at the opening of the guide; it may occur anywhere along its length. Thus, if a pipe of large diameter is tapered to one of smaller diameter, waves that could be propagated in the large pipe might not continue after the diameter has been decreased and thus would be attenuated or reflected; probably both. The point is, however, that they would not continue in the smaller portion of the pipe. In this way, frequencies may be separated, an action that can only be brought about at lower frequencies by coils and condensers. In fact, the analogy to filter circuits can be extended even further by combining sections of wave guides that have different filtering characteristics.

Another use that will be mentioned briefly here (but will be described later in greater detail) is the cavity resonator which can be considered as a section of a wave guide. Just as it was possible to take a small section of an ordinary transmission line and convert it into a resonant circuit, so it is possible to close one end of a small section of a wave guide and get the same results, only at

higher frequencies. The transition from the tuned circuit can be pictured as starting out at the low frequencies with a coil and condenser, becoming a section of a transmission line at the hyper frequencies, and then finally ending up as a cavity resonator at the ultra-high frequencies. The function still remains the same but the means of accomplishing this are changed.

The final use of wave guides to be considered here will also be discussed in greater detail in a later chapter. Let us investigate the action of a wave guide as an antenna. The wave will travel down a wave guide without reflection of any of its energy just so long as no changes occur in the guide itself, a fact equally true in any transmission line. However, should any change occur, it immediately results in a reflection of energy back along the guide.

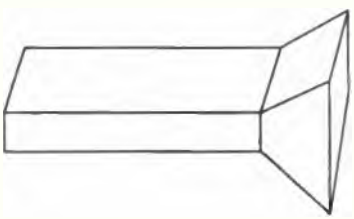


FIG. 5.17. Flaring the end of a wave guide produces an electromagnetic horn.

For transmitting purposes, the open end is of greatest importance. Some of the energy coming to the open end will continue out into free space while some of it will be sent back. However, if the change is not made abruptly, but rather is introduced gradually, the amount of energy reflected will be less and the transmitting efficiency of the wave guide will increase appreciably. This flared type of guide, Fig. 5.17, is known as an electromagnetic horn and has directional properties. It will be considered in greater detail along with other types of ultra-high frequency antennas to be taken up presently.

Summary. Electromagnetic energy at the ultra-high frequencies can usually be carried from one point to another not too far distant by means of wave guides. The guides are hollow tubes, either cylindrical or rectangular in shape, and serve as transmission lines having low attenuation properties.

To deal with these guides, it has been found necessary to use the general equations for the energies involved and this has meant dealing with electromagnetic waves rather than electric currents, as is usually done in ordinary lines. It is closely akin to the

process involved when dealing with the waves that are sent out from any transmitting antenna, except that now the energy is more or less confined to a certain specific space called a wave guide.

Many of the properties that have been discussed in the chapter on Transmission Lines are applicable here, but some new facts are brought in that have not previously been mentioned. For example, the idea of *TM* and *TE* types of waves. These two modes of transmission are differentiated from each other by the longitudinal component of the wave that is being transmitted. If the electric field has components only at right angles to the length of the guide, the name given is $TE_{n,m}$ and, if the magnetic field has only transverse components the wave is said to be of the $TM_{n,m}$ type. Each mode has its own properties and even each type within each mode, such as $TE_{0,1}$ and $TE_{0,2}$, have different characteristics. Of course, all the types under one mode are related by the same general properties. But the differences are enough to distinguish one from the other.

So far, only the simpler types of waves have been used to any great extent since the field patterns obtained by their use are easier to analyze and the frequency of these waves are within a range that we can generate fairly easily. As more is learned about these waves and as the frequency that can be generated is raised, the other types will come under extensive analysis.

Wave guides also find use as antennas if the end of the guide is flared so as to effect an easy transition for the wave from the inside of the guide into open space. The flaring does not make the change too sudden and hence results in less energy being reflected back down the tube. Another important use of guides arises when the tube is completely sealed, except for a small antenna opening; in this form we have a cavity resonator. This type of tuned circuit is used quite extensively with Klystron tubes.

CHAPTER 6

CAVITY RESONATORS

The tuning unit of the future

Why Cavity Resonators? It has been noted in previous chapters that the trend in tuning apparatus was from lumped circuit elements at the low frequencies to distributed elements at the ultra-highs. The necessity of using lines arose from the fact that at the wavelengths generated by the Klystron and magnetron (10 cm.), the size of coils and condensers would have to be much too small to be of any practical use. Besides, the Q of these circuits would have been very low, not at all comparable with values which can be obtained from resonant lines. Now, as the frequency is raised to 3000 megacycles (10 cm.), even the length of tuned lines becomes a little impractical. For example, a quarter-wave line at a wavelength of 10 cm. is only 2.5 cm. long or roughly 1 in. This certainly could not be handled with ease and would bring on the same situation that was encountered with regular tuning coils and condensers.

Another disadvantage encountered with tuned lines lies in the undesired radiation of energy since these systems are not easily shielded. Thus the time was ripe for modifications of the existing equipment which, it was hoped, would lead to more efficient means of solving the problem. The result of further experimentation led to the development of the cavity resonator which was used as the tuning device on the Klystron oscillator discussed in Chapter 3. It is the purpose of this chapter to investigate this new tuning device and see just how its properties compare with those of the resonant circuits that preceded it.

Development of the Cavity Resonator. An understanding of cavity resonators may be attained from two directions. One of

these deals with tuned lines and the other with wave guides. Both methods will be given since it is believed that a better insight into cavity resonators will be had this way. Starting with the tuned transmission lines, consider the quarter-wave section shown in Fig. 6.1. If a generator is coupled to it, the familiar voltage and current distribution shown will be obtained. At the shorted end, the current will be a maximum and the voltage will be zero, or at least very low. At the open end, the reverse will take place, resulting in low current and high voltage. These are the conditions that must exist if the line is to operate as a resonant circuit.

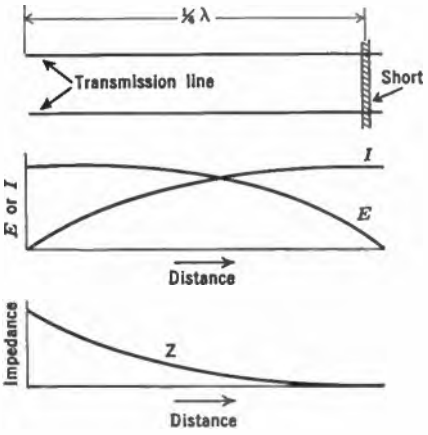


FIG. 6.1. The voltage, current, and impedance values on a short-circuited, ¼ λ line.

Now suppose that it is desired to excite two quarter-wave lines placed in parallel across the a.c. generator, as shown in Fig. 6.2.

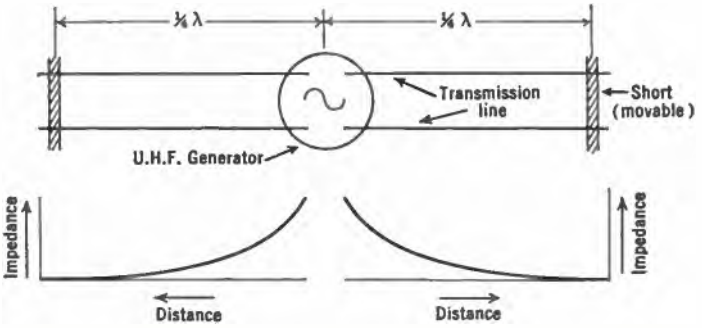


FIG. 6.2. Connecting two quarter-wave lines across a generator.

This combination should have a higher resonant frequency because the inductances and resistances in the two lines are in parallel with each other. The end result is a smaller inductance and

resistance than that of either component. Continuing the process of connecting more and more lines in parallel with the generator would finally result in a container having a shape as shown in Fig. 6.3A and 6.3B. The generator would be at the center, exciting the infinite number of units that have now been blended into one. This, then, is the final product, the cavity resonator. It might be mentioned that, in this process, the capacitance has changed very little. It has been found that inductance is a much easier unit to control, in cases like this, than the capacitance.

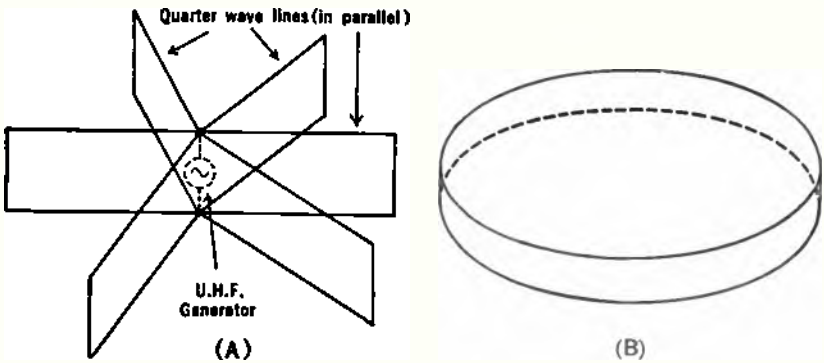


FIG. 6.3. By carrying the addition of short transmission lines to the limit, we would obtain, finally, the cavity resonator shown in Fig. 6.3B.

It will be recalled that, in the Klystron oscillator, instead of actually having a generator at the center of the cavity resonator, holes or openings were made at the center and that bunches of electrons were sent through. These electrons excited the cavity resonator and started its operation. But, whether an actual generator is located at the center of the cavity resonator or bunched electrons are sent through, the end result is the same — excitation of the resonator. The reader will note, at this point, that the electron itself assumes greater importance at the ultra-highs than at the lower frequencies. This situation does not arise because the electron has different functions at the shorter wavelengths, because it has not. Rather it is due to the fact that the phenomena can be more easily explained when the individual electron and its action are considered.

From the foregoing analysis, it can be argued that the cavity resonator is essentially the same as an infinite number of quarter-wave tuning lines placed in parallel and touching each other. Remember the voltage distribution on the quarter-wave line shown in Fig. 6.1. It will help tie in the cavity resonator, as obtained by the above process of reasoning, to the resonator which will be derived now from a wave guide analysis.

The easiest way to begin is to deal with a small section of wave guide, as pictured in Fig. 6.4. One end of the wave guide has been blocked and the other end left open so that, by means of a small antenna, waves can be sent down the guide. The waves will travel unmolested until the closed end is reached. Upon striking this solid wall, a reflection of energy will take place, whereupon there will be waves traveling now in two directions, something quite similar to a transmission line when a change occurs along its length. In order to have the reflected waves reach the opening and arrive there so as to reinforce the electric field set up at this place, the end blocking wall may be made movable and adjusted so that this reinforcement does take place. Standing values of electric and magnetic fields are now set up. This situation is very similar to the quarter-wave tuning line.

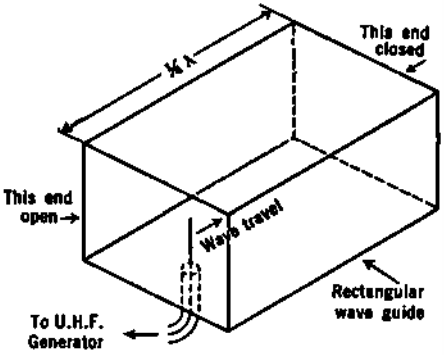


FIG. 6.4. Demonstrating resonance in a closed wave guide. This is quite similar (in principle) to the whistles small boys often used.

To complete the picture, take a similar section of a closed wave guide and place it on the other side of the transmitting antenna as shown in Fig. 6.5. Waves from the antenna will travel in both directions away from the antenna and both will be reflected when they strike the two closed ends. If the lengths of both guides are equal and correctly adjusted, the standing waves set up will reinforce at the antenna and a continuous pattern, as

shown in Fig. 6.6, will be obtained. The apparatus is now a cavity resonator which will resonate best for only certain fre-

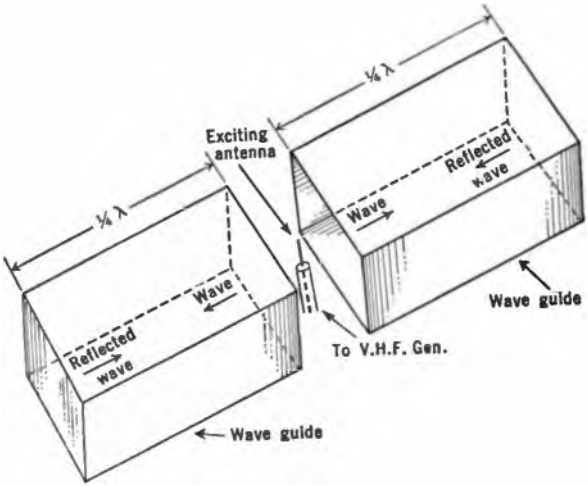


FIG. 6.5. The cavity resonator (entire unit) derived from two quarter-wave wave guides.

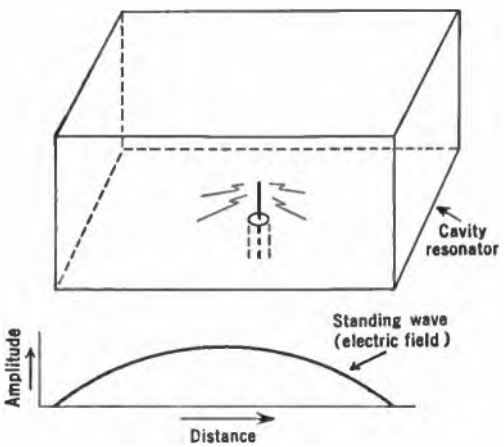


FIG. 6.6. The complete resonator with exciting antenna.

quencies. The standing waves set up in this device may easily be likened to the standing waves of currents and voltages obtained on tuned transmission lines.

The two methods of deriving cavity resonators show a marked resemblance to each other, and it should become apparent to the reader that, although it was not specifically mentioned in the chapter on Wave Guides, the wave guide may be considered as a number of parallel transmission lines carried to the limit. In transmission lines, all relationships deal with the currents and voltages that are found to exist on these lines, whereas in guides only magnetic and electric fields are mentioned. The relationship between the two is very easily brought out by electric theory and will be mentioned at the appropriate places from time to time. By pointing out the similar conditions that exist on our everyday equipment with those encountered at the ultra-highs, the transition in ideas and thinking may be made that much easier. Bear this in mind as these paragraphs progress.

Sizes of Cavity Resonators. Cavity resonators must be at least a half wave in length in order to act as a tuning unit. This fact can be understood when the development process is recalled, for it will be remembered that in order to arrive at the resonator, the distance from one end of the transmission line to the generator was one quarter wavelength. To go on from the generator to the opposite end would mean another quarter wavelength distance, giving a total of one half wavelength. Naturally, if a one half wavelength cavity resonator will work, so will a resonator that is any whole number multiple of one half wavelength, such as a full wavelength, one and one half, etc. This same sort of situation holds on ordinary transmitting antennas where a half-wave Hertz, a full-wave Hertz, or any full multiple of a half-wave antenna will work. The basic principle remains the same. The object in having the length of the resonator exactly an integral (or whole number) multiple of the first half-wave unit is quite obvious. Any wave sent out, say from the antenna, must be reflected and returned so that the maximum amplitude or strength of the standing wave is obtained. This will occur for only one distance or, as stated above, multiples of this distance. The minute any change in distance occurs, the forward and backward (or reflected) waves will no longer act so as to aid each other and

weak standing waves will be set up, resulting in inefficient operation of the line or resonator as a tuning unit. The distance from the antenna to the place the wave is reflected is critical and must always be carefully adjusted.

Electric and Magnetic Fields in Resonators. An understanding of the distribution of the r.f. voltage on an ordinary coil is quite important when coupling it to other circuits. Likewise, the distribution of currents and voltages on transmission lines is very important and was thoroughly investigated for the same

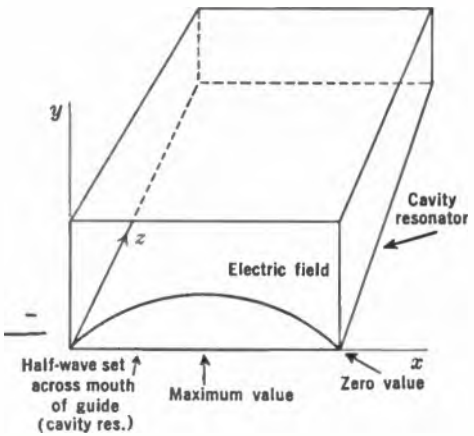


FIG. 6.7. Fitting a half-wave across the width of a cavity resonator so that it conforms to Maxwell's equations.

reason. However, for cavity resonators, the important points to determine are not currents and voltages, but the electric and magnetic fields. A study of the way in which the electric and magnetic fields are distributed in a cavity resonator should not be too complicated if the facts mentioned in Chapter 5 are remembered. The various wave intensities were shown and described for wave guides and, as just pointed out, a cavity resonator may be considered as essentially derived from a wave guide.

Consider the small rectangular cavity resonator shown in Fig. 6.7. Any electric wave set up must be so arranged that at

any of the walls its intensity will be zero, otherwise currents due to this electric field will flow at these boundaries and distort the whole wave. This condition, it will be recalled, was also adhered to in the wave guide. In order to satisfy this rule, a wave will be set up across dimension X (the width of this rectangular box) so that it is maximum in the middle and gradually tapers to zero at the sides. This is shown in Fig. 6.7, and represents a half wave.

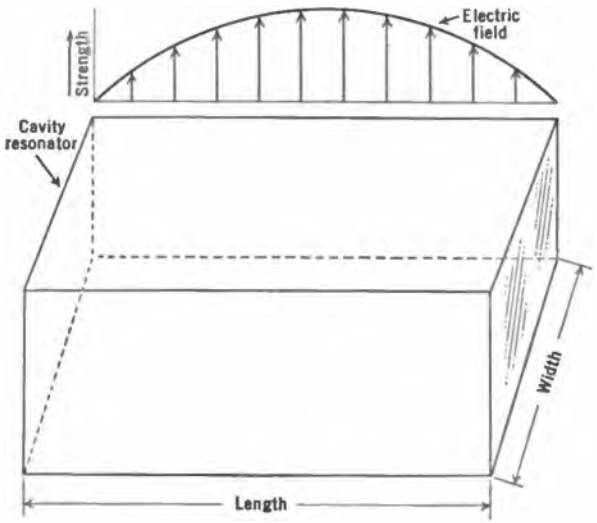


FIG. 6.8. Adjusting the length of a cavity resonator in order that the electric field (and consequently the magnetic) has the correct variation from point to point.

The next point to be considered in the electric field distribution will be the length of this rectangular cavity or, as given in Fig. 6.7, the dimension Z. Since a three-dimensional cavity is being dealt with, the electric field will vary not only along the X axis as described above, but it will also change as it moves down the length of the enclosure, the part that is now being considered. And, as this wave travels in the Z direction, it must eventually reach the end wall of the cavity resonator where, again, the electric field must be zero. This is depicted separately in Fig. 6.8, where the length of the resonator has been adjusted so that it is one half wavelength long. Thus, if a three-dimensional picture

is to be visualized, there would be a half-wave distribution in electric field intensity along dimension X (Fig. 6.7) and one half-wave distribution along the length (Fig. 6.8). Fig. 6.9 is used to illustrate another situation, where the electric field intensity along the X axis is still the same, but the length is greater, a full wave and a half distance. No matter how long or wide these resonators may be, even in different forms, the rules as laid down in previous chapters regarding the distributions of electric and magnetic fields

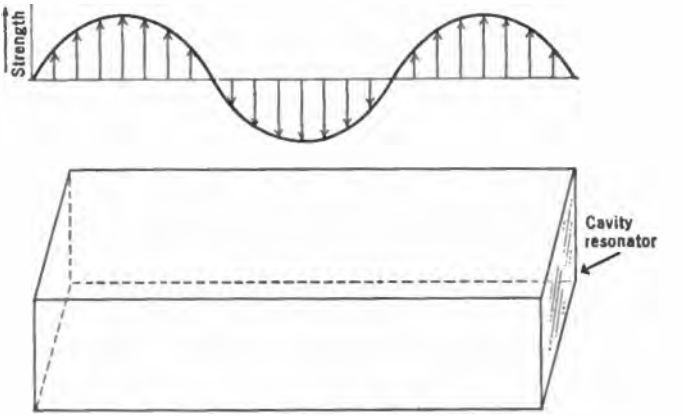


FIG. 6.9. The length is now $1\frac{1}{2}$ wavelengths long. The equations are still obeyed, allowing the cavity to be used as a resonator (width constant).

still hold. This explanation has dealt exclusively with electric fields but, of course, there are magnetic fields present at the same time. These have been omitted purposely in order to simplify the discussion and to avoid a complicated diagram in three dimensions, something not easily comprehended. As in wave guides, or with plain electromagnetic waves in space, the r.f. energy oscillates between the magnetic and electric fields. One instant it is in the electric field, the next it has smoothly changed over to the magnetic field. The rate at which the changes occur is the same as the frequency of the wave itself. And, as pointed out before, it is through this interchange of energy that the wave sustains itself and moves forward. Whether in a cavity resonator, a wave guide, or in ordinary space, the same things happen in the same

manner. The only difference refers to the minor restrictions imposed by the confining walls of the wave guide or cavity resonator.

Exciting Cavity Resonators. Even with all the preceding discussion, one question of great importance remains unanswered. If the cavity resonator is an enclosed hollow cavity, how will energy be introduced to resonate it and, after this has been accomplished, how will energy be withdrawn? To understand how a cavity resonator may be excited, consider the chamber shown in Fig. 6.10. This chamber is similar to the other cavity

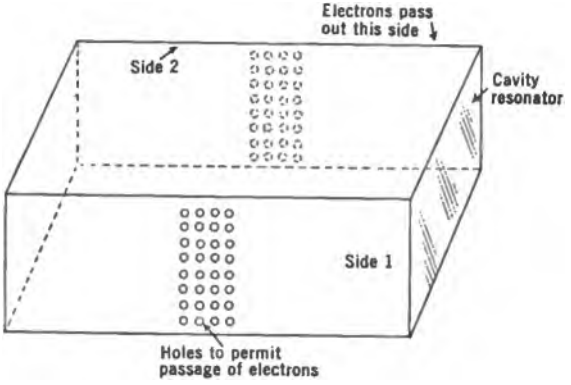


FIG. 6.10. Excitation of a cavity resonator by groups of electrons passing through holes drilled in the sides of the cavity.

resonators previously described except that the sides have been slightly altered. At the center of the sides a series of holes have been drilled to allow passage of a stream of electrons through the middle of the resonator. Every time an electron approaches side 1, there is a movement or displacement of negative electrons away from this side, repelled by the charge on the approaching electron. As this electron passes through the openings in the resonator itself, the induced positive charge may be thought of as moving with it. However, the positive charge is confined to the surface of the metal. When the electron leaves the resonator through side 2, the electric charge that it induced may be considered as released and now it moves back to its equilibrium posi-

tion. Thus there is a flow of current in one direction when the electron passes through and a flow in the opposite direction when the induced charge returns to its previous rest position. This may be compared to a pendulum that is pushed aside every time something passes it and then returns to its former position when the disturbance has gone. In both cases oscillations will occur, which is the desired effect. If the exciting electrons swing through the cavity resonator at definite times, then the oscillations set up in the resonator will continue, just as in the case of an ordinary oscillating circuit. Needless to say, the definite frequency mentioned in the last statement must bear a direct relationship to the resonant frequency of the cavity resonator.

Modified Cavity Resonators — The Klystron. In the Klystron tube described in Chapter 3 the cavity resonator (Fig. 6.11) is connected to and made part of the tube. The section that was

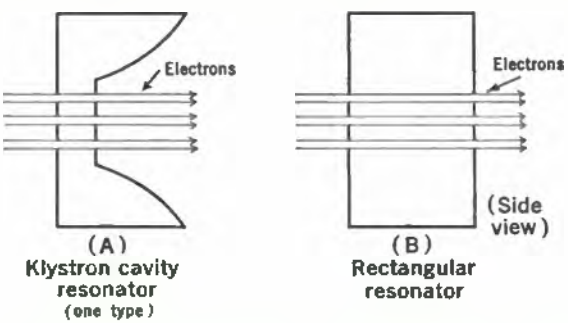


FIG. 6.11. Due to electron transit time the resonator must be narrowed at the points where the electron beam travels through.

drilled full of holes to allow passage of the electrons acts as the first two grids or the buncher. A similar arrangement is set farther along in the tube and constitutes the catcher grids and catcher resonator. The electrons coming from the cathode are first speeded up and focused by the first grid and then pass through the buncher. The action from here on has already been given and will not be repeated. Comparison of the resonator described in this chapter with the one used on the Klystron in Chapter 3 shows that the construction is not quite the same. The Klystron reso-

nator has been reprinted in Fig. 6.11. Alongside of this has been placed the rectangular resonator outlined above. The difference is immediately discernible, for the distance that the electron must travel through the center portion of the Klystron resonator is much less than the distance through the same part of the rectangular resonator. There is a definite reason for this, namely, the electron transit time. The electron requires a definite time to travel from one side of the resonator to the other. It is desirable that the time be very short in comparison with the period of the alternating voltage on the grids. If this is not so, the electron, while in the space between the sides of the resonator, will be acted upon by changing values of the a.c. voltage and its velocity may be altered in the wrong direction. All this, of course, is considered in conjunction with the workings of the Klystron.

To make sure that the electron spends only a short time in the cavity resonator space, the center section of the resonator is

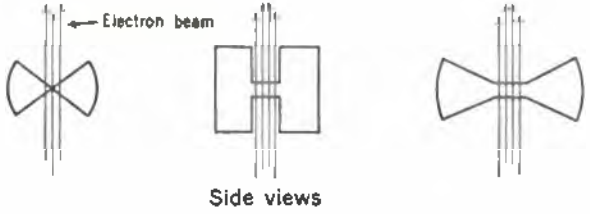


FIG. 6.12. Three additional shapes for cavity resonators.

reduced in size by placing the sides closer together. On either side of this section, however, the distance between the walls is made greater. It is entirely due to the electron transit time that this change in construction is made. There are other shapes which the cavity resonator can assume, as in Fig. 6.12. Note again that in each case the resonator portion which allows the passage of electrons has been made quite narrow. Then very little time will be spent in going from one side to the other. In the foregoing, only electrons have been given as the exciters of the resonator and this phenomenon has been described in detail. It has probably occurred to the reader that if a hole or opening of

some sort is made on the side of the chamber and an antenna inserted, energy in this radiator may also start the oscillations (see Fig. 6.6). This method is used when energy is fed from the catcher to the buncher grids in an oscillator. Although the cavity resonator has been illustrated with the Klystron oscillator only, there is no reason why it cannot be used with magnetron generators.

Output from Cavity Resonators. Output from a cavity resonator which is excited by means of the above mentioned electron stream can be accomplished in three ways. Surprisingly enough, each method has the same action for the ultra-high frequencies as for an ordinary low-frequency transmitter. The most obvious method is to allow the energy in the cavity resonator (or even a coil and condenser) to radiate from the tuning apparatus itself. This, of course, is usually undesirable, but nonetheless offers a method for the transfer of energy from the tuned circuit (or resonant chamber) to some outside receiver.

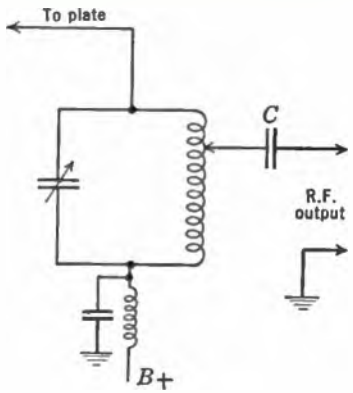


FIG. 6.13. Low-frequency method of obtaining output energy by capacitive coupling.

Capacitive Coupling for Output Energy. The next method of obtaining energy from a cavity resonator may be compared to the capacitive method used on low-frequency circuits and illustrated in Fig. 6.13. In order that a maximum of energy may be transferred, the wire from the condenser C is placed at a point on the coil where the r.f. voltage is high. Moving the tap up or down will either increase or decrease the amount of energy taken off. To apply this to a cavity resonator, it is first necessary to find a point of high value in the electric field, this being analogous to a high r.f. voltage. From what has been mentioned before, a high intensity electric field would not be found near any of the conducting walls, but rather toward the center of the cavity resonator.

To illustrate, refer to Fig. 6.7 and 6.8 where the electric field distribution for both the width and length of a cavity resonator is shown. Any point where there is large r.f. voltage, as shown by the arrowed lines, may be used for the output device.

The device itself that is used to pick up the ultra-high frequency energy is shown in Fig. 6.14. Here, two views are given, one (*A*), showing the small antenna separately and the other (*B*), where it is actually connected to the cavity resonator. To make this type of probe, all that is needed is a coaxial transmission line with the inner conductor extending a little beyond the outer cylindrical conductor. The hole in the cavity resonator through

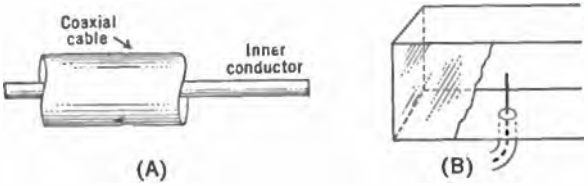


FIG. 6.14. (A) Extended inner conductor acts as capacitive coupler to electric field in cavity resonators and wave guides. (B) Actual position in resonator.

which this cable passes is usually threaded so as to enable the outer conductor of the coaxial line to be screwed firmly into place and at the same time make contact with the inner surface of the cavity chamber. The inner coaxial rod then extends well into the chamber itself and so comes in contact with the electrical field. The electrons in the center rod will be acted on by the electric field and thus energy will be transferred from the resonator. The other end of the connecting coaxial line may be attached either to a suitable wave guide or to some antenna, whereupon the energy will be transmitted out into space. In an oscillator, if the cavity resonator happens to be the catcher, the energy may be fed back to the buncher resonator and make the oscillating action complete. In this case another coaxial line may be attached to the catcher resonator for output energy.

In the low-frequency capacitive coupling method shown in Fig. 6.13, less energy was taken to the output circuit if the tap was moved down the coil. To accomplish the same results in the

case of a cavity resonator, all that is necessary is that the probe be moved to a region where the electric lines of force are less intense, which would mean bringing it closer to one of the walls and away from the center. This requires that another hole be drilled so that the coaxial line may be inserted. But, regardless as to whether the process of changing the energy output is more or less difficult at the ultra-high than at the lower frequencies, the comparison is quite evident. In one case there is an actual voltage being dealt with, in the other, an electric field of force.

Inductive Coupling for Output Energy. There remains one more way of coupling at the low frequencies, namely, the inductive method. Placing a small coil close to an oscillator coil will, as shown in Fig. 6.15, transfer energy from the tank circuit on the left to the output circuit on the right. Moving the small coil closer will result, up to a certain point, in more energy being available at the antenna. Increasing the distance between the coils will naturally produce the opposite effect, the well known action at the low frequencies. For the high-frequency analogy, magnetic lines of force will have to be considered. These form closed circles within the resonator and at the same time are everywhere at right angles to the electric fields. This fact has been mentioned in previous chapters and is referred to here to bring these thoughts back to mind.

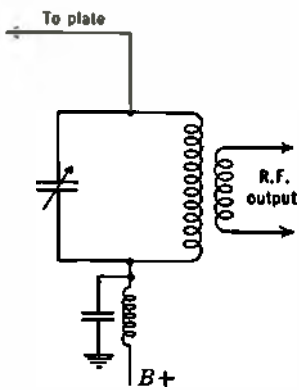


FIG. 6.15. Low-frequency method of obtaining energy by inductive coupling.

Now, in order for a changing magnetic field (such as might be obtained in a resonator) to induce a voltage in a metallic probe, best results will usually be obtained if this conductor is in the form of a loop. The more turns of wire in the loop, the greater the induced voltage. Experimentally this loop can be moved about in the cavity resonator until the point is reached where maximum coupling with the magnetic field occurs. By moving the coil

either toward or away from this position, less coupling may be obtained in much the same manner as is done on low-frequency equipment. A diagram for a magnetic probe that may be made from a coaxial cable is given in Fig. 6.16. Although it might be more desirable to use several loops, they are not so easily made to fit within the small volume of a cavity resonator. One loop, as pictured, will usually provide enough pickup for most purposes.

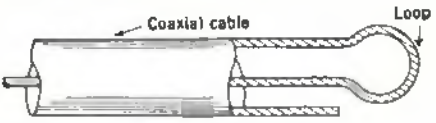


FIG. 6.16. High-frequency inductive coupler.

Q of Cavity Resonators. One reason why ordinary coils and condensers are not used at the ultra-high frequencies is due to the fact that the Q or selectivity (and efficiency) of these tuning units decreases to very small values at the higher frequencies. This has been explained in detail in a previous chapter. It might be interesting, however, to see just how the Q of a cavity resonator is obtained, since nowhere in its make-up can any lumped inductance (as such) be found. Of course, from what has been stated about transmission lines, it can be seen that a resonator comprises distributed resistance, capacitance, and inductance instead of lumped constants. Returning to the symbol Q itself, it was previously shown to be the ratio

$$Q = \frac{\text{inductive reactance}}{\text{resistance (a.c.)}}$$

or, stated in a simple mathematical formula,

$$Q = \frac{\omega L}{R}.$$

The above resistance is not the d.c. resistance as read on an ohmmeter but is the a.c. resistance. This increases with frequency for one or more of the following reasons:

1. Skin effect, which causes the current to be concentrated near the surface of the conductor as the frequency goes up.

2. Dielectric losses and leakage. At the high frequencies,

insulators often act as partial conductors. We are referring here to those substances that act as insulators at the low frequencies but lose this property at the very high frequencies.

3. Eddy currents in the conductors.

It is possible to express Q in slightly different terms meaning, however, the same thing. This follows from the fact that the energy put into the coil is divided into two parts: that which is stored in the magnetic field and that which is lost because of the resistance of the coil. It can be shown that the following ratio

$$\frac{\text{Energy stored per cycle}}{\text{Energy dissipated per cycle}}$$

is also equal to Q. Then we can deduce the fact that a coil which has a high Q is a more efficient device than one which has a lower Q.

To apply the above formula to a cavity resonator, it is necessary to know where the energy is stored and where it is lost. The first question is easy to answer since any energy contained in a

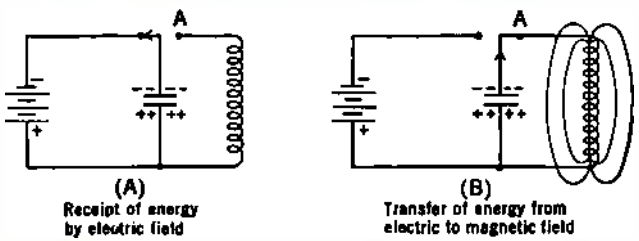


FIG. 6.17. Low-frequency resonant circuit showing how energy alternates between electric and magnetic fields.

cavity resonator can be found in the electric and magnetic fields. As previously mentioned, the energy is alternately stored in both. The comparison, at this point, between the old and the new is excellent and can be brought out by this very simple example. Take the circuit of a coil and condenser shown in Fig. 6.17. By means of a switch it is possible to charge the condenser to the same voltage as the battery. If the switch is opened, the condenser retains its charge. The energy may be considered as

residing in the electric field between the plates. Moving the switch to position *A*, the electrons will flow from the condenser to the coil and build up a magnetic field around the coil. Eventually, all the energy that was formerly lodged in the condenser is now found in the magnetic field around the coil. However, as soon as the current stops flowing, the magnetic field will collapse and force the electrons to continue to the other condenser plate. The condenser is now charged, but in the opposite direction. The process now repeats itself, only the electrons flow in the opposite direction. Thus it can be seen that here, too, at a frequency determined by the size of the coil and condenser, the energy changes from the electric to the magnetic field, just as in a cavity resonator.

There is only one place in a cavity resonator where energy is lost and this is in the confining walls. While the waves in a resonator continually move back and forth they strike the various sides and cause currents to flow. This represents an energy loss since the walls are not perfect conductors with zero resistance. The situation might be compared with the resistance in an ordinary coil and its connecting wires.

With the energy shown to be stored in the volume, or hollow portion of the resonator and the loss occurring at the various sides or surfaces of the resonator, the ratio

$$\frac{\text{Volume of resonator}}{\text{Surface area of resonator}}$$

should be proportional to the *Q* of the resonator. Thus a higher *Q* will be obtained if the Volume to Surface Ratio is large, which is the objective striven for in the designing of these units. Fundamentally, the definition and purpose of *Q* remain the same. However, as this unit is applied to new and different tuning circuits, the forms that go to make up the *Q* ratio may change, just as described above. That it can be altered and still mean the same thing shows at once the relationship between the old and the new and likewise speaks well for a system that is so flexible.

With the conclusion of the discussion on the *Q* of a cavity

resonator, this chapter is brought to a close. Only the basic theory has been covered and applied essentially to a simple rectangular resonator. Needless to say, innumerable forms of resonators may be used, all essentially the same except for the field distributions. To date, the mathematical theory has been applied only to a few of these complex shapes, so the experimental process of hit and miss is very much in vogue when designing these small units for ultra-high frequency oscillators.

Summary. Tuning units, whatever their form or shape, have essentially the same duties to perform. It is on the efficiency of this performance that their usefulness is evaluated. At the low frequencies, an ordinary coil and condenser combination will suffice. As the frequency increases, it becomes necessary to transfer from lumped electrical constants to distributed constants, such as are found in ordinary transmission lines. Generally, some odd multiple of a quarter-wave line is used to give the necessary resonance effects that we call tuning. At the still higher frequencies (the ultra-highs) it was found that hollow box tuners, called cavity resonators, became more suitable than transmission lines and so these were adopted. In each case the change-over was occasioned by the different effects brought about by the varying frequencies.

Cavity resonators may be considered as derived from two seemingly different electrical circuits. We may start with a transmission line and show that if enough of these lines are placed in parallel, we will finally end up with a cavity resonator. Or, we may go to a wave guide, and by suitable means, likewise derive our resonator. Whichever is used, the end result is the same. The new tuning unit that we have evolved is found to have a higher Q and, hence, a greater efficiency than any piece of tuning apparatus with which we were already familiar.

In exciting a resonator, the most widely adopted method involves the direct use of electrons as exciters. The electrons, as they pass through the resonator, induce charges in the walls of this unit, and these charges are forced to move along with the electrons until the electrons finally pass out through the opposite side

of the tuning circuit. Once released, the induced charges quickly flow back to their equilibrium positions. If the exciting electrons move through the cavity resonator at regular intervals, oscillations will be generated and sustained. Incidentally, it may be pointed out here that the electrons, in forcing the charges to move in the cavity resonator walls, give up energy in this process and this represents a transfer of energy from the electron beam to the cavity resonator.

Removal of this energy may be accomplished either by inductive or capacitive means. The first mentioned method requires a loop that is so situated in the cavity resonator that it will come in contact with the maximum number of magnetic lines of force. Moving the loop from this position will result in what might be termed looser coupling. For the second method, a straight length of antenna will suffice. This antenna, which is generally the inner conductor of a concentric cable, is moved about until it is parallel to the electric lines of force. This now represents capacitive coupling.

As a final consideration, we discussed the Q of a cavity resonator. Ordinarily this represents the ratio of the inductive reactance to the a.c. resistance of a coil. The same results, however, can be obtained if we use the ratio of power stored to power dissipated. In a resonator, the storage of electromagnetic power takes place within the resonator space itself, whereas for energy dissipation the confining walls must be considered. Using this as a basis, the Q of a resonator may be thought of as proportional to the ratio of the volume of the container to its surface area.

CHAPTER 7

U.H.F. ANTENNAS

A description of the various radiators that find use at the ultra-highs. Some are familiar, some are new

In the transmission of signals, there are two equally important factors. One is concerned with the generation of the signal itself, as discussed in the first portion of this book. The other, which will now be considered in detail, deals with the antenna systems that are our means of radiating the signal. A good generator with a poor antenna system is just as undesirable as an excellent set of radiators and a poor generator. Each one must be carefully designed for optimum results. In this chapter many of the systems used for micro-wave transmission will be considered with the main purpose of finding out just how this efficiency is attained. When setting up an antenna system or array, it is necessary first to determine whether the signal sent will be nondirectional in nature, or whether it is desired to concentrate its strength in one or more specialized directions. With broadcast stations, for example, the transmitting towers may be located either within the city itself or somewhere in the suburbs. In each case, the particular location will determine just how much directivity will be sought. This is further dependent on how the listening population is arranged.

Simple Antennas. For a nondirectional pattern, it might be advisable to use a vertical antenna, as shown in Fig. 7.1A. Its radiation pattern, plotted on circular coordinate paper, is shown in Fig. 7.1B. This type of graph paper is used to show the amount of energy radiated at various angles with reference to the antenna. If the field strength at each angle is plotted, and then all these points are connected, a continuous curve will result, showing how

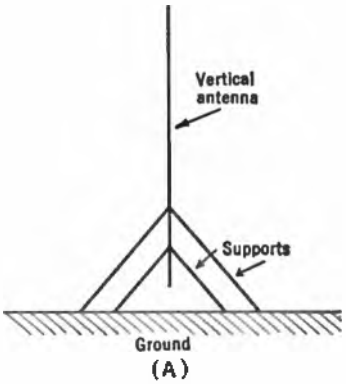


FIG. 7.1(A). A vertical antenna (side view).

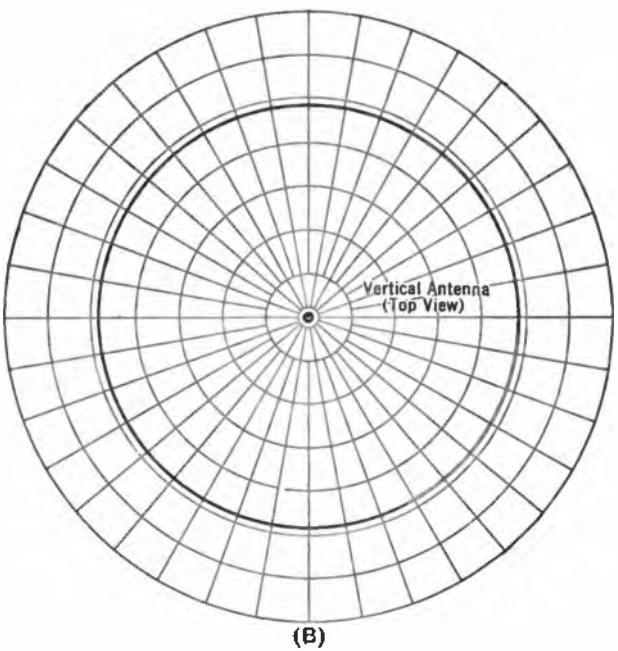


FIG. 7.1(B). The signal strength pattern of a vertical antenna. Notice that it remains constant at all points about this antenna (in the horizontal plane).

the strength of the transmitted or received signal varies at all points about the antenna. Consider, as an example, the curve in Fig. 7.1B. This shows that for a vertical antenna, the signal strength, at least theoretically, is constant at every point equally distant from the radiator. The distance from the antenna for each field strength measurement at each angle was the same. Hence, it is seen that vertical antennas tend to be nondirectional in a horizontal plane. A horizontal plane, according to usage here, refers to any plane which is parallel to the ground. Strictly speaking, this type of horizontal plane could not be flat since the earth's surface itself is not flat. However, for practical purposes, it will be considered as such. The vertical plane is at right angles to the horizontal plane and hence deals with the energy radiated in the skyward direction.

Now consider an antenna that is placed parallel with the ground, a horizontal antenna. Fig. 7.2 shows such a set-up, as seen from above, together with its radiation pattern. Here, in the horizontal plane, there is greater signal strength in the direction at right angles to the antenna, or broadside, as it is called. As the angle made with the center line AA' increases, the signal strength decreases, until at 90°, at one end of this antenna, there is no signal strength or radiation. This, then, is a directional antenna since it tends to favor the broadside direction and radiates nothing off the ends.

The radiation patterns, so far, have dealt only with the signal strength received or transmitted in the horizontal plane. As will be shown in Chapter 9, this is due largely to the direct or ground wave. But it should be pointed out at this place that these antennas also send radio waves upward, toward the sky. For long distance communication purposes, extensive use is made of the sky wave (see Chapter 9). Hence radiation patterns are also desirable showing the variation in signal strength at various angles in a vertical plane. Referring to Fig. 7.3, imagine that you were standing on the ground looking at the side of a horizontal antenna. The circle would represent the signal

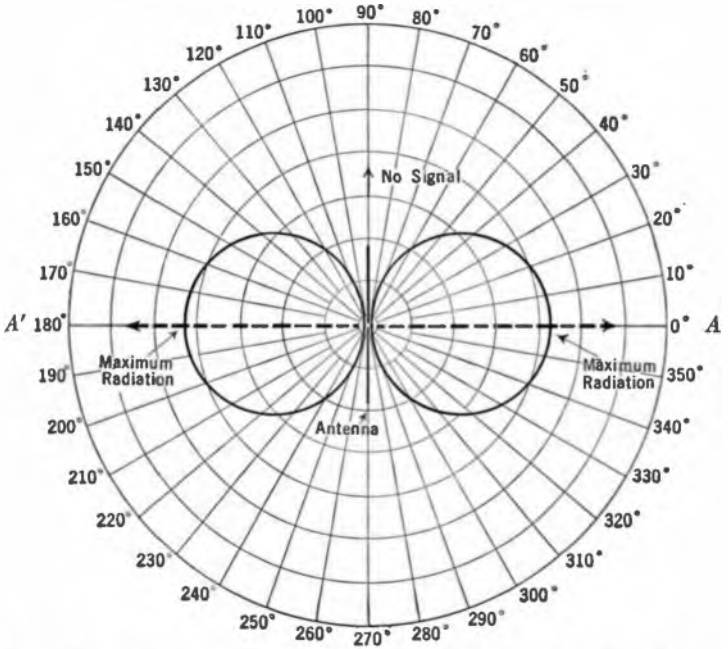


FIG. 7.2. Radiation pattern of a horizontal antenna as seen from above.

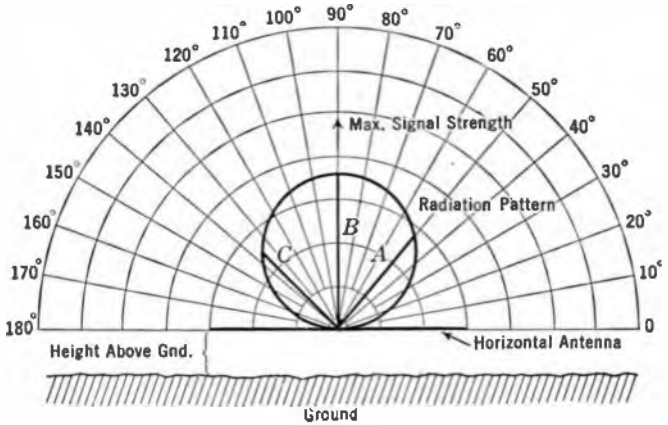


FIG. 7.3. Side view of a horizontal antenna, placed above the ground. The radiation pattern shows how the signal strength varies in the vertical (or upward) direction.

strength at various angles above the wire, with the center of the wire as the origin. Line *A* shows the strength of signal which might be expected at a 45° angle, line *B* at a 90° angle, and line *C* for a 135° angle. Maximum radiation in the upward direction is at right angles, or broadside, to the wire.

Since, as shown in Chapter 9, there is but little useful sky wave propagation at the ultra-high frequencies, patterns depicting signal radiation in the vertical plane will be omitted. Any pattern shown from now on, unless otherwise stated, will refer to the type of signal laid down in the horizontal plane, parallel with the ground. When reading these radiation patterns, it must always be remembered that the type of ground below the antenna will considerably influence and modify the radiation graph obtained for this radiator. For idealized curves, a fictitious, perfectly conducting and reflecting earth is assumed. Actual conditions, however, may differ substantially from this perfect ground. This results in actual patterns that are not of the same shape as the theoretical curves. However, for many locations, it will be found that the actual patterns agree quite closely with the graphs derived from formulas and warrants the continued use of these idealized patterns.

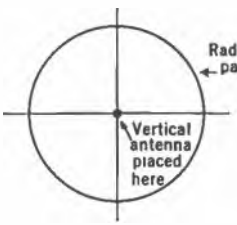


FIG. 7.4. Radiation pattern of a vertical antenna without the various angles shown. Compare this with Fig. 1B.

Another practice, widely used, shows the radiation patterns without the background of angles. A diagram of this type is given in Fig. 7.4 for a vertical antenna. Very little confusion is brought about by this method, since it is simple enough to draw in the necessary angles, if needed. Both this latter type of diagram and the previously mentioned graph will be used for these patterns throughout this chapter.

Antenna Arrays. The two simple types of antennas, the vertical and horizontal, may be used in various combinations to give rise to directive patterns, wherein the signal strength in one

direction is increased and in some other direction kept down to small values, possibly zero. These systems are known as arrays and are of two general classifications, the driven array and the parasitic array. In the driven array, each individual antenna comprising the combination is separately excited whereas, in the parasitic array, only one element is connected to the transmitter, the others receiving any currents they may have by pick-up of some of the radiated energy from the single excited antenna. The parasitic array is the most common at the ultra-high frequencies and so will be the only one described.

Directional systems or arrays are very widely used instead of the simpler vertical or horizontal antenna because of the increased concentration in one or two directions which they give to the radiated energy. This gain is derived from the decreased radiation at other angles to the system. It does for a signal what a megaphone does for the voice. When speaking with normal loudness into a megaphone, it can be noticed that the forward projected sound has greater intensity because of the addition of sound energy that would not ordinarily go forward.

Gain. Gain in itself has very little meaning unless compared with some standard; in this case a half-wave antenna that is placed at the same point in space as our directional array. Thus, a gain of 10 could mean that it would be necessary to put 10 times as much power in the half-wave antenna as in the directional array in order to lay down the same strength signal at some distant point. Or, the power in the standard antenna could be kept constant and the power in the directional system lowered $\frac{1}{10}$, whereupon the same signal strength at the distant receiving point would be obtained. The definitions are equivalent to each other.

Many times, the decibel is used as the unit of gain. In this case, since decibels are defined by

$$\text{db.} = 10 \log_{10} \frac{P_1}{P_2},$$

a ratio of P_1 to P_2 of 100 would give the equivalent decibel rating

of 20. This follows from the fact that the $\log_{10} 100 = 2$, and 2×10 equals 20 db.

Directivity. In speaking of the directivity of a directional antenna, the sharpness with which the signal strength is concentrated in any one particular direction is meant. It may, in a sense, be compared with the selectivity of a receiver in allowing one signal to pass through to the exclusion of all others. The more selective the set, the sharper and more peaked its tuning curve will be. In radio sets, sharp selectivity can usually be attained only after several tuning circuits are used in conjunction with each other. One coil and condenser combination, by itself, would not be adequate. It is much the same with antennas. Ordinarily several radiators must be used before a highly directive pattern is obtained. Just one or two elements, by themselves, might show definite directive effects, but these would not be as clear cut as can be obtained if a greater number of antennas are used. Finally, it might be mentioned in passing that, if a system of wires shows a marked directional pattern for sending, these same antennas would show similar directivity when used for reception.

Parasitic Arrays. It was previously mentioned, in defining parasitic arrays, that this system contains only one driven or excited antenna, all other wires receiving their energy from the radiated waves given off by the driven element. The parasitic wires are placed parallel to and some distance away from the exciting antenna. They receive their energy from the driven antenna and, due to the currents set up by the induced e.m.f., likewise give rise to their own radiations. Since the parasitic antennas are located some distance away from the driven antenna, a phase difference will exist between the two radiations. In some directions the waves will add up, while at other angles partial or total neutralization will occur. When the radiated energy is plotted completely around the system, a typical directional pattern will be obtained.

The phase difference that exists between the current in the parasitic element and that which flows in the driven antenna will

depend on two obvious factors:

1. The length of the parasitic antenna.
2. Distance between the two wires.

The length will determine whether the parasitic antenna is above, below, or at the resonant frequency of the wave being intercepted. As in any ordinary tuning circuit containing capacitance, inductance, and resistance, the current flowing will have a definite relationship with the induced e.m.f., depending on how much the inductive reactance neutralizes the capacitive reactance in the wire. At resonance, of course, only resistance is present. The influence which the distance between the wires plays in determining the phase difference is easily seen. It is directly related to the amount of change which occurs at the driven antenna during the time the energy is traveling the intervening space between the wires. The two factors of length and spacing will determine the directivity of the antenna array and, in addition, the gain in field strength obtainable at some distant receiving point.

Reflector Wire. Consider, for example, the simple type of parasitic array shown in Fig. 7.5 where a half-wave antenna is fed energy from a transmitter. The parasitic element or reflector is placed behind and parallel to the driven antenna. It has been found through experimentation that when the distance between these two wires is 0.20 wavelength, a directional pattern as shown in Fig. 7.6 will be obtained. The radiation is almost wholly in the forward direction, very little appearing behind the reflector. Here there is found a great difference between the driven array and the parasitic combination. Generally, the former has a bi-direc-

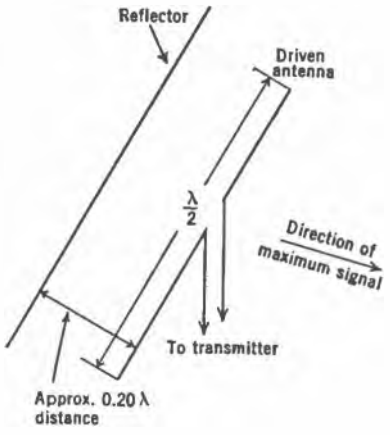


FIG. 7.5. A simple parasitic array containing one driven antenna and one parasitic element.

tional pattern whereas the latter usually has a uni-directional character.

For greatest gain, the length of the parasitic element should be slightly longer than that of the half-wave driven antenna. Unfortunately, the condition for greatest attenuation of the signal strength in the backward direction is not exactly the same as for

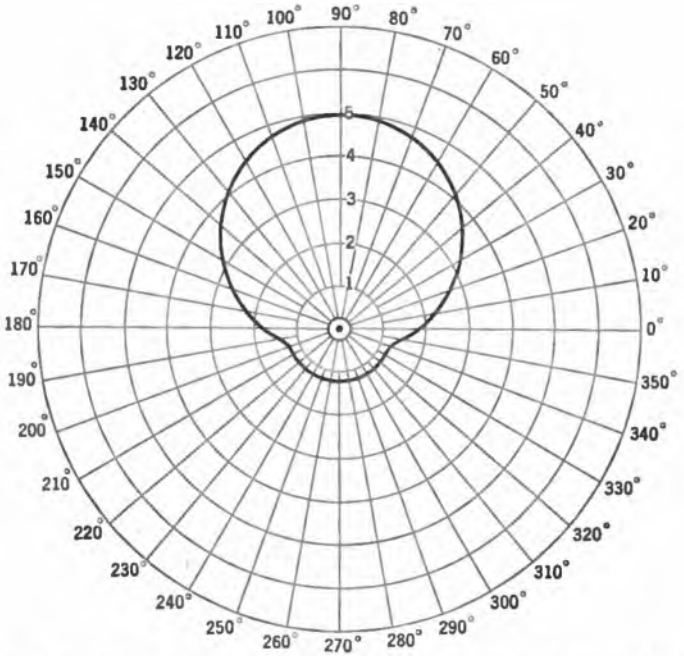


FIG. 7.6. An idealized radiation pattern of the array shown in Fig. 7.5.

maximum gain in the forward direction. To reduce the backward signal to a minimum, it is necessary to increase the length of the director beyond the value mentioned above for maximum forward gain. This increase in length, while aiding the attenuation of any back signal, at the same time reduces the forward gain somewhat. However, for many television or F-M antennas, a clear signal from one direction is more desirable than extra gain. This is because tubes and amplifiers can be added easily to the

set to give the requisite gain. For transmitting purposes, it may likewise be more important to eliminate signal transmission in certain directions than to have maximum gain. Here, again, the maximum attenuation conditions are of greater interest than those of maximum gain.

Directors. The uni-directional effect may also be obtained if, instead of using a wire of slightly longer length behind the driven element, a parasitic antenna is placed in front of (or in the forward direction from) the excited element. This antenna is called a director. Its length is slightly less than that of the driven antenna. The spacing between the two wires in this system is generally restricted to 0.1 wavelength and is more critical than the reflector distance. For maximum backward attenuation, the director is made even smaller in length, at least for spacings of 0.1 wavelength or more. A natural question may now arise in the mind of the reader as to which would be more desirable for a two-wire system, a reflector or director. Results from numerous experiments indicate that better gain is obtained when the parasitic wire is used as a director rather than as a reflector.

The method of adjusting a system of this type is usually the cut and try method rather than any other. For transmission, set up the reflector or director at the spacings given above and then, with some distant detecting device, adjust the length of the parasitic element until the greatest signal strength is received. For reception, the procedure is exactly the same, with the receiver now tuned to some distant station. One great advantage of these parasitic arrays lies in their flexibility. Generally they are mounted on some rotatable device which enables the operator to sit in his room and, by motor control, have the system face in any desired direction. In this antenna system only one pair of feeders, or transmission line, is needed to excite the driven antenna. For a driven array, where all the elements receive power directly from the transmitter, a more complex feeding system is necessary, which reduces the flexibility of the array and restricts its rotatability.

Since the excited antenna in this parasitic array is generally a

half wavelength long, any of the methods used for feeding this type of antenna may be applied here. A variety of such methods is given in Fig. 7.7. It is difficult to state which method finds

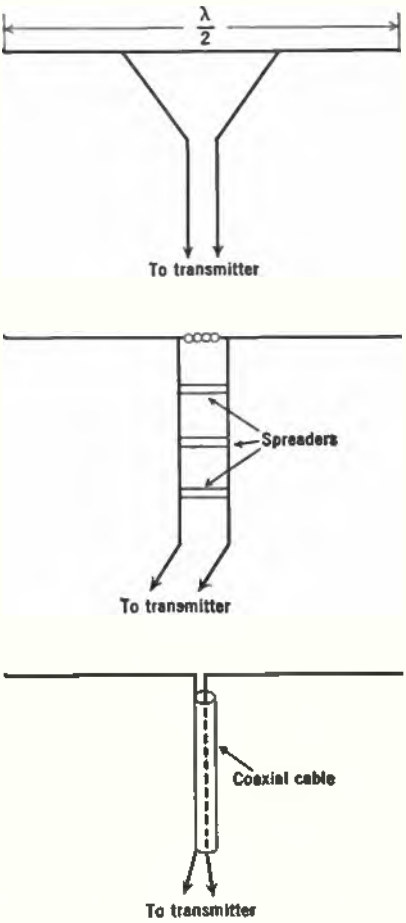


FIG. 7.7. Various methods of feeding antennas.

widest use, the requirements differing with each location. These antenna systems may be placed many feet in the air, and under these circumstances simple feeders would be favored. Complex transmitting line combinations require too much care and adjust-

ment and add to the problem since the tall structures are not easily accessible.

Several Parasitic Elements. For increased directivity and gain, several reflectors and directors may be used, as shown in Fig. 7.8. The distances to be used between elements are the same as given above, i.e., 0.1 wavelength between directors and 0.20 wavelength between reflectors. As before, parasitic elements behind the driven antenna are made longer while those in front are made shorter than the excited antenna. The frequency

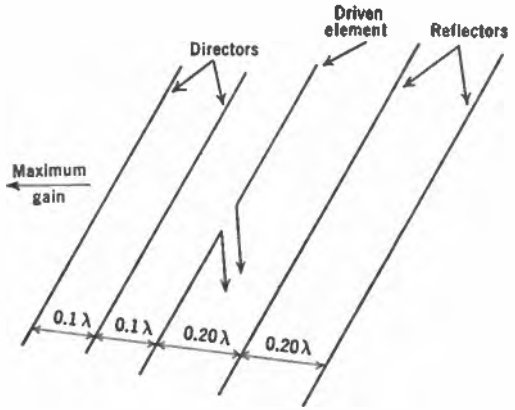


FIG. 7.8. A parasitic array containing several elements. Reflectors are longer and directors shorter than the driven element.

characteristics of this type of antenna array are more selective than those of two-element arrays. If used for either transmission or reception, they will tend to respond to a much smaller range of frequencies confined to a narrower direction. The last reason, together with decreased ease in construction, generally limits the number of wires used to 4 or 5. Beyond this, it proves easier to utilize some of the other directive systems soon to be described. When the foregoing systems are to be used for transmission or reception over a band of frequencies, the array is tuned for maximum signal strength at the intermediate frequency of the range. The gain decreases the farther away we get from this center frequency, but generally this reduction is small for moderate band widths.

Instead of having the parasitic elements placed directly behind the excited antenna, an arrangement such as shown in Fig. 7.9 may be employed. Here, the driven element is placed at the focus of a parabola of wires, this position resulting in the greatest forward gain. The adjustment is not too critical whereas with the solid reflector, soon to be described, the requirements become more exacting.

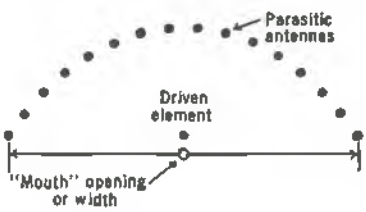


FIG. 7.9. A parabolic parasitic array.

Of interest in the design of this type of antenna array for maximum gain there is: (1) the effect of the number of reflectors placed behind the driven element; (2) the length of these reflectors; and (3) the width of the mouth opening (the distance *O* in Fig. 7.9). Fig. 7.10 shows the effect of the number of elements in the array while Fig. 7.11 shows the variation of gain with the length of these elements. For the first set of curves we see that the greater the number of elements, the more gain we obtain. The limit, of course, would be a sheet of metal, which would result in the optimum condition for variation of this factor. The second group of radiation patterns (Fig. 7.11) shows that the length should be increased for greater gain. However, as soon as the length becomes large in comparison with the wavelength transmitted, any further increase will not appreciably change the results. This type of array is not often used where the wavelength is large (lower frequencies), for here

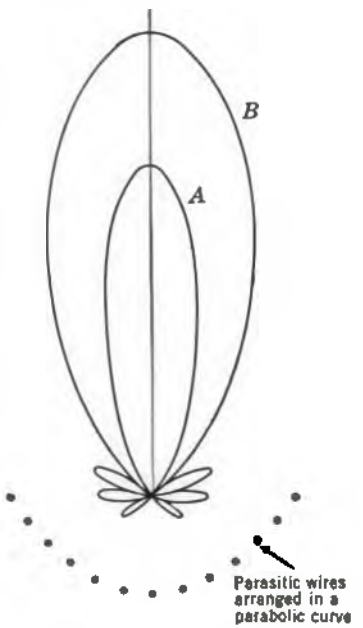


FIG. 7.10. Comparison of two parabolic parasitic antenna systems. (A) is the pattern for a system that has fewer parasitic elements than (B)

the lengths required would result in unwieldy elements, much too large for practical use. However, at the ultra-highs, in the absence of other types of radiating systems, the overall length needed is quite small and the total array is compact and easily moved. As to the width of the mouth of the parabolic system, the general rule states that there will be a narrower radiated beam from a smaller mouth width.

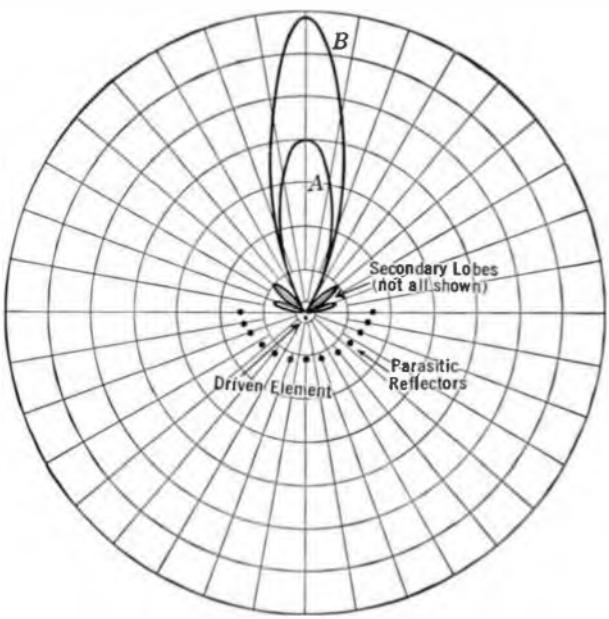


FIG. 7.11. The length of the parasitic reflectors is greater for curve *B* than for *A*. Extending the length of the reflectors beyond several wavelengths does not appreciably change the results.

Metallic Parabolic Reflectors. Metallic parabolic reflectors are not at all new. For many years, before anyone thought of using them for the concentration of radio waves, they were employed for light beams. Their action depends on the fact that, for reflection, the incident ray and the reflected ray make equal angles with any line that is drawn perpendicular to the surface at the point of reflection. To understand this, refer to Fig. 7.12A.

Here we have the ray AB hitting the parabolic surface at point A. Let the line N be perpendicular to the parabola at point A. Then, for reflection, angle α must equal angle ϕ. It will be found that the reflected ray AC will pass through the focal point F.

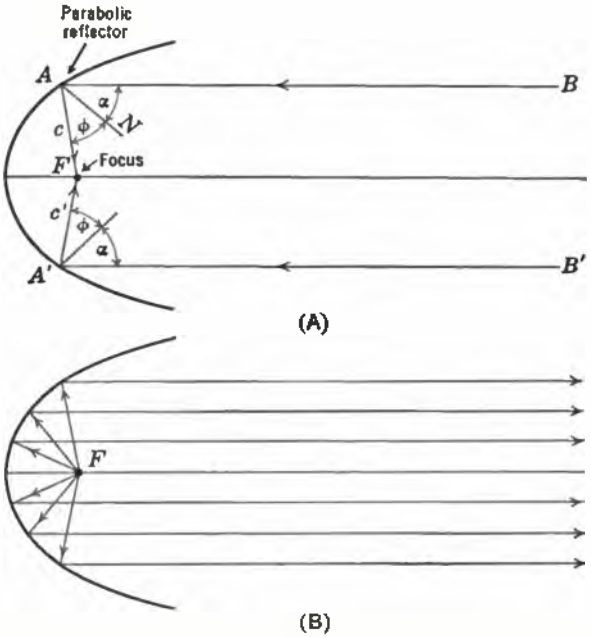


FIG. 7.12. (A) Action of a parabolic reflector in concentrating all parallel rays at its focus, F.

(B) Rays originating at the focal point, F, leave the parabolic surface in parallel rays.

It can be shown, too, by mathematical reasoning, that all parallel rays coming toward this parabolic surface will likewise meet at point F. Going the other way, any rays starting from F will hit the parabolic surface and travel outward in parallel lines, as in Fig. 7.12B. A spherical surface does not have this property. The parabolic reflector may be used either to concentrate a set of rays coming toward it or to produce a group of parallel rays. For ultra-high frequency radio, both are desirable.

It must be remembered that the concentrating action or

focusing only occurs when the rays arrive at the reflector along parallel lines. For all other angles of incidence, the rays do not

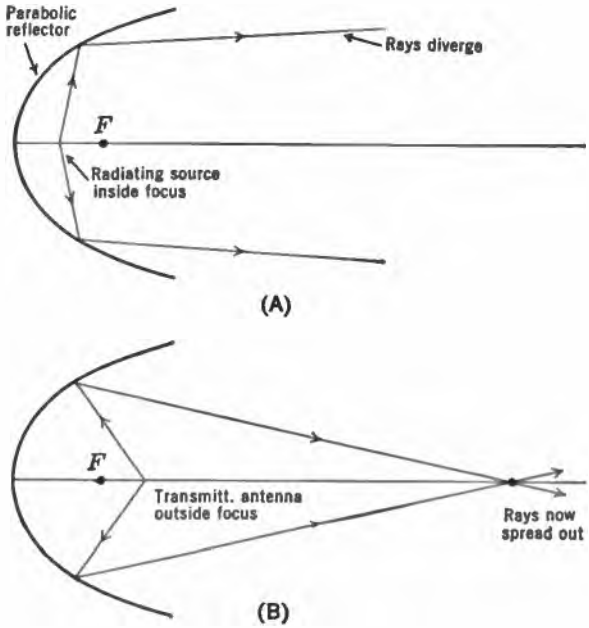


FIG. 7.13. Illustrations of what occurs when the antenna is not correctly placed at the focus, F, of a parabolic reflector.

pass through the focal point F. The reader might think this would impose a limitation on the unit, but experience has shown that radio waves coming from points 10 miles away arrive as essentially parallel rays. At greater distances the effect is even more positive.

For transmission, it is necessary that the antenna be placed at the focal point, F, and that the dimensions of the parabolic reflector be large in comparison with the wavelength used. The latter is one reason why these units can only find practical application at the ultra-high frequencies. The adjustment of the transmitting antenna is quite critical. If placed between the focal point F and the surface, it will give rise to rays that diverge or spread out after reflection from the parabolic surface. For dis-

tances greater than the focal point, the rays meet at some point P and then spread out. Both of these conditions are shown in Fig. 7.13, and both are generally undesirable. Hence, the need for a correctly adjusted antenna position.

Another difficulty encountered with the system just described is due to the observed fact that, when the antenna radiates energy, not all of it goes toward the reflector. Some of it travels directly forward from the antenna itself. To reduce this to a minimum, a semicylindrical cup may be placed behind the antenna as in Fig. 7.14 to cut off any direct radiation and force all the rays to hit the reflector before speeding forward. The cup must not be too large or it will intercept many of the reflected rays and lower the radiating properties of the system.

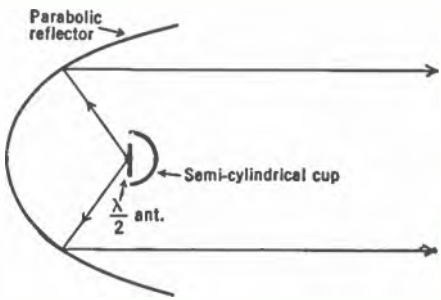


FIG. 7.14. The use of the semicylindrical cup prevents direct radiation from the antenna. All rays now come from the parabolic reflector.

It is perhaps due to the critical adjustments mentioned, coupled with a rather restricted frequency range, that other devices, soon to be mentioned, have been devised.

The Corner Reflector.[5] Instead of using curved surfaces as reflectors in a parasitic array, J. D. Kraus has suggested the use of two flat, solid, conducting sheets which are so placed as to intersect each other at some angle, forming a corner. He has called this type of reflector, shown in Fig. 7.15, the "corner reflector" antenna. The driven element, usually a dipole antenna, is placed at the center of this corner angle and at some distance, S, from the vertex of the angle. Fig. 7.15B shows the antenna as it could be used for the transmission of vertically polarized waves. By turning the system on its side, horizontally polarized waves could be sent (and received) as well. These two flat surfaces may form any angle up to the limiting value of 180°, at which position the reflector becomes a flat surface.

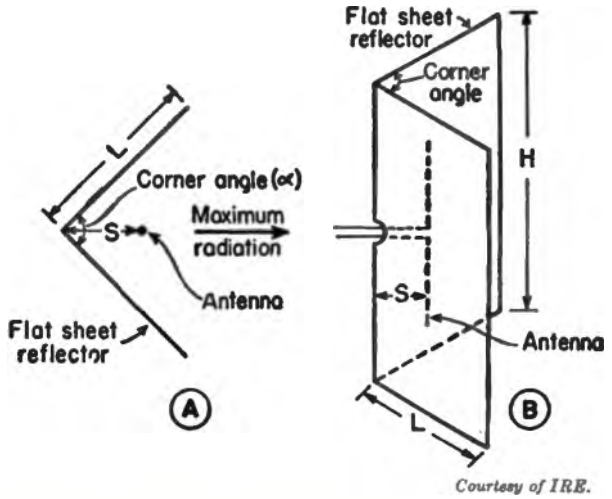


FIG. 7.15. Corner reflector in cross section (A) and in perspective (B).

The radiation pattern obtained for this reflector depends not only on the corner angle (greater gain with smaller angle) but also on the distance between the antenna and the vertex of the reflector corner. For relatively large distances (large compared to the wavelength used) a pattern containing more than one main lobe is obtained as, for example, Fig. 7.16B. As this is generally undesirable, there are limits for which maximum gain will be had with only one lobe in our directional pattern (Fig. 7.16). Kraus recommends the following limits:

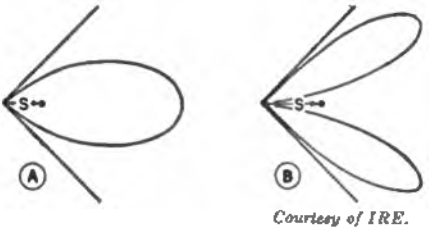
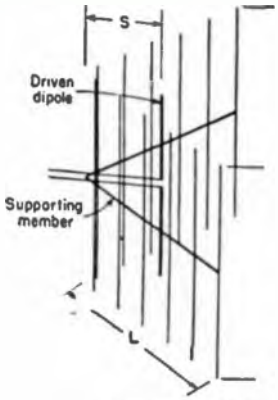


FIG. 7.16. Typical directional pattern for the corner reflector antenna (A). When S is large, a double-lobed pattern is obtained, as in (B).

Angle of Corner in degrees	*Limits of S*
180	0.1 –0.3 wavelength
90	0.25–0.7 wavelength
60	0.35–0.75 wavelength
45	0.5 –1.0 wavelength

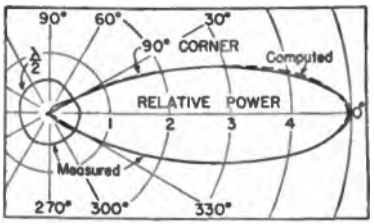
It can be noted from this table that none of the values are critical. This applies also to the dimensions of the reflector which may be used for a relatively broad frequency band.

Instead of using a solid sheet, a set of parasitic wires may be placed behind the driven dipole, as shown in Fig. 7.17A. Essentially the same results will be obtained using this grid-type corner reflector as would be obtained with a solid metallic sheet.



Courtesy of IRE.

FIG. 7.17A. Grid-type corner reflector having spaced parallel wires or conductors.



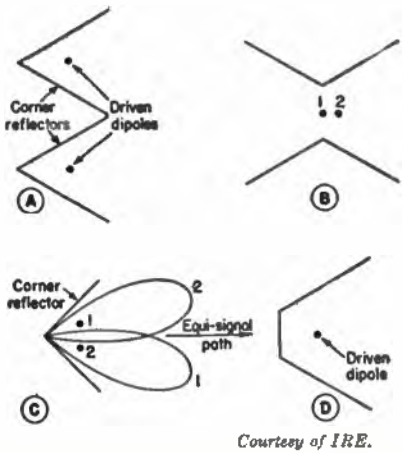
Courtesy of IRE.

FIG. 7.17B. The directional pattern of a square corner reflector. Included also, for comparison, is the pattern of a half-wave dipole antenna.

For an angle of 90° and a spacing, S, of 0.6 wavelength, the radiation pattern of Fig. 7.17B was plotted. The center circular pattern was obtained when the corner reflector was removed. Notice that, with this reflector, a gain of approximately 10 times was obtained over a half-wave antenna used without this aid.

At the ultra-high frequencies all the physical dimensions, both for the driven antenna and for the parasitic wires or sheets, are quite small, resulting in a compact, portable unit that is easily folded and transported from place to place. It is claimed that parabolic reflectors and electromagnetic horn radiators provide little or no improvement over corner reflectors of comparable size, and that the latter are less critical to adjust than the parabolic

system. Other suggested applications are given in Fig. 7.18. In Fig. 7.18A, two corner reflectors are combined in a multiple-unit structure, with both antennas fed in phase. At B, a bi-directional pattern is obtained. If the driven dipole is moved from position



Courtesy of IRE.

FIG. 7.18. Two-unit corner reflector (A) and a bi-directional type (B). Single corner reflector with two driven dipoles for producing double beams (C). At (D) we have a modified corner reflector with three sides.

1 to position 2, the right-hand direction would receive more of the signal than the left. For position 1, both directions, right and left, would be equally favored. At C in Fig. 7.18, with two antennas placed at equal distances from the central bisecting line, the radiation patterns shown will be obtained. One suggested use for this would be with radio beacons for landing aircraft. And, finally, Fig. 7.18D shows a modification of the two-sided reflector.

Metal Horns.[6] It was only natural for the men who first investigated the properties of wave guides (both cylindrical and rectangular) to wonder how efficient these devices might be as radiators. And it was just as natural to start with the usual types of wave guides, using them as they were, with the open end for the emission of the ultra-high frequency waves. In Fig. 7.19 is shown the type of radiation curve that may be obtained with a

cylindrical guide. A similar pattern may be plotted for a rectangular guide. The waves in these guides traveled along until they reached the open end where part of the energy continued out into space while part was reflected back along the guide.

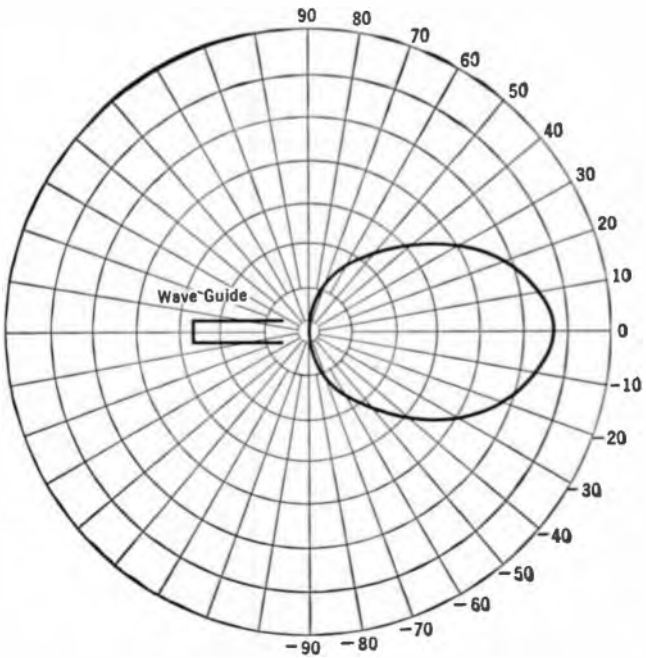


FIG. 7.19. The radiation pattern of a cylindrical wave guide. The waves were just emitted from the end of the guide. No flare was employed.

It was reasoned, however, that perhaps better results could be obtained if the transition from the wave guide to the open space was made less abrupt. To effect a slow transition, horns of varying diameter were resorted to, as shown in Fig. 7.20, for cylindrical wave guides. These horns actually are sections of wave guides, except that the diameter of the cylindrical forms increases gradually, i.e., flares. In the diagrams that follow, the various properties of these units, used as receivers, will be shown. However, the same relationships prevail for the transmission of waves,

so that we are imposing no limitation. A good transmitting radiator is likewise a good receiver of electromagnetic energy (for the same band of frequencies). Although circular horns will be used as illustrations, we will at the end of this section generalize

Courtesy of IRE.

FIG. 7.20. Various electromagnetic horns that were used in tests.

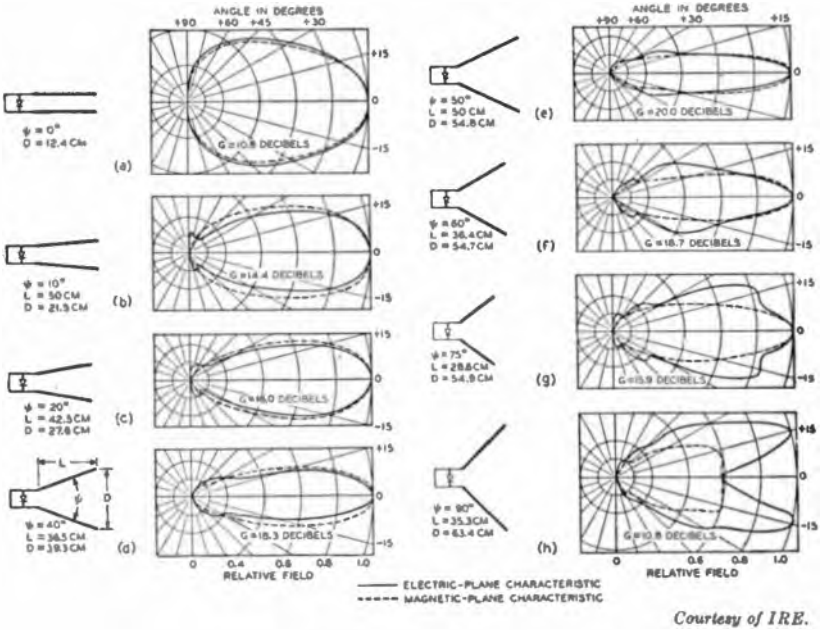
from the given diagrams to those that would be obtained if a so-called pyramidal horn is employed. The waves to be transmitted down the guides must, of course, be above the critical frequency.

In the investigation that followed the early work, three conditions were of general interest:

1. The type of radiation pattern obtainable with various lengths of the horn.
2. The variation of directivity and gain with change in flare angle.
3. The effect of frequency.

With patterns for these three factors worked out separately, it should then be relatively easy to arrive at conclusions that would enable us to obtain best results for any type of combination that might be used. The first set of diagrams in Fig. 7.21 show how the directional properties of a circular horn vary with change in the angular openings. For this set of readings, the length of each horn was kept approximately constant while the frequency used was 15.3 cm. Note that the greatest gain is obtained at flare angles close to 40°, with decreases above and below this value. For very large openings, the radiation pattern starts to form other small lobes. These are called secondary lobes. The same

result is found to occur with the rectangular horn and the sectoral horn soon to be described. The gain, as given in these diagrams, should have 2.15 deducted if we wish to compare them with a half-wave antenna. The data were taken relative to a nondirectional

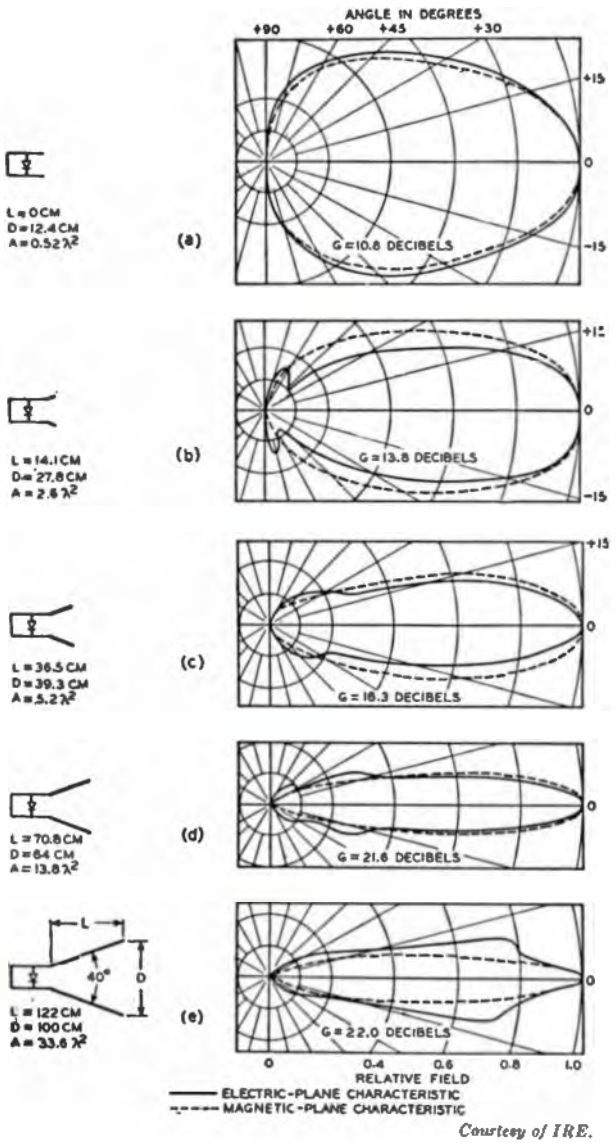


Courtesy of IRE.

FIG. 7.21. The directional properties of metal horns of approximately the same length but with different flare angles.

radiator. The half-wave antenna is 2.15 times more directional than the nondirectional element. Southworth calls this an absolute standard and suggests its use in place of the previous reference level. Gain is given in decibels.

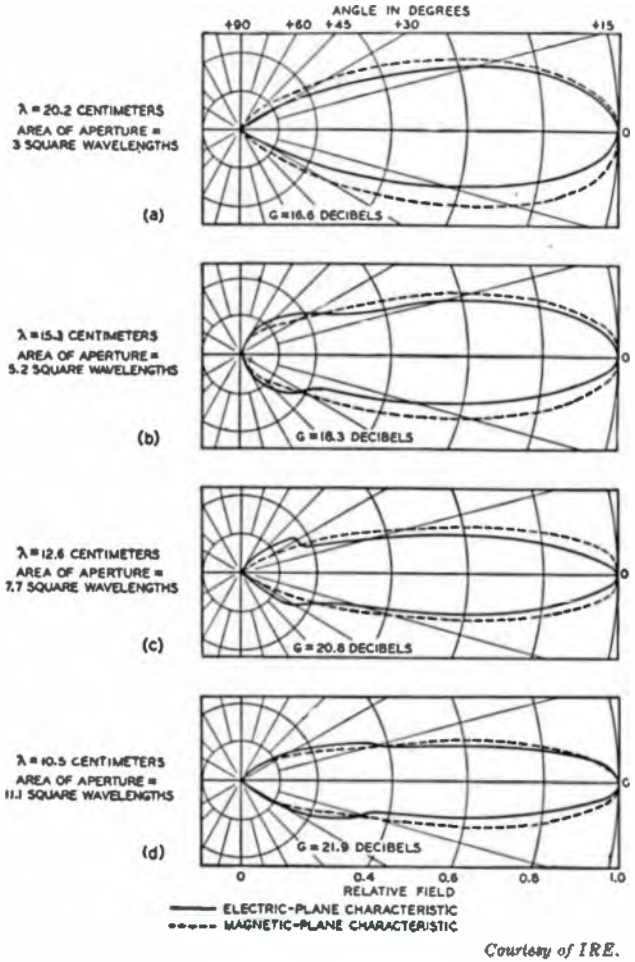
For the effect of length of the horn, with the angle set at the above optimum value of 40°, refer to Fig. 7.22. In the first four diagrams in this figure, the greater gain available with increase in length is shown quite distinctly. Still greater lengths, however, seem to show very little increase in gain and are accompanied by a lessening of the sharpness of the beam. The reason for this



Courtesy of IRE.

FIG. 7.22. These radiation patterns are for metal horns of the same flare angle but of different lengths. Wavelength used — 15.3 cm.

relates to the optimum angle, as tested in the previous set of patterns (Fig. 7.21). It should be pointed out that for each length



Courtesy of IRE.

FIG. 7.23. Directional properties of a metal horn as affected by changes in frequency (wavelength).

there is a "best angle," so to speak. For the lengths used in Fig. 7.21, the angle was found to be approximately 40°. In

general, as the length increases, the optimum angle needed becomes smaller so that, in the set of graphs given in Fig. 7.22, the radiation pattern improved with the length until the best length for that angle was reached. Beyond this, for improved results, the optimum angle should be decreased, which explains the fifth pattern in Fig. 7.22.

For the last condition, the only variable is the frequency; the flare angle and length of the horn are kept constant. Fig. 7.23 gives the results in this case and seems to indicate that the shorter the wavelength, the greater the gain. Actually, better results are obtained with the higher frequencies because, then, the relative dimensions of the horn increase. The length of a horn may be relatively small at one wavelength and yet quite large when compared to some higher frequency. When we stated that the gain rose with increase in length, we meant that the physical size of the horn was large in comparison with the wavelength used. A horn only 10 in. long might be a large unit if the frequency used is high enough.

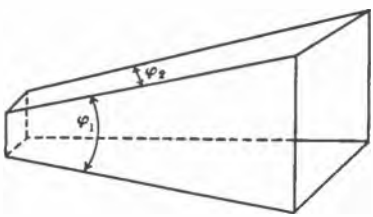
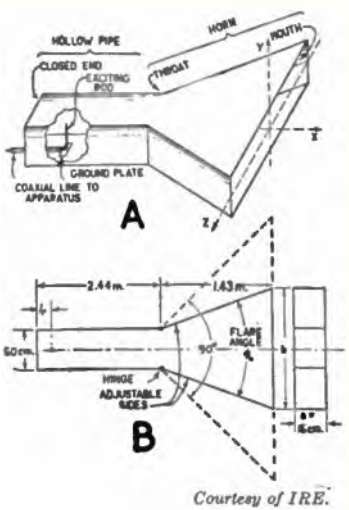


FIG. 7.24. A pyramidal horn showing the two ways the flare angle may vary.

Instead of conical horns, we might have performed these experiments with a rectangular wave guide and a pyramidal horn, as shown in Fig. 7.24. Note that here, when we talk about the flare of the horn, we may mean either one of two angles, ϕ_1, or ϕ_2, or both. And it follows naturally, from what has been presented before, that the resultant radiation patterns will depend on both these variables, in addition to the length of the unit itself. For equal angles, symmetrical radiation patterns are derived. On the other hand, increasing ϕ_2, for example, while ϕ_1 remained constant, gives a broader pattern in the horizontal plane. Likewise, if ϕ_1 is increased, a less well-defined beam is obtained in the vertical plane. The observations regarding length apply here, as well as to circular horns.

The Sectoral Electromagnetic Horn.[7] Instead of having the walls of an electromagnetic horn flare outward on all sides, it is possible to have only two walls flared, with the top and bottom walls parallel to each other. This type of horn, shown in Fig. 7.25, is called by its originators a " sectoral horn." While the top and bottom portions remain fixed, the two side walls may be arranged to form any angle between 0° and 90°. The angle is measured as shown in Fig. 7.25B, with the sides straight for an angle of zero degrees. The movable sides are hinged at the throat. The antenna may be placed close to the throat, as shown in this figure, in which case an adjustable backboard would probably have to be moved back and forth until the maximum intensity of radiation was obtained. However, there is no reason why this antenna could not be placed in a wave guide at some distance from the horn arrangement and the energy made to travel the length of the guide before it reached the attached horn. As before, the backboard tuning arrangement would still be behind the antenna. The backboard and the antenna may be considered as a directive array in itself. Certainly it is similar to some previously mentioned arrays, such as the corner reflector or the solid parabolic reflector. Its proper adjustment, which only takes a few minutes, is well worth the effort and may add several decibels to the wave.

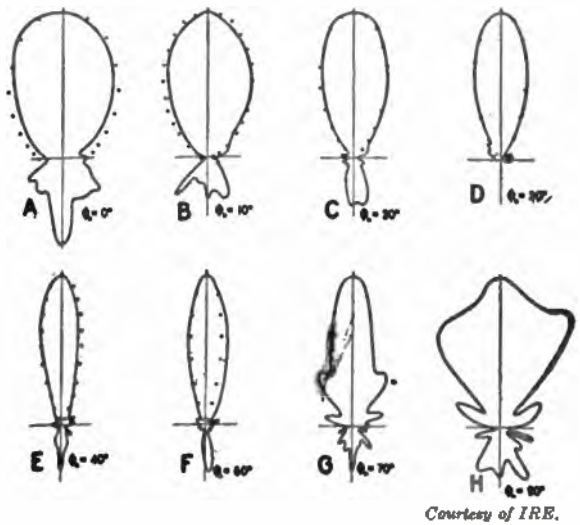


Courtesy of IRE.

FIG. 7.25. Perspective view (A) and cross-sectional view with dimensions in the x, z plane (B), of a sectoral electromagnetic horn.

Working with the sectoral electromagnetic horn, Barrow and Lewis obtained a set of radiation patterns like those in Fig. 7.26. The data were taken at the various angles indicated below each plot. It is seen that, as the flare angles increase up to the range of 40° to 60°, the radiated beam becomes sharper and more

defined. Above these angles, the sharpness is decreased and again there is the appearance of secondary lobes. In each of the above cases the mouth of the horn may be considered as placed at the origin of the curves, which is the place where the two straight lines intersect at right angles. With this in mind, it may seem strange to the reader that any radiation would occur in the backward direction, since this is behind the apparatus. Barrow and



Courtesy of IRE.

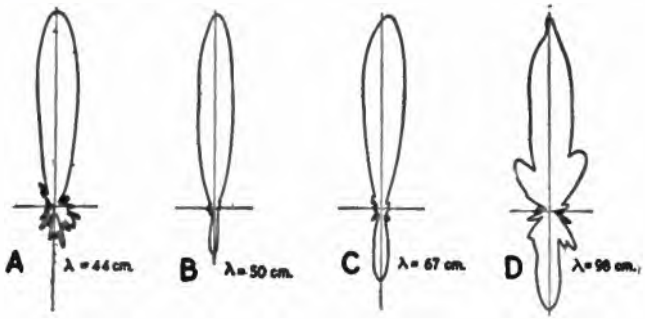
FIG. 7.26. The radiation patterns of the sectoral horn for the various flare angles noted.

Lewis investigated this effect and after experimentation came to the conclusion that it was due to diffraction, or bending of the waves around the mouth of the horn. With suitable precautions, it may be eliminated altogether.

With the flare angle kept at 40°, the next desired information pertained to the operation of this horn at various frequencies. The guide that was used had a critical wavelength of 100 cm. The results, using several different wavelengths (those given are 98, 67, 50, and 44 cm.) are shown in Fig. 7.27 and emphasize the useful fact that the horn may be employed over a wide band of frequencies without having its directive powers altered. The

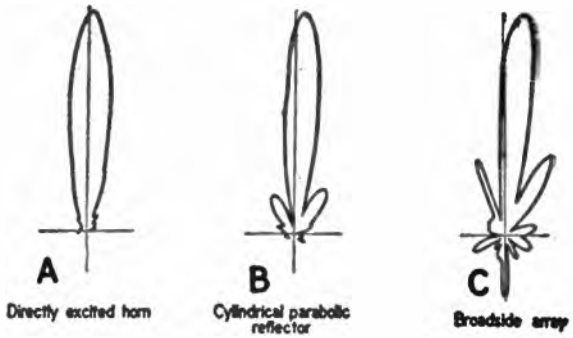
pattern changes appreciably only when the critical frequency of the wave guide is approached.

This type of radiating system has been compared with that obtained with a parabolic reflector. The results seem to indicate



Courtesy of IRE.

FIG. 7.27. The above directional patterns are for a sectoral horn with flare angle constant at 40°. Disregard the curves that go below the horizontal line as this signal is due to diffraction around the mouth of the horn. The frequencies used are given with each plot.



Courtesy of IRE.

FIG. 7.28. A comparison of the radiation patterns of three different types of antennas.

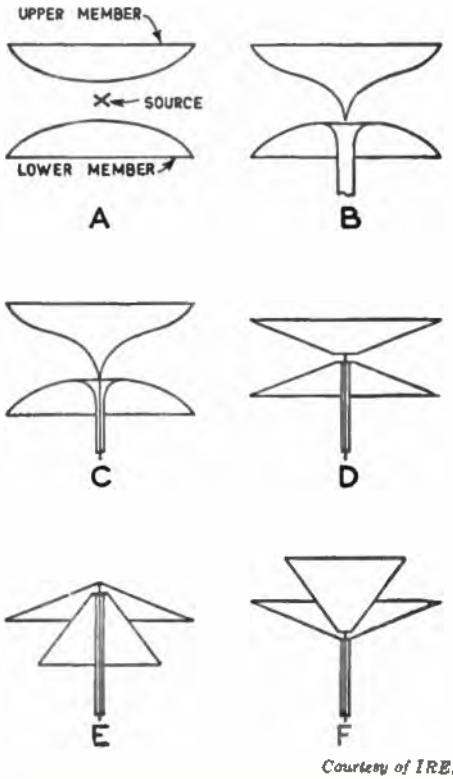
that the horn gives much better performance and is to be preferred. For example, comparing typical radiation patterns for each type, as given in Fig. 7.28, it will be immediately apparent that the one derived from the horn is devoid of any appreciable

secondary lobes, whereas they are evident in the parabolic radiation pattern. Furthermore, the parabolic system, as noted previously, is far more critical to wavelength variations than the corresponding horn and, even with any one frequency, requires careful adjustment before optimum operation is obtained. The antenna must be placed at the focus of a parabolic reflector for best results. Antenna arrays fare no better in comparison at these extremely short wavelengths. An array is seldom used at the very high frequencies, due in part to the difficulty of adjusting the system. The radiation patterns of the arrays usually give rise to appreciable secondary lobes and these can only be removed by adding more parasitic elements. At the longer wavelengths, however, the horn or parabolic reflector becomes too large to handle and so the antenna array must be employed. The same situation exists here as with transmission lines and wave guides. The method to be used depends upon the ease with which it can be assembled and adjusted.

It is interesting to note that the radiated beam of the horn may be made sharper by lengthening the sides of the horn. For compactness, it has also been suggested that small transmitting units be placed in the short rectangular wave guide connected to the horn and in this way decrease the loss that is usually occasioned by transmission lines that are needed to transfer the power from the unit to the guide.

Biconical Electromagnetic Horns.[8] The patterns obtained with the directional apparatus considered so far concentrated the beam into a relatively narrow angle in one direction. There are times, however, when it is necessary to send out radiation in all directions, giving rise to a fairly uniform beam throughout the entire 360°. This might be true when it is desired to have the signal received by a large population scattered over a considerable area. One solution to the problem might be to use a vertical antenna, which has been shown to be nondirectional in the horizontal plane. There is another method, however, where the same nondirectional property may be obtained with the added advantage of additional gain. This type of system has been

described by Barrow, Chu, and Jansen. The results of their experiments will now be considered. They combined two electromagnetic horns in various ways, some of which are shown in Fig. 7.29. Because of the use of only two elements in this system, it has been given the name of " biconical electromagnetic horns."



Courtesy of IRE.

FIG. 7.29. Various combinations of biconical electromagnetic horns.

Note from the illustrations that the flaring of the horn may be gradual, as in B and C of Fig. 7.29, or much more sharp, as in illustration D. For E and F, one horn is partially placed within the other and for E the radiation would be in the downward direction. For F the beam is upward. The energy is led up to the horns by means of a coaxial cable. Between the horns, the

center conductor of the cable is extended. It is from this antenna that the radiated energy is obtained. The other end of the cable would be attached to a magnetron or Klystron generator.

In discussing the radiational patterns of biconical horns, many points of similarity will be found between these and the other types of horns previously examined. For example, just as with the sectoral horn, the sharpness of the beam of the biconical

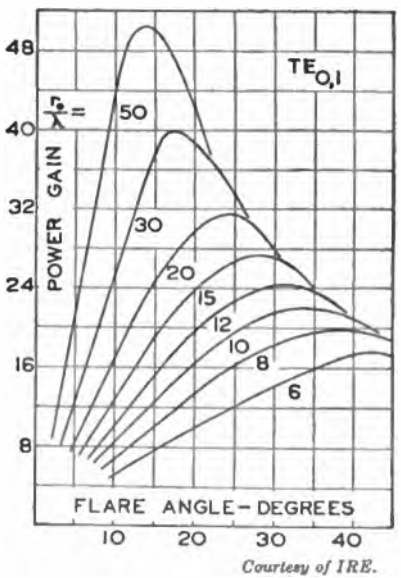


FIG. 7.30. Curves showing the variation of power gain (not in db.) with flare angle for various lengths of horn. Each number given with each curve represents the ratio of the length of the sides of the horn to the wavelength. The wavelength, of course, remains constant.

arrangement will depend on the length of the flared sides and the flare angle. As the length is increased, the optimum flare angle will decrease and, at the same time, the resultant beam will be sharpened. Also, if we keep the length constant, there is one angle which will give best results. An interesting diagram showing the relationship between flare angle and power gain is given in Fig. 7.30. Each curve is for a different length of horn.

When the length of the sides is large with respect to the wavelength used (in Fig. 7.30, this is a ratio of 50 for the top curve), a gain of approximately 50 times that of a half-wave dipole may be

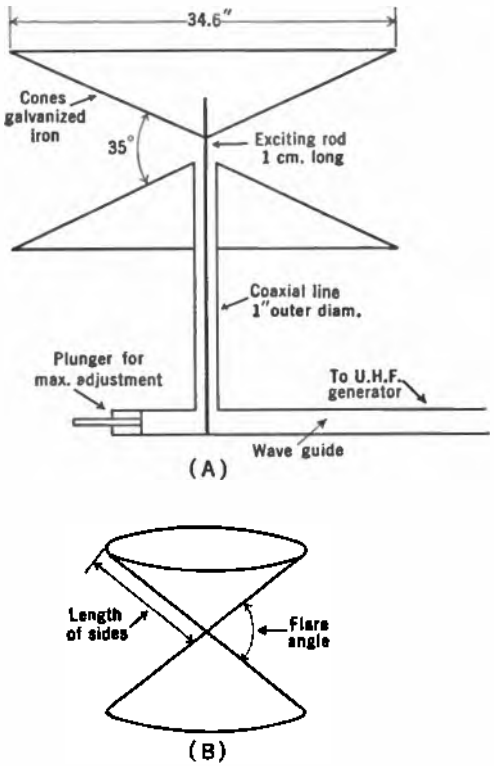


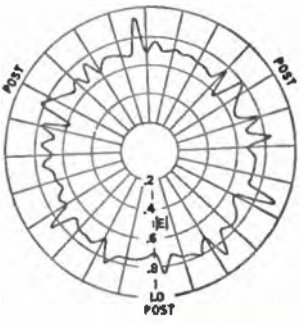
FIG. 7.31. (A) A diagram of the set-up used to obtain the results in this section. (B) A view of a biconical horn showing the important factors that determine its radiational properties.

obtained. As the sides become smaller (with the same transmitted frequency), the flare angle must increase and the gain decrease.

To obtain the foregoing information, a set-up like that shown in Fig. 7.31A and B was used. The general results obtained at some distance from the radiator are shown in Fig. 7.32. Note that the pattern is fairly circular, with slight variations introduced

by the support posts of the biconical horn and other extraneous surfaces connected with the radiating system, such as nearby objects. The signal energy sent upward may also be plotted, but is of little consequence here. Vertical angle radiation may be reduced by decreasing the flare angle. If set out in the open, a covering may be devised to protect the cones from the elements.

The reader may be struck by the resemblance between the conical horn and a similar arrangement suggested by Kraus in his corner reflector. Essentially these are the same except that the corner reflector, because of its structure, does not give rise to a radiated beam for all 360°. The biconical horn does. To accentuate the beam more in one direction than at any other angle, it is possible to move the apex of one horn (say, the top one) a little to one side. With the apex of the top horn moved to the left, increased signal strength is observed to the right.

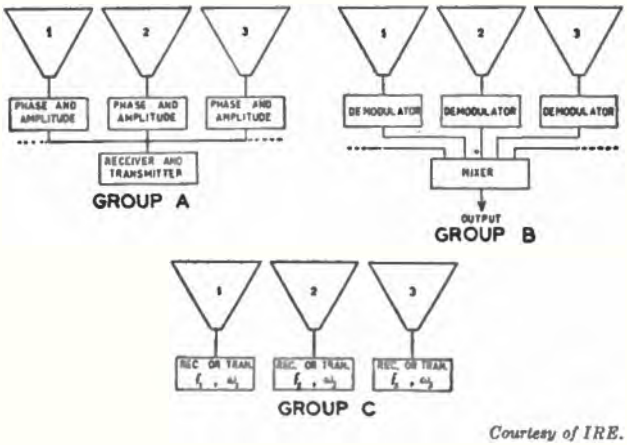


Courtesy of IRE.

FIG. 7.32. Radiation pattern of a biconical horn in the horizontal plane.

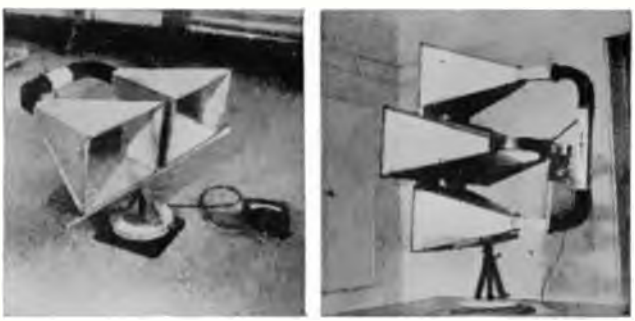
Multiunit Electromagnetic Horns.[9] We have shown several radiators for the ultra-high frequencies and have referred to their various useful characteristics. In this concluding section, more than one unit will be considered and an array of horns will be used. It can be called a multiunit radiator. These systems have been broadly classified into three groups, A, B, and C, as shown in Fig. 7.33. The horns may be stacked in almost numberless combinations, two of which are given in Fig. 7.34. Fig. 7.34A shows two units placed side by side; 7.34B, four units, one above, one below, and two alongside each other at the center. The horns used in these arrays are the previously mentioned pyramidal horns and differ from the sectoral horn in that all the sides flare, none being parallel to each other. Excitation may be accomplished either by hollow pipe conductors (guides) or concentric cable radiators placed in the neck connected to these horns.

In treating an array of this type, the characteristics of each element in the array are directly responsible for the final form of the overall radiational pattern of the group. If each horn does



Courtesy of IRE.

FIG. 7.33. Different combinations of multiunit horn arrays, showing how the signals leaving or entering each unit may be modified for a particular purpose. At (*C*), each horn responds to a separate frequency.



(A) (B)

Courtesy of IRE.

FIG. 7.34. Two possible multiunit arrays.

not directly affect the transmitting or receiving properties of any of the other units in the array, the signal field laid down by each will be maintained independently of the other horns. If this occurs, then it is said that the mutual impedance between the

horns is zero. This condition may be obtained if the length of each horn is large in comparison to the transmitted (or received) wavelength and if the horns are so placed as not to face each other directly, i.e., " mouth to mouth." This set-up is desirable when signals of different frequencies are to be transmitted from separate units in the array. When the arrays are closely stacked, each pattern will blend with that of the others to give a resultant signal that tends to become quite concentrated in a small angular distance. As the spacings between units increase, however, small secondary lobes become more prominent, resulting from the overlapping of the radiated fields.

The uses to which these multiunit assemblies can be put are legion, restricted only by the ingenuity of man himself. There is always a definite need for radiating mechanisms of this type in the aircraft installations of an airport and they will, no doubt, find wide usage there. The foregoing is meant to show only the underlying equality of many radiating systems and to aid the reader to understand any application to which they may be put.

Summary. At the ultra-high frequencies, dimensions of antennas and antenna systems decrease to such an extent that it becomes feasible to use such radiators as parabolic reflectors, corner reflectors, rectangular and conical horns, and multiunit arrays. Each has its peculiar properties and yet there is an underlying pattern that each system follows. For horns, it can be seen that sharper beams will result if the length of the sides of the horn is made large in comparison with the wavelength being transmitted. In addition, it was found that the best flare angle likewise decreases as the length of the horn becomes greater.

Electromagnetic horns have some desirable qualities which will probably result in their wide adoption. They are quite readily constructed and give almost identical results over a relatively wide frequency band. Adjustments on horns are not at all critical and can be accomplished in a short time. Generally, any position inside the throat of a horn may be used for antenna placement.

In discussing the gain that is possible with a directive array

as compared with one that has less directivity, two possible basic standards have been proposed. One is a half-wave antenna placed in the same position in space as the array whose gain is to be computed. The other is a totally nondirectional antenna. The difference between these two standards is approximately 2.15 db., this number referring to the gain of the half-wave standard over that of the nondirectional system. Generally, the average gain of a horn over a half-wave antenna is 20 db., which corresponds to a power increase of 100.

Most of the units described in this chapter give rise to unidirectional patterns. For broadcast coverage, a nondirectional antenna might be desirable. For this purpose the biconical horn radiators could be used. Note that, although this system is nondirectional in a horizontal plane, it is quite directional in the vertical plane. There is very little sky wave signal strength when this system is employed. This fact probably explains why the array gives greater gain than a nondirectional vertical antenna, as mentioned in the first few paragraphs of this chapter.

Finally, we have combinations of electromagnetic systems called multiunit horn arrays. If properly spaced, each unit will give rise to a signal pattern which is independent of the radiation of any of the other units in this array. By changing the phase of the various signals, almost any desired directional pattern can be obtained.

CHAPTER 8

U.H.F. MEASUREMENTS

The newer methods of measuring voltage, power, and wavelength

Factors Governing U.H.F. Measurements. One reason that radio science advanced as rapidly as it did can be laid directly to the fact that excellent measuring equipment was always on hand. For, without proper meters to guide the engineer constantly and to show him success or failure, there would not or could not be the effective precision-type radios of today. The meters were developed for the low frequencies at which measurements of inductance, capacitance, and resistance do not change measurably with frequency and at which these quantities can be measured quickly and easily with high accuracy. But, as the frequency increases, so does the number of factors that must be accounted for, and soon we find a system of measurement that has lost its speed and accuracy. Meter readings that ordinarily could be taken at face value must be corrected for the many errors introduced through the peculiarities of the ultra-highs. As we study the errors that occur, it will become increasingly obvious why the conventional types of meters had to be either modified or completely discarded.

Some of the reasons that make it impossible to utilize low-frequency meters (in their conventional form) at the ultra-highs will now be reviewed. In the ideal case, when a meter is inserted in a circuit to measure current or voltage, the quantity indicated on the scale will be the true value, the presence of the meter having no effect on the circuit constants. To achieve this goal, all current indicating meters should have zero resistance and all voltage indicating devices should possess infinite resistance. The more the meters in actual practice deviate from these ideals, the less

accurate the results will be. At the low frequencies, good results can be obtained because these rules are approximately attained easily and quickly. With increase in frequency, no longer is it certain that our fundamental units (capacitance, resistance, and inductance) will keep their values constant. Naturally this will affect any voltage or current reading taken. Consider, for example, the resistance of a length of wire. At the low frequencies, currents flowing through this conductor do so in a uniform manner, each small section of wire carrying the same amount of current as any other similar section of wire. Increasing the frequency causes the currents to vary and to crowd toward the outer surface of the wire. This is the well known skin effect which results in an increase in the effective or a.c. resistance of the conductor.

Skin Effect. The reason for the increase in resistance can best be understood if we consider in greater detail just what occurs within a length of wire when a current flows through it. It is common knowledge that current flow has associated with it a magnetic field or, what is the same thing, circular magnetic lines of flux. These are everywhere encircling the current. The definition of inductance depends upon these flux linkages and is given by the formula

$$\text{Inductance (Henries)} = \frac{\text{Flux linkages encircling conductor}}{\text{Current producing these linkages (in amps.)}} \times 10^{-8}.$$

Consider now the end view of a small round section of wire that has a current flowing through it See Fig. 8.1. Each small section of current flowing through this wire has magnetic lines of flux encircling it, but the sections of current at the outer edges of the wire obviously have fewer lines of flux around them than the currents at the center of the wire. This is due to the fact that the flux produced inside the wire by the central currents does not encircle the outer currents and so cannot influence their flow. The flux produced by the currents at the surface of the wire does, however, encircle the currents at the center and hence exerts an

influence upon them. From the definition just given for inductance, it is seen that, since there are more flux linkages encircling the center of the wire than the outer surface of this wire, the inductance at the center will be greater than at the surface.

The consideration here is not the entire inductance of a piece of wire, but the small components that add up to this total. It is only in this way that a thorough understanding of the skin effect and its results can be obtained. Returning to the discussion

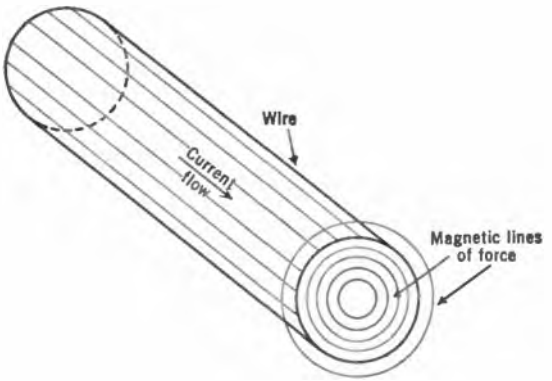


FIG. 8.1. A small section of a wire carrying a current.

above, as the frequency increases, the inductance at the center of the wire will offer more opposition (reactance) than the outer sections of the wire where there is less inductance. Hence, the current, seeking the path of " least resistance," will tend to concentrate more at the surface (or skin) of the conductor as the frequency rises. But now the current, which formerly spread uniformly throughout the entire area, concentrates near the surface. This has, in effect, reduced the useful cross-sectional area of the conductor. The resistance, due to this decrease in effective area, will rise, a fact generally accepted as the explanation of skin effect. If the wire is used as a resistor, its resistance will change as the frequency varies, increasing more and more as the frequency increases.

In addition, inductive and capacitive effects will also be included in this once pure resistor (at the low frequencies). The

latter is by far the greater of the two and arises from the close proximity of the conducting wires. After all, a condenser is by definition only two conductors separated by a dielectric. The equivalent circuit for a resistor at the higher frequencies is given in Fig. 8.2. It is possible to manufacture certain resistors so that the resistance is fairly constant up to high frequencies.

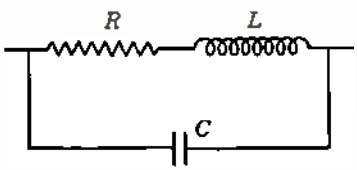


FIG. 8.2. Equivalent circuit of an ordinary resistor at the ultra-high frequencies.

Inductance and Capacitance. Inductance changes slightly from the low to the high frequencies, due again to the effect of flux linkages and the currents. As the frequency increases the distribution of the currents in the wire will change and, because inductance depends upon this distribution, it too will change. The result will be a decrease. It is true that the decrease in inductance is slight but still it is measurable. Here, then, is another factor that must be considered when using meters at the ultra-high frequencies.

Finally, there is the matter of capacitance. The capacitance of any circuit may change because of stray couplings (which also affect the inductance), couplings that might arise between adjacent wires, a wire and ground, points between which there exists a difference of potential, and in almost innumerable other ways that were never considered important before. Leakage conductance can also be included in connection with capacitance. This becomes quite large, many times, at the very high frequencies. Condensers and insulators may, because of certain conditions, allow small amounts of currents to flow through them. They then act as high resistances, or poor conductors. The conductance is low and is termed leakage conductance. This effect is quite prevalent on transmission lines. For a poorly constructed line it may rise to large values. Care must always be taken with so-called insulators which are to be installed in these high-frequency systems. They should first be tested for leakage at the short wavelengths.

The discussion so far is not a complete story but should enable the reader to gain a better idea of why our equipment must be modified to meet the new conditions. There may, in addition, be combinations of factors; for example, the diode voltmeter which will be described presently. Here the combination of interelectrode capacitance and lead inductance gives rise to a resonant circuit and introduces sizable errors in readings taken with this instrument at the ultra-high frequencies. Fortunately it is possible to make adjustments for these in this particular voltmeter. Furthermore, at the short wavelengths, current and voltage values of a small piece of wire in the circuit will change appreciably in a very small distance. This fact must be known before any measurements of these quantities can be undertaken intelligently. Such limitatons will be noted as the various instruments are described.

In this chapter only those instruments will be presented which fulfill the necessary requirements of usefulness and accuracy to a reasonable degree. The field is thus restricted to wavelength, voltage, and power measuring instruments. There are good methods available for the indication of resistance and reactance, but the mathematical formulas involved would, from the standpoint of immediate usefulness and the knowledge assumed for the reading of this book, be more involved than desirable. Hence these are omitted. Alternating-current measuring instruments are not included for two reasons: (1) so far, very little use has been found for them. From known values of power and voltage (or impedance) it is easier to calculate the currents than if these measurements were taken directly; (2) there is the difficulty of securing readings with any reasonable degree of accuracy using only modest equipment. Due to the inherently low resistance of a current meter, it is subject to stray pick-up from nearby electric and magnetic fields. This factor necessitates elaborate shielding and very careful placement in the circuit before a true indication can be obtained.

Do not confuse these current measuring instruments with those that will be used in conjunction with the diode and triode

vacuum-tube voltmeters or other rectifying instruments. The latter actually measure only an average d.c. value and are low-frequency instruments adapted in some special way for the ultra-highs. Such meters retain adequate accuracy.

WAVELENGTH

In studying the history of some particular development in any of the sciences, a common pattern can usually be found. As new problems arise, engineers in these fields try to use existing equipment. When these fail, the same equipment is tried again, this time with modifications as demanded by the particular needs of the case. If still unsuccessful, new methods will be evolved. Examples may be found in the field of radio. Take the case of the u.h.f. wavelength meter, a piece of apparatus that early achieved prominence because experimenters were primarily interested in reaching the shortest wavelengths possible, considerations of power coming second. The common coil and condenser absorption type of wavemeter was widely used because it could be constructed quickly and simply and calibrated with good accuracy. For use at the higher frequencies the dimensions were reduced in accordance with the formula

$$F = \frac{1}{2\pi\sqrt{LC}}$$

until the desired frequencies were reached. The former multiturn solenoid became — in the final analysis — simply a one-turn loop closely attached to a small condenser built with a few semicircular plates, all mounted on insulators that had very little leakage at the wavelengths under measurement. A good example of a meter of this type is shown in Fig. 8.3, the General Radio wavemeter, type 758-A. This instrument must be placed close to the u.h.f. generator so that there is an adequate energy transfer to the wavemeter. Resonance may be indicated generally in one of three ways:

1. By observing when the light bulb becomes illuminated.

2. By replacing the bulb with a sensitive thermocouple-microammeter, and noting its reading.

3. By eliminating the bulb altogether by means of a small shorting bar and obtaining the indication of resonance from meters in the oscillator circuit.

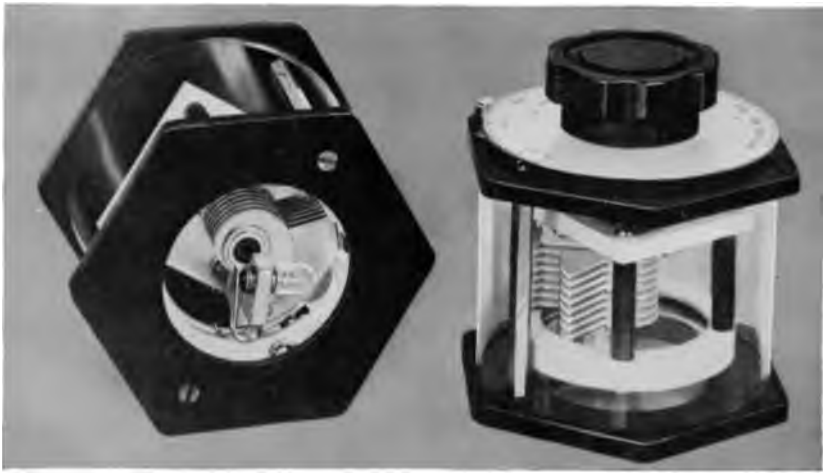


FIG. 8.3. The General Radio Wavemeter — 55-400 mc. (Type 758-A)

The range of this wavemeter is from 55 to 400 megacycles; values are read directly from the calibrated scale. An accuracy of 2 per cent is possible with reasonable care.

Wavemeters of the type just described cannot, however, be used much above 900 megacycles, even if the condenser plates are reduced to a minimum. The Q or sharpness of resonance would soon decrease at these higher frequencies to the point where reasonable accuracy could not be obtained. At this point engineers turned to other equipment; in this case an ordinary transmission line. It will be remembered that, as the frequency of oscillators rose and the same decreasing selectivity mentioned above was encountered, simple forms of open-wire and coaxial transmission line could be used with good results. They had a high Q, even at the very high frequencies, moderate lengths, and the ability to act as a resonant circuit at different frequencies by

the simple expedient of changing their length — usually accomplished by a movable bar. The length needed at the low frequencies prevents its use there, the coil and condenser being much more compact and capable of high accuracy.

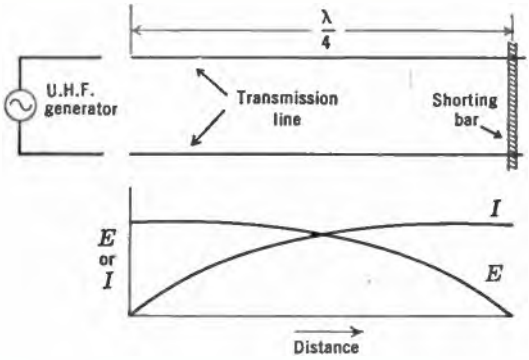


FIG. 8.4A. Current and voltage distribution on a shorted ¼ wavelength line.

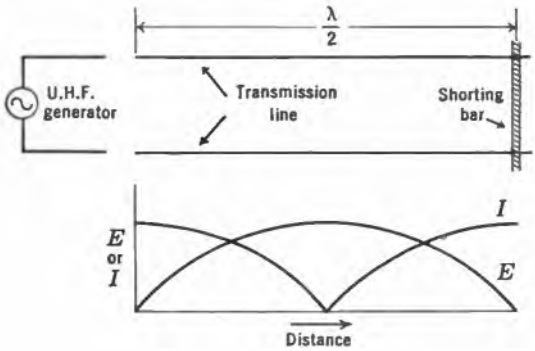


FIG. 8.4B. Current and voltage distribution on a shorted ½ wavelength line.

Transmission Line Wavemeters. In order to understand how the transmission line may be used to measure wavelength, review a few of the facts given in Chapter 4. Here it was shown that if a short length of line was excited by means of a generator, then, for all terminations of the line other than its characteristic impedance, standing waves would be set up. Fig. 8.4A reviews the

distribution of these standing waves for a quarter-wave line and Fig. 8.4B shows the same conditions on a half-wave line. The wavemeter wires do not necessarily have to be either one or the other of the above lengths; they may have any value. The important point is the formation of the standing waves. Generally a short-circuited line is used, for, then, it is easy to move the shorting bar back and forth.

The determination of the wavelength may now be accomplished in two ways. With a generator placed close to the transmission line, adjust the movable shorting bar until the greatest standing voltages (or currents) are formed along the line. This

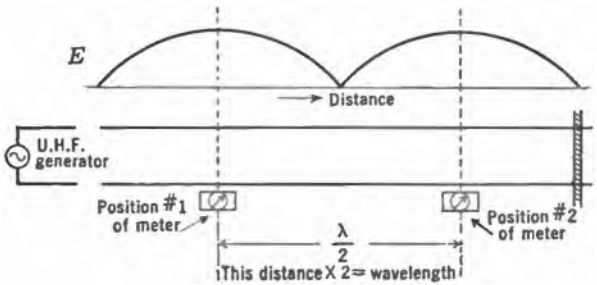


FIG. 8.5. Using a transmission line and a movable meter to measure wavelength.

indication may be detected by means of a crystal detector and a d.c. microammeter placed in series with each other and capacitively coupled to the line. The meter does not have to be calibrated for absolute readings; relative values are sufficient. The instrument is now moved along the line and the distance between successive maxima noted. As shown in Fig. 8.5, this distance, multiplied by 2, will give the wavelength of the generator output. The higher the Q of the line, the sharper will the resonant maximum point be and the greater the accuracy with which this wavelength may be obtained.

The second method is simpler and finds greater application. Here the crystal and microammeter are placed at a point of maximum voltage and then, by means of a screw arrangement, the

shorting bar is slowly moved along until the meter records a second maximum. The distance through which the shorting bar was moved, multiplied by 2, will give the wavelength of the wave under test. Fig. 8.6 shows essentially what has occurred. At *A*

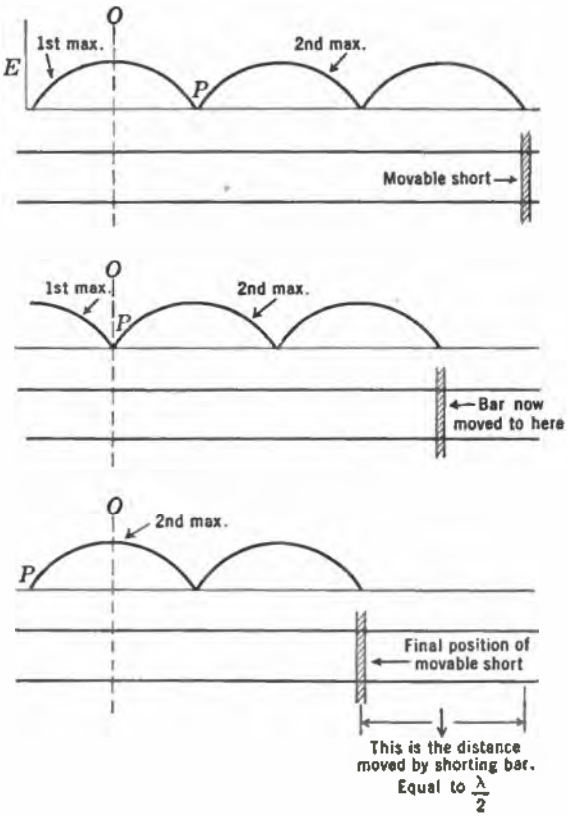


Fig. 8.6. Determining wavelength by means of a movable shorting bar and a meter positioned at *O*.

the line is shown as it would be if it were 1½ wavelengths long. The indicating device is placed at point *O* where it indicates a maximum. Then, as the shorting bar is moved along, the wave can be considered as being pushed before it. Since the voltage at *O* was a maximum at the start, moving the shorting bar along

will cause the voltage standing wave, moving past this point, to decrease. The voltmeter indication will likewise decrease and reach zero when point *P* passes the meter positioned at *O*. Beyond this, the meter indication will again increase until the second maximum is now at point *O*, opposite the meter. The distance moved by the shorting bar in this procedure should be exactly one half wavelength long. The distance is read from a calibrated scale and multiplied by 2. Greater accuracy can be attained by this procedure than with the previous method in which the voltmeter changed position. A vernier screw mechanism will further aid precision.

In practice it is not even necessary to place the voltmeter at a maximum point. If the line is several wavelengths long, the meter can be placed at almost any position, although preferably near the beginning of the line. Note its reading and start turning the screw drive. When the first maximum point is reached, take the reading on the vernier scale. Continue turning the crank handle in the same direction until a second maximum is reached. Make a note of the reading now on the scale. If the previous reading is subtracted from the latter, and the result multiplied by 2, the answer will again be the wavelength. This is the most widely used method when using transmission lines for wavelength measurement.

At the very short wavelengths the inductance of the shorting bar must be taken into account. If it is neglected, the readings will be slightly too low. The easiest way to adjust for this error is by calibration against known frequencies. Another method involves the determination of the inductance of the bar by the usual bridge methods and then using this value to obtain the correction factor needed. This alternative may be used if there are no known frequencies available for calibration.

In all of the foregoing measurements, care should always be taken to make sure that the coupling between generator and transmission line is loose. There is less chance that the frequency of the generator will be affected, and a sharper resonant point can be secured on the line.

VOLTAGE

In making voltage measurements, the greatest accuracy is obtained by using meters that have a large impedance. For the low frequencies, the serviceman finds meters with 1000 ohms per volt and 5000 ohms per volt satisfactory for most measurements. Occasionally, when tracing voltage paths in the a.v.c. system of a superheterodyne, it is necessary to use vacuum-tube voltmeters. This particular circuit employs very low voltage and very high resistance. The type of meter movement is only useful directly on d.c. voltages, but may be adapted for a.c. measurements by the use of small copper oxide rectifiers. For the latter type of measurement, voltages with frequencies up to 20,000 or 25,000 cycles per second may be read with good accuracy. For higher frequencies, the capacitance inherent in the meter combination leads to serious inaccuracies, the effect increasing with frequency.

Thermocouple Voltmeters. For extremely high frequency measurements, the thermocouple type of meter finds extensive application. This instrument works on a principle which is quite different from anything encountered previously. It is based on

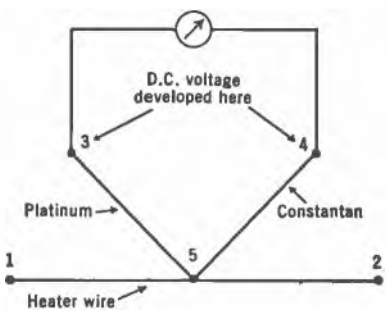


FIG. 8.7. A thermocouple meter arrangement.

the fact that when two dissimilar metals are joined together and the junction heated, a d.c. voltage will be produced which is proportional to the amount of heat generated. For practical use, a current is sent through this junction point and heat is thus generated. Of greater interest are the facts that the heat is pro-

portional to the square of the current and the voltage generated by this heat is directly proportional to the temperature rise. Thus, by proper calibration, it is possible to measure currents and voltages of very high frequencies. Fig. 8.7 shows the arrangement of the usual type of thermocouple meter. The junction, point 5, is called the thermocouple. The thermocouple wires are generally made of platinum and constantan. The heater strip conducts the current to and from the junction and should be quite short in length so that the current will tend to be uniformly distributed and so that each point will be evenly heated.

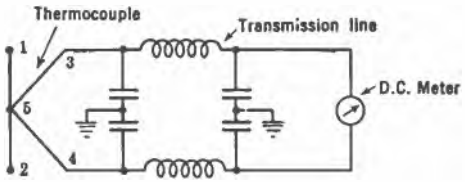


FIG. 8.8. A thermocouple meter arrangement where the d.c. meter is placed at some distance from the thermocouple. A transmission line conducts the d.c. voltage from points 3 and 4 to meter movement.

The component metals were chosen because they give rise to a greater potential difference than most other combinations. It is not at all necessary to have the meter movement connected directly to the points 3 and 4; a transmission line may be inserted and the meter kept at some distance. Fig. 8.8 shows the necessary set-up. For very low currents, some manufacturers enclose the couple and heater strip within a glass envelope and evacuate the air. One model available from an eastern manufacturer is shown in Fig. 8.9. It is called a vacuum thermocouple.

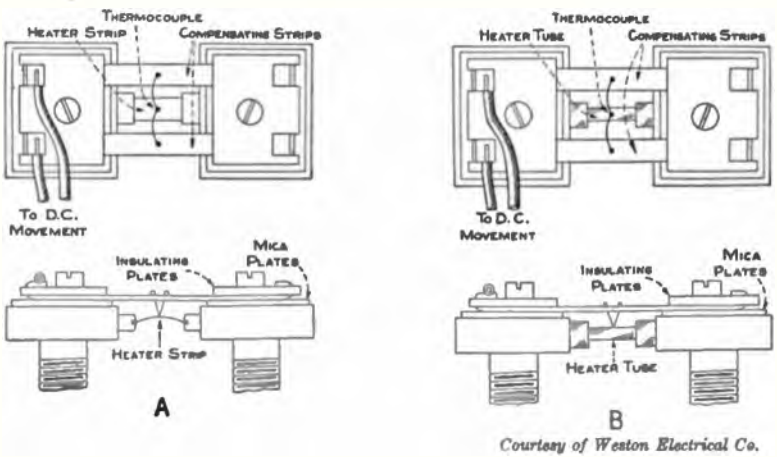
Calibration of the meter scale is made at the low frequencies where it is simpler to obtain known currents. Since the d.c. current produced by the thermocouple will be proportional to the square of the current flowing through the heater wire, the scale will be more crowded at one end. The range of any one scale will be limited, but, by the addition of shunts, accurate readings over wide ranges may be obtained by changing the shunt whenever the indication is at the crowded end of the scale. The use of this

meter at the high frequencies necessitates that there be practically no skin effect in the heater strip because this would invalidate

Courtesy of Weston Electric Instrument Co.

FIG. 8.9. A vacuum thermocouple.

the low-frequency calibration. Two common methods are employed. One method involves the use of a hollow conductor



Courtesy of Weston Electrical Co.

FIG. 8.10. Two types of thermocouples which are accurate up to fairly high frequencies.

while the other calls for a very thin strip of metal. Both are illustrated in Fig. 8.10, together with a temperature compensating strip to eliminate error due to changes in the surrounding tem-

perature (ambient temperature). The compensating strips have the same thermal characteristics as the heater strip itself. The cold ends of the couple are connected to the compensating strips and the hot junction is placed on the heater wire. Since the compensating strips and the heater strip are made thermally equivalent, there is a temperature balance. However, when current heats the heater strip, there is a resultant difference in temperature between the hot and cold ends. This temperature difference produces a thermoelectric voltage due solely to the current flowing. Incidentally, it should be recognized that any type of current flow in the heater strip will cause this thermoelectric voltage. Hence d.c., a.c., or r.f. currents may be measured this way.

For use as a voltmeter, all that is necessary is a high resistance in series with the heater wire. The electrical arrangement is given

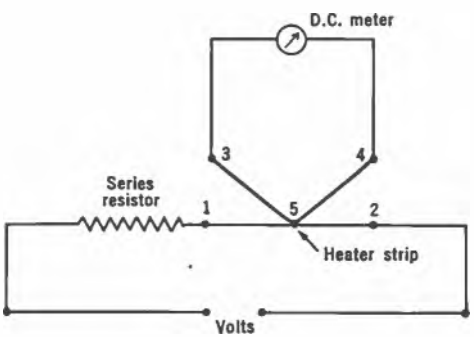


FIG. 8.11. A thermocouple voltmeter.

in Fig. 8.11. The resistor should have as high a resistance as the movement will permit. The construction of the resistor is very important because it is actually equivalent at the higher frequencies to a pure resistance with a shunting condenser. This combination will present a lower impedance than at the lower frequencies and might even require additional calibration under actual operating conditions. Fortunately, many voltage measurements taken at the shorter wavelengths will serve as well if they are merely relative values, the absolute quantities being

unnecessary. In this instance, the low-frequency calibration will serve effectively.

Vacuum-Tube Voltmeters. Another type of voltmeter frequently employed for the u.h.f. measurements is a vacuum-tube voltmeter. To understand better the various corrections necessary in this type of meter when used at the short wavelengths, it might be instructive to investigate its operation at the ordinary frequencies where it is considered one of the most sensitive measuring devices available. Consider, first, the simple diode tube arranged as a half-wave rectifier, as shown in Fig. 8.12. With the voltage positive at the plate, current will flow and charge condenser C to the peak value of the voltage being measured. A high resistance and a sensitive milliammeter or microammeter are placed across the condenser (essentially a voltmeter arrangement). A higher resistance is possible with the last mentioned meter. If the series resistance is high enough, the peak value of the voltage under test will be obtained. This relates only to peak value of the positive peaks, since there is no conduction through the tube during negative plate voltage periods. For smaller values of series resistance, the condenser will discharge too rapidly and never reach peak value. In this instance, less than the peak value will be read, the amount depending upon the value of the resistance. It is well to note that in this combination, when the current flows, the meter has a finite value of resistance. Hence, the meter will introduce inaccuracies due to the shunting effect across the circuit under test. For the negative half cycle, when no current flows, the resistance is very great. The more current that flows in this circuit, the lower the tube resistance and the greater its shunting effect on the circuit under test.

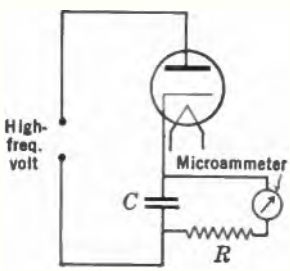


Fig. 8.12. A diode vacuum-tube voltmeter.

For more sensitive arrangements there are several triode tube circuits available, three of which are given in Fig. 8.13. The first circuit, Fig. 8.13A, is the popular slide-back type vacuum-tube

voltmeter and it is used to measure peak a.c. or d.c. voltages. To operate this meter, the input leads are first short-circuited and the position of the movable arm on resistor R_1 adjusted to give zero plate current reading on meter M_1. Meter M_2 registers the grid bias voltage needed to cause this zero plate current. The leads are now placed across the voltage to be measured and the

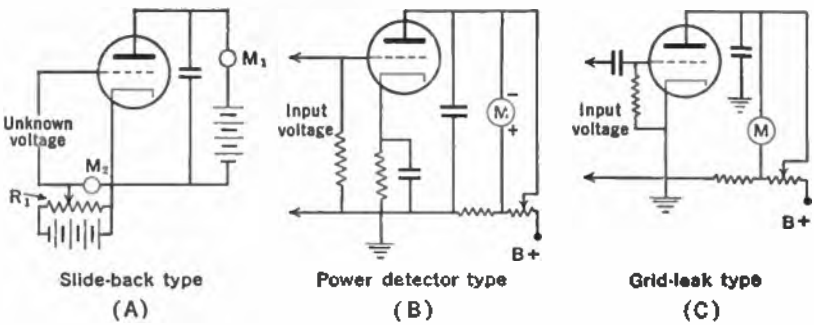


FIG. 8.13. Triode vacuum tube voltmeters.

grid bias voltage adjusted, by means of R_1, to again give zero plate current. Meter M_2 should now read some new value and it is the difference between this latter reading and the first reading of M_2 that indicates the peak value of the a.c. voltage under test. Naturally, for d.c., this will be equivalent to the average value also.

The great advantage of this meter is that it requires no calibration. With careful design and construction, accurate readings can be obtained. For sinusoidal waves, r.m.s. values can be computed by multiplying the peak values by 0.707. Other shaped waves, though, cannot be handled so easily. In practice it is quite common to use one meter for both purposes of measuring current and voltage. Appropriate shunts, of course, are necessary.

The second and third types of voltmeters illustrated in Fig. 8.13 are essentially detector (or rectifier) circuits. At B we have the so-called power detector while at C we find the famous grid-leak detector so widely used in former radio sets.

For the vacuum-tube voltmeter at B, the cathode resistor is generally of a high value so that the tube is kept near the cut-off point with only a small amount of plate current flowing. The operating point on the characteristic curve is shown in Fig. 8.14. When the test leads are placed across the voltage to be measured, an increase in plate current will be observed, the increase depending upon the strength of the input voltage. With proper calibration against known voltages, the plate current meter readings can be used to measure unknown voltages in the same range.

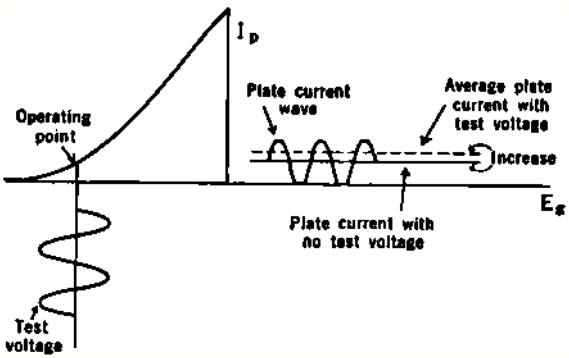


FIG. 8.14. Illustrating how a voltage placed across the input of the power detector vacuum voltmeter causes an increase in plate current.

Instead of having the rectification occur in the plate circuit as just described, the grid-leak detector accomplishes this in the input circuit. The end result, though, is the same, namely, an increase in plate current. The power detector circuit is perhaps the better one to use since it can handle high voltages with less loading upon the circuit under test.

In both B and C of Fig. 8.13 provision has been made to reduce the no signal plate current to zero before any readings are made. This insures wider use of the plate current meter scale and provides a common point from which all readings may be taken. Pentodes are seldom employed in vacuum-tube voltmeters because of the extra trouble brought on by the additional voltages required for the other electrodes. They may, however,

be used as triodes. The smaller the voltage to be measured, the more constant must the tube operating voltages be and the more sensitive the milliameter inserted in the plate circuit.

The essential theory of the basic vacuum-tube voltmeter has been covered. There are many variations and applications of these meters, both at the low frequencies and at the ultra-highs. Since it is with the latter that we are primarily concerned, they will be considered now. The vacuum-tube voltmeters described so far can be calibrated at 60 cycles and used up to 20 megacycles with good accuracy. If many measurements are to be made at the high frequencies, it is usual to insert the tube in a probe so that the grid cap can be placed directly at the point where the voltage is to be measured. This eliminates any shunting capacity which a pair of long leads would introduce and also eliminates standing waves on these wires. A standing wave would result in a reading that would not be a true indication of the voltage under test.

For higher frequencies, small acorn tubes (like those mentioned in Chapter 1) are advisable. The shunting capacity is reduced due to smaller electrodes and special construction. But shunting capacity is not the greatest trouble by any means. As the frequency increases, two effects come into play and it is largely due to these that the readings become inaccurate. One error is caused by the series resonant combination of lead inductance and tube interelectrode capacitance. When measuring voltages with frequencies near the resonant point of this series circuit, the voltage that actually appears across the tube is greater than the voltage under test because of the resonance. This, of course, is characteristic of all series resonant circuits. The error is greatest when the voltage under test is of the same frequency as the series circuit. For practical operation, the resonant frequency of the leads and the tube capacitance must first be determined and a proportional equation set up to compute the effect at the frequency under test. It is not very difficult to apply this correction since the result is independent of the applied voltage. Usually this error is the larger of the two[10] mentioned above.

The second cause for inaccuracy is none other than the previously discussed phenomenon of electron transit time. To see why this causes low readings, consider the following explanation of tube action[11]. A diode vacuum-tube voltmeter (Fig. 8.12) will be used for illustration although the same line of reasoning will apply equally to any other tube. When a high-frequency voltage is applied to the diode, the condenser is charged by the current that flows through the tube during the positive half cycles. If the electrons from the cathode reached the plate in zero time,

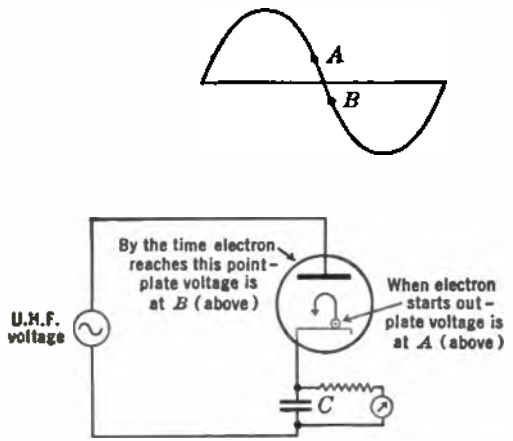


FIG. 8.15. When the voltage under test has a very high frequency, transit-time errors will cause C to be undercharged.

the condenser would charge to the peak value of the voltage under test. However, as it takes the electron some measurable time to travel the filament-to-plate distance, the positive voltage on this anode will change while some electrons are still in flight. Thus the electrons that left the cathode when the plate was at point A, Fig. 8.15, may not reach the anode before the voltage goes negative (point B). Hence they will be repelled. The condenser, deprived of their added charge, will read low. Remember that this effect is present at all frequencies. At the longer wavelengths, the time it takes an electron to travel the interelectrode distance is negligible in comparison with the time of one cycle and

so the condenser is charged to essentially full value. In calibrating the meter for actual use, a correction must be computed and applied to all readings in order to obtain the true value.

Crystal Detector. One type of meter that enjoys wide application in field strength measurements is a crystal detector and a microammeter. Generally, only relative readings are desired. Hence there is no need for accurate calibration of this combination. The greatest difficulty in adjusting the meter lies with the crystal,

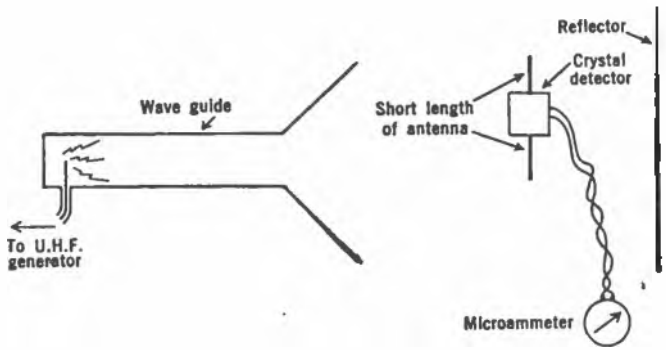


FIG. 8.16. Illustrating the use of a crystal detector for field strength measurements.

which is almost always extremely critical. Care must be taken to see that no jarring of the crystal occurs while it is in use. To provide good pick-up of the signal, a small half-wave antenna is incorporated into the system, and the crystal is connected at the center of this dipole. See Fig. 8.16. Any of the field strength patterns shown in Chapter 7 may be obtained by this process. The same type of meter may be used to investigate the electric field distribution inside wave guides and cavity resonators.

It is desirable to calibrate all of the foregoing instruments at 60 cycles or some other low frequency where it is possible to obtain accurate voltages. Direct calibration at the ultra-high frequencies is not easily accomplished due to a lack of some fixed standard. Consequently, false readings at the ultra-high frequencies will probably occur. Good qualitative results, however, may be obtained by calibrating the meter at some low frequency

and then checking this calibration at the short wavelengths by either one of the following methods.

Calibration of U.H.F. Voltmeters. The first method calls for a length of transmission line and a source of variable high-frequency power. The power is fed to the line and, if the line is not terminated in its characteristic impedance, standing waves of voltage and current will be formed. At regular intervals the voltage will pass through a maximum, and a little farther on, through some small finite minimum. Take the voltmeter whose calibration is to be checked and measure the ratio of the maximum voltage to the minimum voltage at several points along the line so that an average set of values is obtained. Now change the power fed into the line and again get the average ratio of $E_{max.}$ to $E_{min.}$ If the frequency is kept constant, the ratio should also be constant and the point at which this ratio starts to change indicates the end of the linearity of the meter scale. This method indicates just how much of the meter scale to use and quickly shows up any changes in the low-frequency calibration. When the meter has not been calibrated, the method is still useful if only relative values are desired. Fortunately, such is the case in many u.h.f. measurements.

A variation of this method consists in finding the maximum voltage point on a suitably operated transmission line and then moving the indicator down the line for a distance of 45 degrees or ⅛ wavelength. At this point the voltage should be 0.707 of the maximum voltage. This ratio is taken at several points along the line in order to obtain an average. Again a test is made to see whether this ratio is the same when the amount of power fed into the transmission line is varied. When the ratio ceases to agree with the above 0.707 value, the meter is no longer linear. Sinusoidal waves are used to obtain the 0.707 ratio, and the line is terminated so that appreciable mismatching will occur. A simple method of varying the applied power to the line is to couple the generator loosely to the line and then increase the distance between the two for less and less power transfer.

The meters described represent what might be called the

beginning of ultra-high frequency voltage measurements. Undoubtedly more and better instruments will be devised as this science progresses. The principles, however, will remain the same, even if their applications should differ.

POWER

For the measurement of power there are a variety of methods, all based on the principle of dissipating the power in the measuring device. While few of these methods can achieve the accuracy attainable with low-frequency equipment, they still are capable of results comparable with those obtained with u.h.f. voltage meters. As with the previously described measuring devices, low-frequency calibration is desirable. To achieve this goal for power instruments it is necessary for the power dissipating element to have the same resistance at both the low and high frequencies. Another important factor relates to the size and shape of the heater element. It is necessary to have this short and straight so that the current will be uniformly distributed along the wire and so that the heating will be even. Some areas will become hotter than others, if large standing waves are allowed to form on the heater wire, and the low-frequency calibration will no longer be accurate. These are the important considerations that must be carefully checked when these meters are to be used at the ultra-high frequencies. Other points of lesser importance will be mentioned with each instrument.

At the present time, power instruments at the ultra-high frequencies are used solely to determine the power output of transmitters. With this in mind it has been found easiest to accomplish the measurement by coupling the power into the power instrument and here dissipating the entire output. The element where the power is dissipated is generally a small, thin, uniform resistance wire, this being capable of constant characteristics over wide frequency limits. Mechanical movements, such as found at the low frequencies, are entirely out of the question. Short lengths of transmission lines are used to couple the power from the transmitter to the meter.

Vacuum Thermocouples. For the measurement of small amounts of power (below 1 watt) vacuum thermocouples, perhaps of the type shown in Fig. 8.9, may be employed. The heater element should have a comparatively high resistance and yet be so constructed as to allow uniform current distribution along its length at the frequency in use. The resistance is made larger than that of any of the other wire resistances in order that prac-

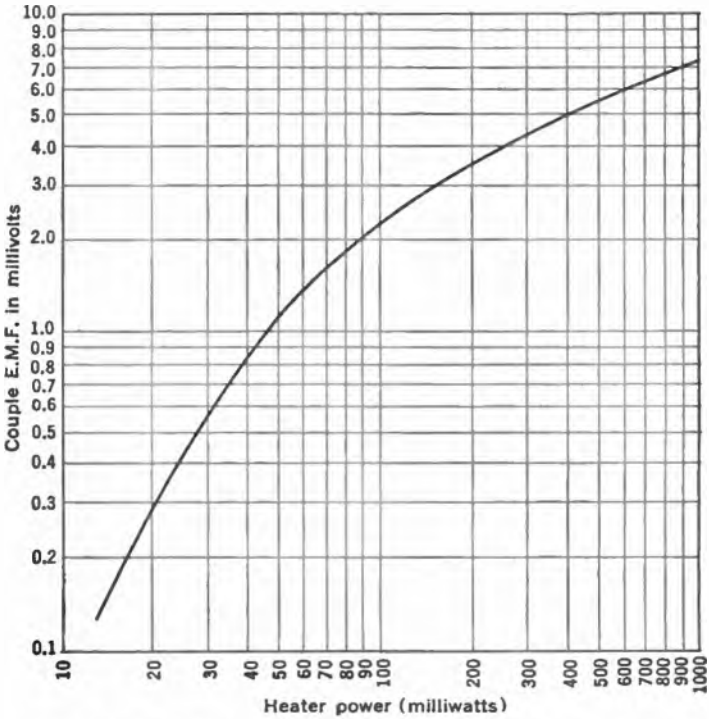


FIG. 8.17. Low frequency calibration of a vacuum thermocouple.

tically all the power will be dissipated in the meter. Generally, the junction is placed so that it will minimize any tendency toward coupling. The heater and couple unit may be placed close to the indicating meter or it may be placed at some distance. The necessary connections have already been explained under voltage measurements.

For calibration, low-frequency currents are generally employed. Since the resistance of the heater element is considered constant, the e.m.f. registered on the d.c. meter connected to the couple will be proportional to the square of the current. Power is likewise proportional to this quantity, and so a calibration chart between the e.m.f. of the meter and the power may be drawn. A typical graph is given in Fig. 8.17. In measuring power beyond the 1-watt range, the heat dissipated in the heater element might injure the couple or burn out the heater wire. There are means whereby the construction of the thermocouple may be altered to handle larger amounts of heat, but they give results which are less accurate than other available methods and hence are not generally used.

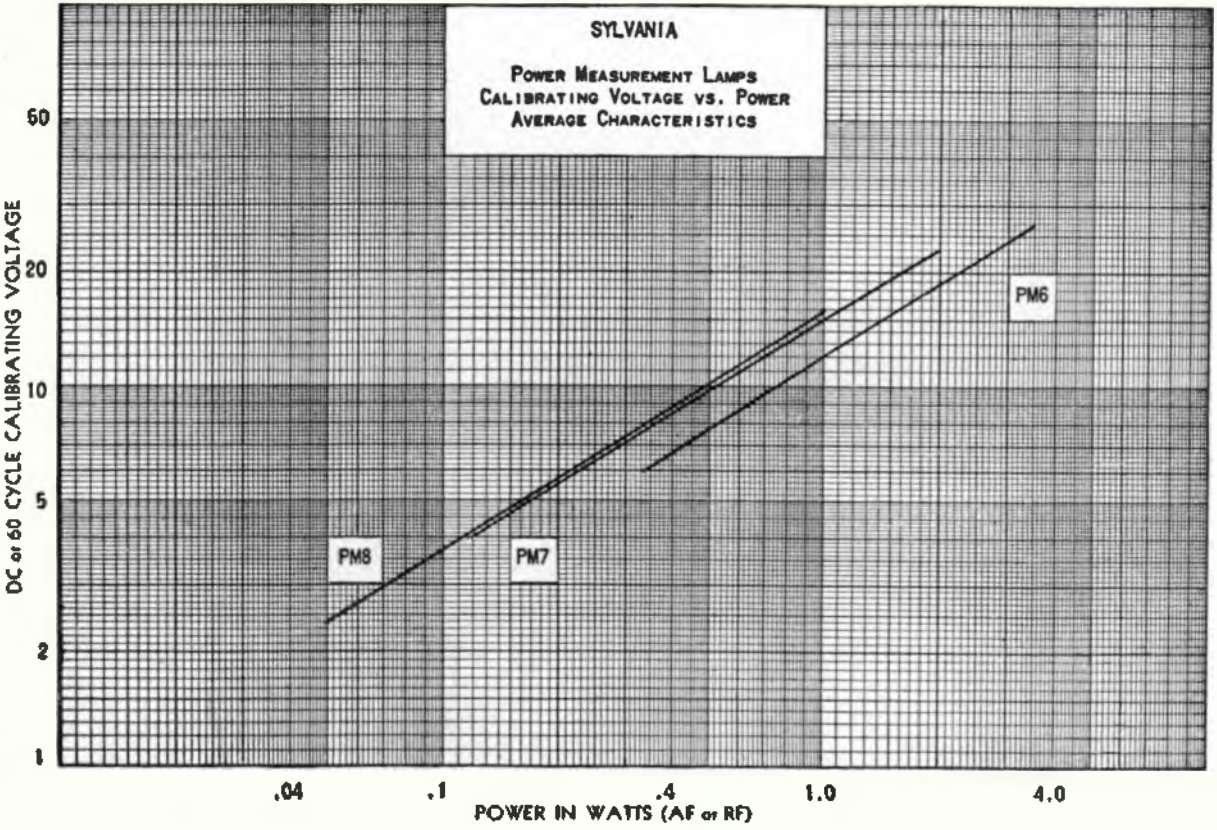
Power Measurement (P-M) Tubes. Recently a set of power measurement tubes have been constructed especially for the higher frequencies.* The tubes, six in number, are shown in Fig. 8.18. Each is seen to consist of two similar filaments, one of

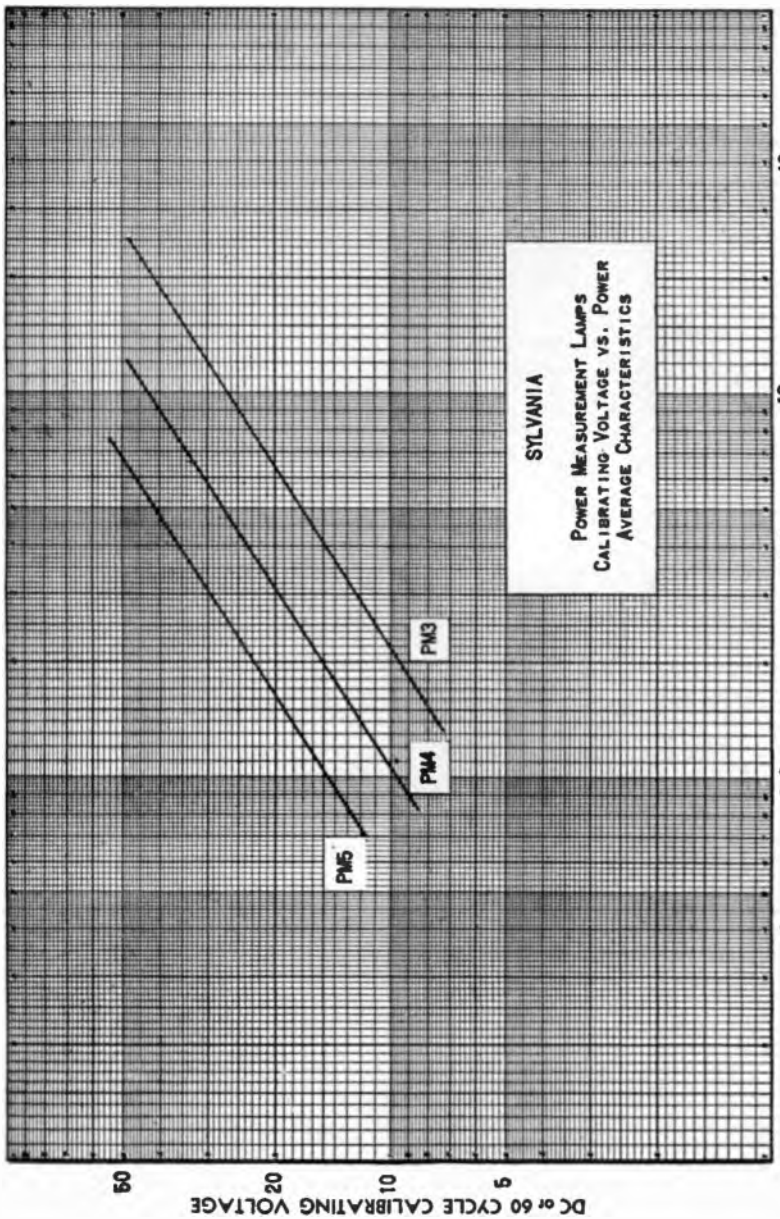
Courtesy of Sylvania Electric Products Co.

FIG. 8.18. Six Power Measurement Tubes.

which is used as the load, where the unknown power is dissipated. The other filament is for calibration, either on d.c. or 60-cycle a.c., and is adjusted until its brightness is equal to that of the load filament. The unknown power then equals the d.c. or 60-cycle a.c. power. Calibration curves applicable to these six tubes are given in Fig. 8.19. It may be seen from these graphs and the

*The author is indebted to the Sylvania Electric Products Company, of Emporium, Pa., for graciously making this material available to him in time for inclusion in this book.


SYLVANIA
POWER MEASUREMENT LAMPS
CALIBRATING VOLTAGE VS. POWER
AVERAGE CHARACTERISTICS
DC or 60 CYCLE CALIBRATING VOLTAGE
50
20
10
5
2
1
PM8
PM7
PM6
.04
.1
.4
1.0
4.0
POWER IN WATTS (AF or RF)



Courtesy of Sylvania Electric Products Co.

FIG. 8.19. Calibration curve for six P-M tubes.

rating chart given in Table 1 that the measurable power range extends from 0.05 watt to 25 watts. The normal power range of the lamps will give a brightness variation from red to yellow which allows direct visual comparison. The lamps may be used up to the maximum power rating, provided a dark glass filter is used for visual comparison or a photocell detecting device is resorted to. Measurements can be made to 5 per cent accuracy without special

	PM 3	PM 4	PM 5	PM 6	PM 7	PM 8	
Maximum freq. for $R_{a.c.} = R_{d.c.}$	100	100	150	200	400	900	Mc.
Normal power (min.)	1.35	0.8	0.7	0.33	0.12	0.05	Watt
Normal power (max.)	10.00	5.5	3.8	1.75	0.95	0.41	Watt
Power dissipation (max.)	25.0	12.0	7.5	3.5	2.0	1.0	Watt
Applied d.c. volts (max.)	48	48	55	28	24	16	Volts

Courtesy of Sylvania Electric Products Co.

TABLE 8.1. Characteristics of the six P–M lamps.

calibration of either filament. For more accurate results, the lamps can be calibrated by either impressing a known voltage or current in the calibration side, applying a voltage to the load side and measuring this load voltage when the two sides are equally bright. The power in the load is then known in terms of voltage or current in the calibration side and the accuracy will be independent of any difference between the two filaments.

Because of the high heat conductivity and small diameter of the filament material, the temperature and color will be constant for a given amount of energy dissipated, regardless of whether the heat is liberated uniformly throughout the cross section, as with d.c., or non-uniformly due to skin effect at the ultra-high frequencies. Generally, power can be measured at any frequency at which it is possible to couple the energy to the lamp filament. In connection with the last statement, a transmission line will serve nicely.

The high-frequency resistance of any one of the power lamps of Fig. 8.18 does not differ appreciably from its d.c. resistance up to the frequencies given in Table 1. This is due to the fact that the wire is so small that the depth of penetration is greater than, or at least equal to, the wire radius. When the lamps are used at higher frequencies, where skin effect becomes a factor, the resistive component increases. However, the power will be correctly indicated because the r.f. current will be less than the d.c. for equal filament brightness. Then the values of I^2R will be the same for r.f. as for d.c.

A variation of the method for fairly large power measurements involves the use of incandescent light lamps. Two such lamps are necessary, one for the high-frequency power dissipation, the other for comparison, using 60-cycle a.c. The adjustment for equal brightness is made by direct visual comparison or by means of a phototube. Unless the lamp filament used for the ultra-highs is of special construction, large errors will be introduced at the very short wavelengths because of uneven heating.

Diode Power Meters. For small or moderate values of power, diodes can be used with good accuracy. The power to be measured is dissipated in the filament of this tube. With sufficient positive potential applied to the plate, all the electrons emitted by the filament will be drawn toward the anode. The plate current can then be taken as an indication of the power being used to heat the filament. Up to a certain point, determined by the physical structure of the tube, the current will increase as the power dissipated by the filament increases. Two limitations are imposed by this method: one relates to the maximum heat loss that can be safely dissipated at the plate and the other is the inaccuracy introduced by the nonuniform heating of the filament wire. The filament should be constructed so that the conditions given at the start of this section are met.

One precaution must always be observed whenever this method of power measurement is employed. The plate voltage must be sufficiently high so that substantially all the electrons emitted by the filament are attracted to the anode. This means that the

plate voltage must be placed at point 1, Fig. 8.20, where a typical E_p-I_p curve for a diode is given. If more power is dissipated in the filament, curve B will result and the anode voltage must be increased to point 2 or some higher value. Points 1 and 2 refer to minimum voltage values needed in each case to secure saturation currents. Naturally any plate voltage higher than these will also be satisfactory.

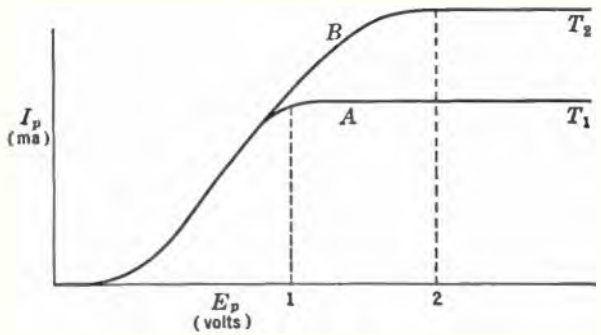


FIG. 8.20. As the power dissipated in a tube filament is increased, the plate voltage must correspondingly be increased to insure that all the electrons emitted are attracted to the plate.

Transmission Line Power Meter. Advantage can be taken of the fact that many transmission lines at the ultra-high frequencies have a characteristic impedance which is entirely resistive, a fact previously mentioned in Chapter 4. The generator whose power output is to be measured is connected directly to a short

FIG. 8.21. A transmission line power meter.

length of transmission line. The line will then transport the power smoothly to the terminating resistance at the other end, and here all the power will be absorbed and dissipated. Since the terminating resistance is, at the same time, also the charac-

teristic impedance, no standing waves will be set up on the line to distort the current and voltage distribution. From known values of resistance and of either voltage or current, it is possible to obtain the power output of the u.h.f. generator. Fig. 8.21 shows how such a set-up might appear. For short lengths of line very little power will be lost in the inherent resistance of the cable.

Summary. In adapting low-frequency meters to ultra-high frequency measurements, various correction factors must be introduced before true readings can be obtained. The correction factors arise from the changes which occur in the resistance, inductance, and capacitance values of a circuit as the frequency is raised. Due to the comparatively limited experience which we have had with the ultra-highs, few accurate measuring instruments are available at present. Of those that are now in use, only the meters that measure voltage, power, and wavelength have been discussed in this chapter.

For wavelength measurements up to a few hundred megacycles, it is still possible to use the conventional coil and condenser absorption-type wavemeters. Above this figure, the decreasing selectivity requires the adoption of a transmission line, which now assumes reasonable lengths. The procedure with the latter meters calls for the setting up of standing waves of voltage and current along the line and then measuring the distance between successive maxima (or minima). This distance multiplied by 2 is equal to the wavelength of the generator under test. Generally, the indicator is kept in one fixed position and a movable shorting bar is adjusted to give the necessary readings.

For voltage measurements there are three commonly used methods. The first meter described consists of an ordinary thermocouple arrangement with a large series resistor. Current, in passing through the thermocouple junction, gives rise to a certain quantity of heat, which in turn is responsible for the appearance of a d.c. potential difference across the cold ends of the couple. The deflection of a meter placed across this d.c. potential will be a measure of the current flowing in the heater, or of the external voltage across the meter combination.

A second method involves the use of either a diode or triode vacuum-tube voltmeter. The method of operation is much the same at the ultra-highs as at the lower frequencies. Care must be taken in the construction of this meter and correction factors must be introduced to correct for transit-time and resonance effects. Finally, there is the crystal detector and microammeter for field strength measurements. With a short length of antenna to aid in the pick-up of electrical energy, this device is capable of good results for comparative measurements. Its most critical adjustment is in the crystal, which will change calibration upon being jarred.

Power-measuring meters are based on the principle of complete dissipation of the power in the instrument itself. Vacuum thermocouples again find use, although only for small values of power. Power up to 25 watts may be measured by the set of P-M tubes recently made available. Here the brilliance of the load filament is compared to a second filament which has a known amount of power dissipated in it. When the filaments are equally bright, the unknown power is exactly equal to the known power in the standard filament. With sufficient care, even voltage and currents may be measured in this manner.

It is further possible to measure power with small vacuum tubes, the filament being the recipient of this power. Plate dissipation and filament current-carrying capacity impose two limitations upon this method.

CHAPTER 9

WAVE PROPAGATION

The various means whereby radio waves travel from transmitter to receiver and the results thereof

It is not surprising to find that the ultra-high frequencies act differently than the low frequencies upon being propagated through the atmosphere; not surprising, that is, after growing accustomed to having these frequencies show peculiarities in so many other ways. It is believed that the difference may be more clearly seen and better understood if a contrasting picture is drawn between the highs and the more familiar lows. Consideration will be given mainly to the propagation of electromagnetic waves after they have left the particular antenna system that happened to be used and are sent out into space.

Information on the propagation of frequencies above 200 megacycles is very meager since it is only in the past few years that attention has been focused on them. However, from the few results that have been published, it can be reasonably assumed that the behavior of waves above 200 megacycles does not appreciably differ from that of waves from 50 to 200 megacycles.

Polarization. Quite familiar, by this time, is the fact that electromagnetic waves have two components and that in free space they are always at right angles to each other. (See Fig. 9.1.) The direction of the electric vector has been taken by definition to be the direction of polarization of the whole wave. Thus a vertical antenna, which gives rise to a vertical electric field, is said to be vertically polarized, and an antenna which sends out waves with the electric field horizontal is horizontally polarized. These waves may and often do change their sense of polarization as they travel along. Hence it is difficult to determine the original

polarization of a wave from observations at the receiver. If the sense of polarization of a wave is known, it can be used to advantage because the receiving antenna will usually receive the signal strongest when its polarization is the same as that of the signal being received. The amount of polarization shift between

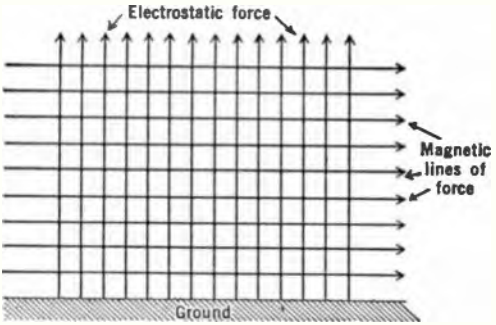


FIG. 9.1. The two components of a radio wave. In this case the wave is vertically polarized and is travelling along the ground.

the transmitting and receiving antennas is one of those factors that can only be predicted with a great many ifs. Also, what is true one day may not be true on the morrow. Polarization shift is very important, yet the conditions determining it are beyond our control.

Radio Wave Propagation — Ground Wave. Investigation of the subject of radiation will now be made to see what occurs at the various frequencies. Fig. 9.2 shows an antenna that is used to transmit the radio signals. Waves leaving this system can travel outward in all directions. We shall focus our attention along two general directions — either along the ground or upward into the atmosphere. The first mentioned wave is generally called the ground wave. At the low frequencies it will travel for distances up to 500 miles or more. The actual area covered depends on several factors, such as frequency, polarization of the wave, and the type of country over which it is moving. So far as frequency is concerned, the best ground wave propagation is obtained at the low frequencies, 20–500 kc., with a steady falling off

as the frequency is raised. This is supposedly due to the rapid rise in ground losses as the frequency increases. The ground wave is used almost entirely by commercial broadcast stations for transmitting programs since they are more interested in local coverage than in reaching people many hundreds of miles away.

With regard to polarization, it has been found at the low frequencies (20–500 kc.) that the radio waves are almost always

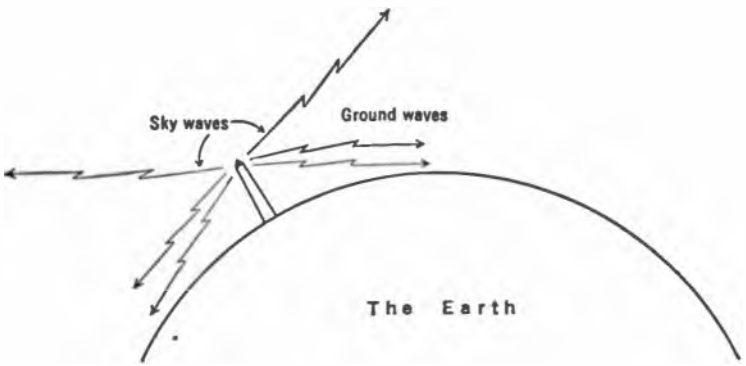


FIG. 9.2. The two general types of radio waves. The ground waves usually die out rapidly while the sky waves travel for long distances.

vertically polarized. This is because a horizontally polarized radio wave would have the electric vector flat along the ground and would be short circuited, the ground being a good conductor at these wavelengths. It has also been found that the ground does not exhibit this property of conductance at all frequencies, generally only up to 10 megacycles. Beyond this frequency the earth exhibits capacitive properties.

Sky Wave — Ionosphere. For the sky wave, investigation has shown that it travels upward at various angles. If the medium did not change sharply, the wave would continue until lost somewhere in space. The sharp change in medium is due mostly to the ionization of gas molecules and atoms that are found in the upper regions of the earth's atmosphere. Ionization consists of the separation of a gas molecule into a positive ion and one or more electrons. It can be brought about by the ultra-

violet radiations that have been sent out from the sun. The ionized layers are responsible for long distance radio communication in the range of wavelengths extending from 10 to 160 meters. The process whereby this is accomplished is due largely to refraction of the radio waves. In reflection, a wave hits an obstacle and bounces back whereas, in refraction, the wave on entering the new medium is bent so much that it finally returns to earth. Both methods are illustrated in Fig. 9.3. Since it is the process of refraction that is responsible for the usefulness of the ionized layers in our stratosphere, it is advisable to investigate more closely the various effects of these ions on radio waves.

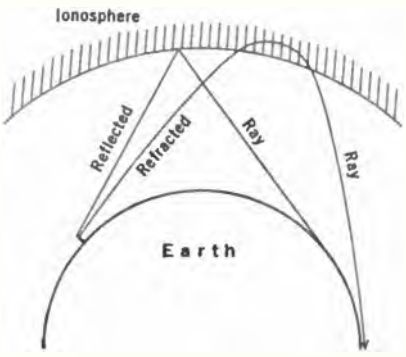


FIG. 9.3. Showing reflection and refraction of radio waves at the ionosphere. Reflection occurs less frequently and then only for low frequency waves.

E Layer. The layers of ions, or ionosphere, as they are collectively called, are found to exist at distances of from 70 to 250 miles above the surface of the earth. Analysis of this region has disclosed that, although there is ionization of the gas molecules throughout the entire area just mentioned, there are distinct layers in which greater concentrations of these ions exist. The layers are separate from each other and so have been given specific designations. The lowest layer, called the E layer, is usually found about 70 miles above the earth, and this height has been shown to be practicaliy constant. Since the formation of the ions in this layer is dependent partly on the amount of ultra-violet rays received from the sun, it is not surprising to find that the density of the E layer varies from hour to hour. The variation is not quite as great here as that in the higher layers, but it can still be detected. The greatest density occurs about noon and the least value occurs during the dark hours of the morning.

F Layer. At higher levels, the next region of maximum ion concentration occurs at a distance of about 185 miles above the earth. This layer has been labeled the F layer, but it can only be found as one layer at night, after the sun has set. During the day it breaks into two separate groups of ions, called the F_1 and F_2 layers. The first named layer is usually found 140 miles above the ground while the F_2 layer occurs at about 225 to 250 miles.

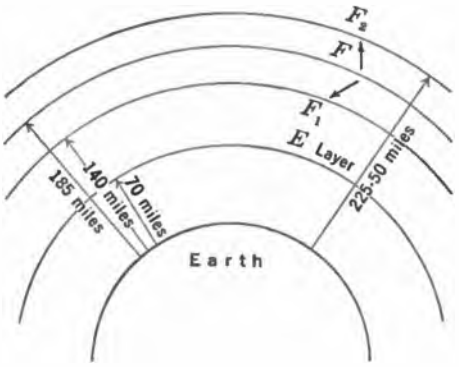


FIG. 9.4. The various layers of the ionosphere. The distances given are only approximate values since these heights fluctuate from day to day.

At night, of course, these two regions combine again to form the single F ionized layer. The three are shown in Fig. 9.4 together with the lower E layer. The F_2 layer has the greatest degree of ionization and electron density. It is this layer that refracts the radio waves of higher frequency, the waves that other layers cannot turn back. If a wave suffers no refraction, or at least not enough bending upon entering this final (F_2) layer, it continues into empty space.

Wave Bending. When the radio waves first enter the lower ionized region, they act on the free electrons present there. Because of this, the wave suffers a bending effect. It is possible, in a way, to visualize the refraction by comparing it with an ordinary light wave leaving one medium and entering another at some angle as in Fig. 9.5. Let the first region be called number 1,

and the second number 2. When point A of the wave from $(A-A')$ reaches the intersection between these two regions, it will increase in velocity because region 2 is rarer than region 1. This is the condition that gives rise to greater velocities. The rest of the wave front, still in the denser area, will continue to travel at the lower speed. Hence the net effect will be that the wave front is swung around. If medium 2 becomes successively less dense as the wave penetrates it, the front of the wave will even-

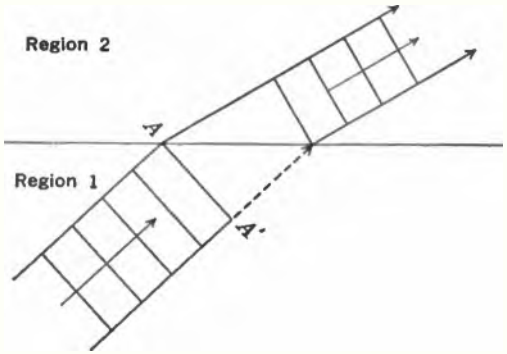


FIG. 9.5. Illustrating refraction of light rays when going from denser to a rarer medium. The same type of conditions may be applied to radio waves.

tually be bent so far around that it will end up facing medium 1 again. It is this condition that is desired, since the wave will now be returned to the first medium whence it came; in the radio case, it is in the earthward direction.

In the ordinary atmosphere, as a radio wave goes up, it suffers only a slight bending due to the increasing rareness of the air. The bending is not great enough in itself to cause sufficient refraction of the wave and make it return to earth. The ionosphere, on the other hand, does have enough refracting power. It is due to this agent that we get our long distance communication. In the ionosphere, the actions of the free electrons are such as to bend the radio waves away from regions of high electron density (which is the upward direction) toward regions of lower free electron density (which is toward the earth).

Whether or not a radio wave will be bent back depends upon three conditions: the angle at which the wave enters the ionosphere, the frequency of the wave, and the electron density of the layer. The importance of the angle may best be seen by reference to Fig. 9.6. The greater the angle ϕ, the more the wave has to be bent, or refracted, in order for it to return to earth. Waves entering at very small angles need only slight bending to have their direction changed enough to cause them to return to the

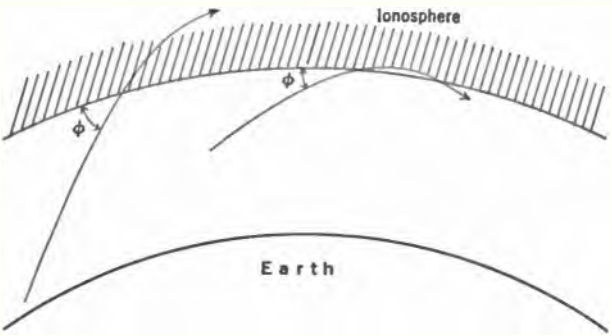


FIG. 9.6. Illustrating that as the angle ϕ which the radio wave makes with the ionosphere increases, more bending must occur before the radio wave is bent back toward the earth.

earth again. The frequency of the radio wave causes an alteration in the vibration of the free electrons in the ionosphere and so it, too, is directly related to the degree of refraction suffered by the radio wave. Experimental observations have shown that the strong refracting power of the ionosphere is directly due to these free electrons. Anything affecting their motion will likewise affect the refractive power of the ionosphere. Finally, the density of the charged particles in the layers themselves will, in the final analysis, determine just how much the wave is bent. As the density increases, so will the amount of refraction increase for any one wave. Although mentioned separately, there is a dependence between the above mentioned conditions, as shall now be shown.

Critical Frequencies of Ionosphere. Radio waves may be sent upward at almost any desired angle, but the wave that will require

the greatest amount of bending is one that is sent vertically upward or, what is the same thing, at right angles to the earth's surface. The frequency of the wave that is bent back after entering a certain layer at this vertical angle is known as the critical frequency for that particular layer. If a wave of higher frequency tries to enter this ionosphere layer at right angles, it will usually not be refracted enough to make it return to earth and will continue upward. However, this wave of higher frequency may enter the ionosphere at a smaller angle, say 50°, and have a better chance of being returned to earth. Of course, if the frequency is very high, not even a small angle of incidence will help, in which case the wave will continue into space. At various times in the day and during the year the density of a given layer will change, whereupon its critical frequency changes. Should the density increase, so would the critical frequency rise in value. If the reverse occurred, it would be necessary to lower the critical frequency. This critical frequency may thus be looked upon as an index of the ability of the ionosphere to return a wave to earth. All lower frequencies will be refracted to earth no matter what the angle used to transmit them upward, whereas higher frequencies may be returned only if the proper angle is used. The lowest possible angles achievable practically are around 4° to 6° with the horizon. If a wave cannot be returned at this angle, sky wave transmission cannot be used for this particular wavelength.

Skip Distance. Now that the normal conditions, which exist when a transmitting antenna emits radio signals, have been discussed, both ground and sky waves may be combined into one figure to determine a composite picture. Refer to Fig. 9.7. It is obvious that unless the sky waves come back to the surface of the earth just where the ground waves end, there will be a region in which signals from this particular antenna cannot possibly be received. The distance from the end of the ground wave to the place where the sky wave first comes back is referred to as the zone of silence, whereas the total distance from the transmitter to this sky wave is called the skip distance. The only remedy for this situation is to change the angle of the transmitting antenna

until a point is found at which the sky waves will be brought down in this area.

Many times sky waves have come down at points well within

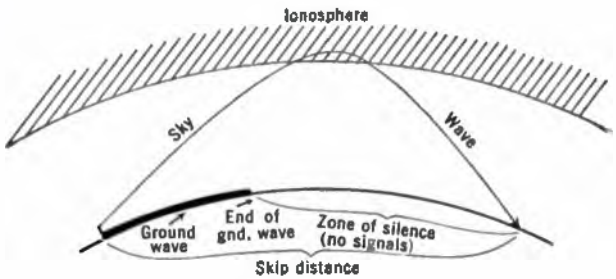


FIG. 9.7. In this figure are shown the conditions that give rise to large areas where no signal from a particular antenna is heard.

the region of the ground wave itself. Then a given antenna receives both these signals at the same time, as shown in Fig. 9.8. This phenomenon usually results in interference, since the paths

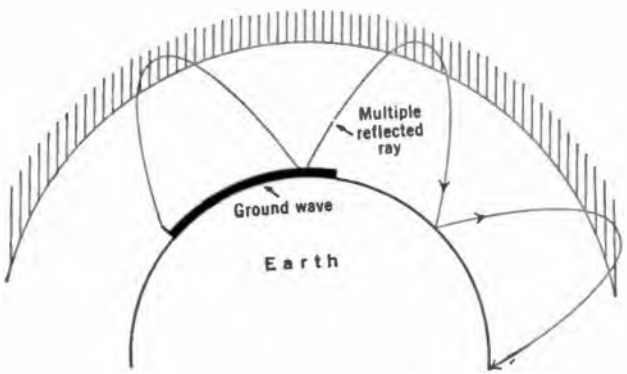


FIG. 9.8. When a large degree of refraction occurs, or when a radio wave enters the ionosphere at large angles, then it may happen that the sky wave will come back at points where there is likewise a ground wave.

followed by the ground and sky waves in reaching any one common point are almost always different in length and give rise to different phase conditions at the time of arrival. Also shown

in Fig. 9.8 is the fact that multiple refractions and reflections can take place before the wave reaches a certain point on the earth. In fact, signals have been known to hop their way around the globe and arrive at their starting point much later. It takes approximately 0.14 second for a signal to go once around a major circle of the earth. Each time one of these refractions (at the ionosphere) and reflections (at the surface of the earth) take place, energy is lost by the traveling wave. After a few of these hops, the wave is much too low in intensity to be picked up. Usually a signal must have a strength of 10 microvolts per meter to be intelligible so far as entertainment is concerned. For code reception, however, values as low as 1 microvolt per meter will suffice. These are the lowest limits. For effective reception, much higher intensities are to be desired.

Summary of Wave Propagation up to 40 Mc. To generalize on the entire field of ordinary communication frequencies (up to 40 mc.), it can be stated that for extremely long distance sending a higher frequency can be used during the day than at night and in summer rather than winter. When the sun has set, it is advisable to lower the frequency, for then the ionospheric layers have a lower electron density and therefore a smaller refracting power. Of course, the frequency could be kept constant and the radiation angle lowered, but usually there are fewer mechanical difficulties involved if the frequency is made the variable.

Sporadic E. The statements just made represent the so-called normal conditions. Under these circumstances, frequencies as high as 40 mc. can regularly be refracted by the ionosphere. It is the F_2 layer, with its greater density, that does the refracting on these higher frequencies. However, every once in a while it is possible to have 56 mc. or even slightly higher frequency waves refracted to earth by the ionosphere. As this is an unusual occurrence, it is necessary to look for conditions in the ionosphere that are present only at these times and then disappear. Upon investigation it was discovered that certian portions of the lower or E layer contained an unusual amount of ions. It is from these

patches that the higher frequencies are refracted. The name given to these small, well defined regions within the *E* layer is Sporadic E spots. These spots occur quite frequently, especially in summer, and are used to good advantage by the amateurs who use the 56-mc. band for their transmissions. The time of the day or night does not seem to have too great a bearing on the matter since long distance communications have occurred at all times of the day.

Another phenomenon of importance that takes place quite often are the intense magnetic " storms " that disturb long distance communications for periods of from 3 to 5 days. During these storms the earth's magnetic field fluctuates in intensity and then gradually settles down to its normal value. These are but two of the many influences that occur and play havoc with our communication systems.

Ordinary Losses in Radio Waves. Coupled with the losses that may be due to the unusual conditions stated above, there is the absorption of energy which ordinarily takes place at various points along the path of a radio wave. The first place to consider is the earth itself, as it is here that the greatest amount of energy loss occurs. For this reason, multiple hop transmission, as shown in Fig. 9.8, is not advisable. By the time the third hop takes place there may not be enough energy in the wave to allow an intelligible signal to be received. The character of the earth's surface determines how much loss occurs at any one place. It is impossible to generalize, because of the great variety of soils that may be found. And to add to the difficulty is the fact that at the lower frequencies the ground is principally resistive, whereas at the high frequencies it seems to act as a dielectric. The ground wave is of very little importance as a means of communication at frequencies above our broadcast band because of the increased attenuation.

The second greatest point of loss in the path of a radio wave is in the ionosphere. There is very little loss of energy in the intervening distance between the earth and the lower or E layer. All the dissipation of energy occurs within these layers themselves.

Ionospheric Losses. To understand why any loss should be introduced by the ionospheric layers, it is first necessary to understand what happens when the radio wave is projected into this region. As the wave enters the ionized medium, the free electrons that are found there in abundance receive energy from the passing wave and begin to vibrate, which in itself represents an extraction of energy from the signal. This energy, known as kinetic energy, as it is one of motion, is entirely returned to the wave except when the electron gives up some to a gas molecule in the vicinity by the process of collision. Any energy lost this way cannot be returned to the radio wave and so represents a loss. It should be emphasized at this point, however, that the mere process of setting of these free electrons in motion by the passing wave does not represent energy loss in itself. It may be looked upon as a loan which is repaid to the wave by the electrons. The loss occurs when something happens to the electrons so that they are unable to give back as much energy as received. One way they can lose energy is by collision with the gas molecules in their vicinity.

The more collisions an electron has, the more energy is lost. The region where the largest number of gas molecules are present is, of course, directly proportional to the atmospheric pressure, and this will be greatest at the lower edge of the lowest layer, namely, the E layer. It is true that, although the electron density increases with height, the pressure becomes less and the possibility of collisions also decreases. From this line of reasoning it is easy to see why the greatest amount of energy loss by a sky wave occurs when this wave enters and leaves the E layer.

U.H.F. Propagation. Up to this paragraph, the discussion has been exclusively about conditions in radio communication that are applicable to frequencies up to approximately 40 mc. This is the highest frequency that will be refracted by the F_2 layer under what may be termed normal conditions. For Sporadic E refractions it is possible to go as high as 56 mc. although then odd conditions are met. While Sporadic E refractions occur quite frequently, they are not to be depended upon for any regular scheduled traffic. From here to the end of this chapter fre-

quencies in the ultra-high region will be emphasized, extending from 40 mc. up. Most of the work done in this portion of the radio spectrum has dealt specifically with the frequencies from 40 to 200 mc. but there are many reasons to believe that, for the wavelengths covered by the 10-cm. waves generated by the magnetron and the Klystron, the same general rules apply.

Three Components of U.H.F. Propagation. Ultra-short wave propagation can be broken down into three categories, each one of which is responsible for the reception of these signals by a different means. First is the so-called "line-of-sight" method which embodies both a direct ray and a reflected ray; secondly, there is diffraction around the curvature of the earth; and lastly there is refraction in the region of the troposphere. The latter is quite different from the inospheric refractions although the end result is the same, namely, bending of the radio wave. These three methods are generally independent of each other. Although it may happen that all of these methods combine to lay down the same transmitted signal at the same point, it does not occur often. Usually the last two modes of propagation are responsible for signals appearing beyond the horizon, whereas the first method deals with distances such that the receiving and transmitting antennas are in a direct line above the curvature of the earth.

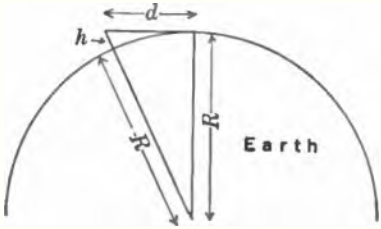


Fig. 9.9. The "Line-of-Sight" method of reception at the U.H.F.

Line-of-Sight Method. The direct ray, Fig. 9.9, has been drawn to illustrate the first method, the direct ray. For purposes of discussion, the height of the antenna will be called h, the radius of the earth R, and the distance from the top of the antenna to the horizon d. The reason for the following simple derivation is to

find just how far a wave could go if sent out from this transmitting antenna, i.e., the direct line route to the horizon. To simplify the derivation somewhat, it will be assumed that the earth is flat over that small portion of its surface. This assumption results in a

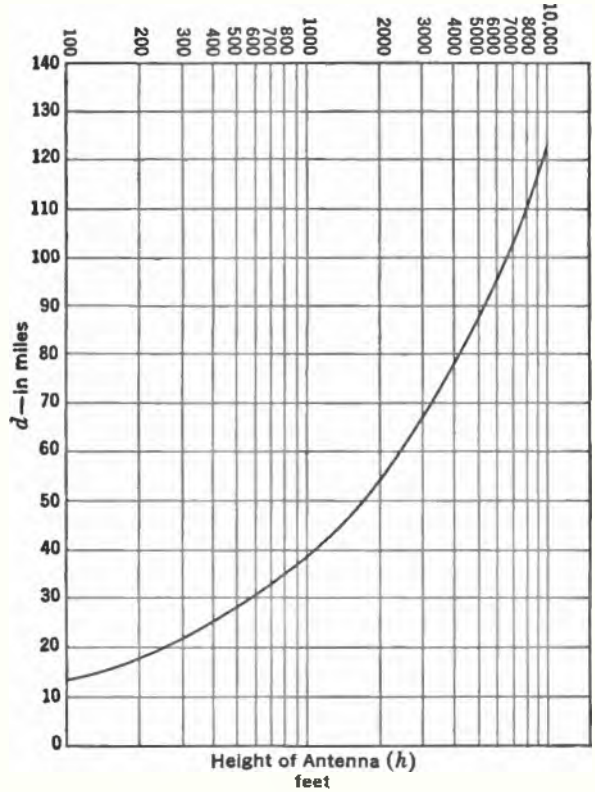


FIG. 9.10. Graph showing relationship between height of antenna and distance of maximum transmission.

right triangle. From elementary geometry it is possible to write the following equation:

$$d^2 + R^2 = (R + h)^2 = R^2 + 2Rh + h^2.$$

h is very small compared to the radius of the earth. Hence, the

h^2 term can be neglected and gives us,

$$d^2 = 2Rh.$$

It is possible to substitute the numerical value of the radius of the earth for R and obtain the following simplified equation:

$$d = 1.23\sqrt{h}$$

where d is in miles and h is in feet. To show the relationship between d and h for various values of h, the graph of the last equation has been drawn and is shown in Fig. 9.10.

It is assumed in the above discussion that the receiving point is at ground level. All that the formula gives us is the distance from the transmitting antenna to the horizon. However, if the

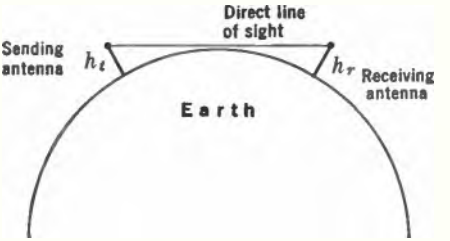


FIG. 9.11. The usable range between antennas can be increased by raising the transmitting and receiving antennas.

receiving antenna is not on the ground but is raised a distance into the air, it should be rather easy to see that the direct line between the two antennas can be made greater before the curvature of the earth again interferes with the direct line. See Fig. 9.11. By means of simple geometrical reasoning, the maximum distance between the two antennas is now increased to

$$d = 1.23(\sqrt{h_t} + \sqrt{h_r})$$

where d is the distance between antennas in miles,
h_t is the transmitting antenna height in feet,
h_r is the receiving antenna height in feet.

Because of the facts just mentioned, transmitting antennas for F-M and television stations are placed atop tall buildings or on

high plateaus. There may be other obstacles in the path of the direct rays which would result in absorption of the signal energy and so tend to weaken or distort the received sound or picture. Then allowance has to be made between the received signal and the calculated value. Such factors must be dealt with separately in each specific location. They assume great importance for television stations where the interfering influences may result in distorted images on the cathode-ray tube screen.

The type of signal laid down at the receiving antenna, however may be due not only to the direct ray just described but, in addition, to a ray which came by way of a reflection from the earth. In Fig. 9.12, the waves leaving the transmitting antenna are broken up into two parts: one wave going directly to the receiving

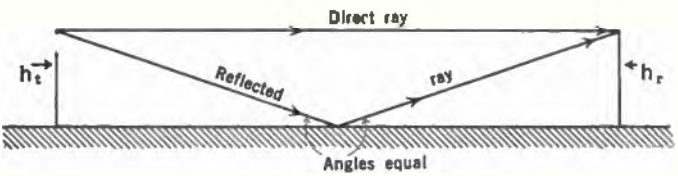


FIG. 9.12. Paths of the two waves from transmitting to receiving antennas.

antenna; the other arriving there after it has been reflected by the ground. The receiver, then, will be subject to both these rays. Whether or not a good clear signal will be heard depends upon the phase relationships of the combining signals. To illustrate this point further, note what happens to the reflected wave. If any change occurs, it can only be due to this portion of the final received signal and not to the direct ray. At the point where the reflected ray impinges on the earth, a phase reversal of 180° has been found to take place and, in addition, an absorption of energy in an amount dependent upon the conductivity of the earth at this point. The phase shift results in a wave at the receiving antenna which acts in an opposite manner to the direct ray. The overall effect is a general lowering of the resultant signal level. However, two conditions are acting against this decrease due to the reflected ray. One is the weakening of the wave due

to the absorption at the point where it grazed the earth, and the other results from an additional phase change (not the 180° just mentioned) arising from the fact that the length of the path of the reflected ray is longer than that of the direct ray. Thus there is a total phase shift of 180° plus whatever else may have been added because of the longer path. These factors all combine to lower the strength of the direct signal to a value less than one would at first expect. It has been found that the received signal strength increases with height of both antennas. At the same time, a decrease in noise pick-up occurs. For F-M and television signals this is most important. With increased height of the antenna, one of the many directive devices described in Chapter 7 is also employed further to increase the intensity of the signal.

Vertical Versus Horizontal Antennas. As to the relative merits of vertical versus horizontal polarization: George H. Brown[12] has found that for antennas located close to the earth the vertically polarized rays yield a stronger signal. Upon raising the receiving antenna about one wavelength above the ground, the difference usually disappears. Then either type can be used. Further increase up to several wavelengths has shown that the horizontally polarized waves give a more favorable signal-to-noise ratio and are therefore to be desired. However, as with everything else in this field, the structure of the terrain may modify the conclusions greatly.

If signals were received only in accordance with the foregoing formulas and reasoning, the study of ultra-high frequency propagation would have ended there. But signals were consistently received at points beyond the horizon, some in great quantity and evidently not due to any unusual phenomena. Further investigation was indicated. Finally the reasons for these added field strengths were laid to two definite causes: one is called diffraction; the other, refraction in the lower atmospheric regions. Each is due to different reasons, as will now be shown.

Diffraction — Second Method. So far as is popularly known, light seems to travel in straight lines. Ask most people whether it is possible for light to bend around a corner and the answer will

invariably be no. Ask the same question about a sound wave and the opposite reply will be obtained. The reason for one and not the other all stems from the observed fact that, where the wavelength of the wave being transmitted is comparable to or larger than the size of the obstacle in its path, the waves will be bent or diffracted around this object. However, as the wavelength gets smaller and smaller, less and less bending occurs until, at the very high frequencies of light, none seems to be present with any ordinary sized objects. Even here it has been demonstrated that, if light from a distant source is sent through a narrow slit, the pattern obtained on the other side of this opening will show that the light did bend in going through. Thus, it is a matter entirely dependent on the relative sizes of the wave and the obstacle in its path. Since the ultra-highs are longer in wavelength than light rays, some bending does take place. The bending, then, is responsible for the fact that the ultra-high signals are heard at points beyond the line-of-sight distance. The waves follow the curvature of the earth for a short distance beyond the ordinary direct ray path of the radio signal. Naturally, as the frequency is increased and the wavelengths decreased, less and less of the bending should occur.

The diffraction effect is quite independent of any weather variations that take place. Hence, the signal heard at short distances beyond the horizon are steady in intensity and can be relied upon for continued use. Although the phenomenon of diffraction increases the distance of propagation somewhat, by far the more important reason for relatively long distance u.h.f. transmission depends on the refraction effects in the lower regions of the air just above us. These effects are not independent of the weather; on the contrary, they are present because of a certain set of conditions in the masses of air that gives us what we call weather.

Refraction in Troposphere — Third Method. For the last method of ultra-high frequency wave propagation, an entirely new set of conditions must be investigated. It is now necessary to deal with the variations in the lower part of the earth's atmos-

phere, that portion which is technically referred to as the troposphere. See Fig. 9.13. The layers of air surrounding the earth may be broken up into three categories. First is the layer that contains most of the air that we breathe. In this region most of the weather conditions take place. It extends above the ground for a distance of 8 to 9 miles. Above the troposphere is found the stratosphere where the air is extremely rarefied and where the atmospheric pressure is low. This region is seldom affected by changing weather conditions; the same, unchanging calm prevails

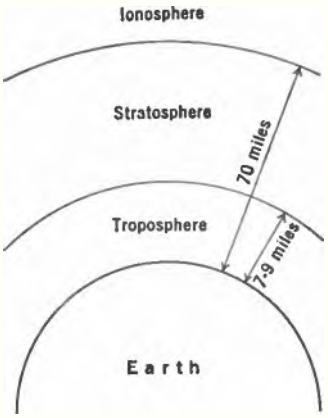


FIG. 9.13. The composition of the earth's atmosphere.

the year around. This region starts where the troposphere ends and continues until the ionosphere is reached, about 70 miles above the earth. The ionosphere, as just described, is responsible for the return of the low-frequency waves to the earth's surface.

To understand how the extremely short radio waves are refracted or bent back to earth in the lower region (the troposphere), it is first desirable to note that, in order for a radio wave to bend in the correct direction (back to earth), several suitable conditions must be present. First, the velocity of the wave must increase as it travels into regions that are less dense. When this condition prevails, that part of the wave front entering the rarer medium will have its velocity speeded up, whereas the remainder of the

wave is still traveling at the slower speed in the denser medium. The result, as demonstrated in the first part of this chapter, is that the wave is bent. This action usually affects all waves (any frequency) traveling upward through the thinning atmosphere, but is not, in itself, sufficient to return the waves to the earth.

Secondly, it has been demonstrated by means of experiments that the velocity of a wave will increase if it travels into a region of increasingly higher temperatures. Now if both of these conditions (higher temperatures and decreasing atmospheric pressure) occur at the same time, it might be possible for the wave to bend sufficiently to return to the ground. Neither effect by itself may be sufficient to accomplish this feat, but both together may be effective in providing the necessary bending or refraction of the radio waves. It might be wondered where these two conditions could occur, as it is common knowledge that as the atmospheric pressure goes down (with increase of distance above the ground) the temperature likewise decreases. This latter combination is the opposite of that desired, and represents the normal situation. Quite frequently, however, other conditions occur which are contrary to the usual run of things and which give the desired combination, namely, decreasing atmospheric pressure and increasing temperatures.

Take, for example, a clear day during which the sun beats down fairly steadily. There is but little absorption of heat as the rays pass through the atmosphere, and the air is only slightly heated. The ground absorbs most of the heat and consequently rises to a comparatively high temperature. After sundown, the ground starts to cool. Then, if the process occurs quite rapidly and to a large degree, it is quite possible that the air above the ground will, after a short period, be warmer than the cooled-off ground. While the ground is becoming cooler, so is the air, but at a much slower rate. Therefore, a condition would be obtained where the temperature increases at increasing distances above the ground. This will aid the refraction of short waves and, as shown by Fig. 9.14, will result in the reception of signals at points below the horizon which the direct rays cannot reach. The weather

conditions that give rise to the desired temperature inversion (increase of warmth with increase in height) occur rather frequently in summer, when the sun is hot during the day and the sky is clear at night.

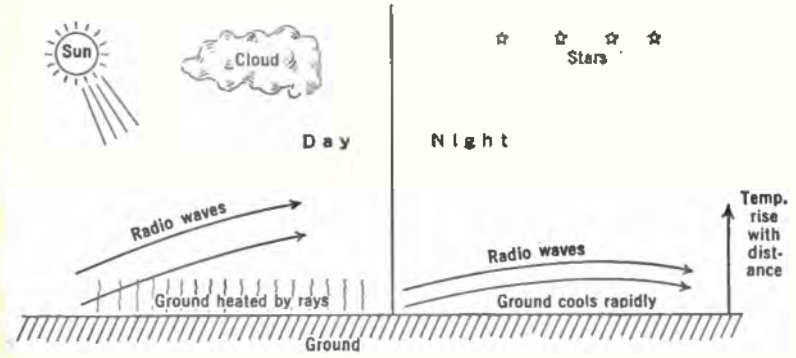


FIG. 9.14. Conditions that give rise to refraction of radio waves in the troposphere.

Air Masses. There is still another set of conditions that will give increasing temperatures with height and that results in stronger signals than those due to the effect described in the preceding section. The temperature inversion is brought about by the ever changing air masses that travel from the Pacific to the Atlantic ocean across this country. If the air masses originate in, or get their characteristics from, the north, cold, clear weather may be expected. It is this kind of air mass that is responsible for the extreme cold that freezes everything throughout the winter night. There is another kind of air mass that may emanate from the warmer climates and which might result in soft, balmy weather mixed with showers. Such masses of air are sometimes responsible for the hot, oppressive heat in the summer when they become stagnant over one portion of the country. In winter they cause what is called unseasonable weather. Both are distinct from each other and may occur at any time. These large air masses are continually on the move in an easterly direction (Fig. 9.15) although it happens from time to time that they remain in one place for several days.

Quite frequently a warm air mass moving forward encounters a cold mass of air which starts pushing into it. The denser, cold air will stay closer to the ground and the warm mass of air will be

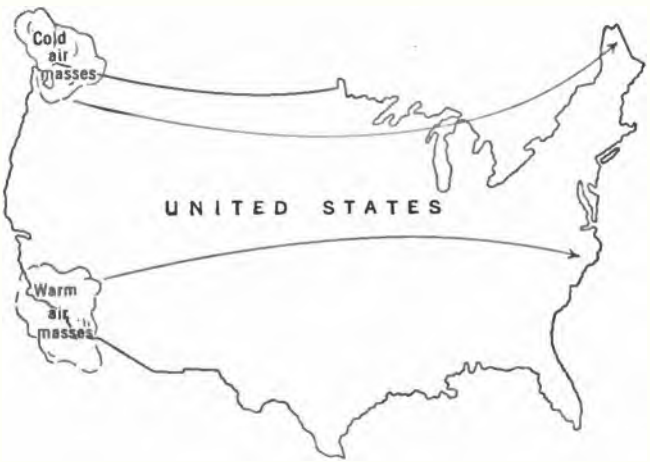


FIG. 9.15. Routes that the air masses take when travelling across the U. S.

pushed upward. The line separating the two is distinct and tends to remain so because there is very little mixing. This situation is shown in Fig. 9.16, with the arrows pointing in the

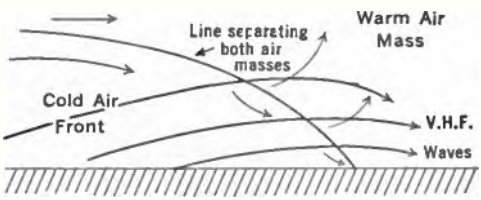


FIG. 9.16. Direction of movement of a warm air mass when overtaken by a cold air mass.

direction of movement of both the cold and warm air masses. Notice, also, that the line separating the two types of weather is not vertical but tends to assume an angle with the ground. The result of all this maneuvering on the part of the air masses has given us something that may be quite useful, namely, a tempera-

ture inversion. The warm air is now higher above the ground than the cold air. Ultra-high frequency waves coming up through the cold air from the transmitting antenna on the ground will meet the warmer temperatures above and so be refracted earthward, somewhat in the manner shown in Fig. 9.16. At 100 mc., transmission of signals has been obtained over distances as great as 225 miles. The amount of bending that occurs depends upon several factors, such as:

1. Sharpness of the line separating the cold from the warm air.
2. Height of this line of separation above the ground.
3. Angle which this line makes with the earth.

Should a cold mass of air be moving forward and then be overtaken by a mass of warm air, the conditions shown in Fig. 9.17 can take place. The lighter, warm air will rise above the cold air and again a temperature inversion occurs. The resulting conditions occur most frequently in the summertime, for the air masses tend to move slower than in winter, and conditions are more favorable for wave bending over a large area.

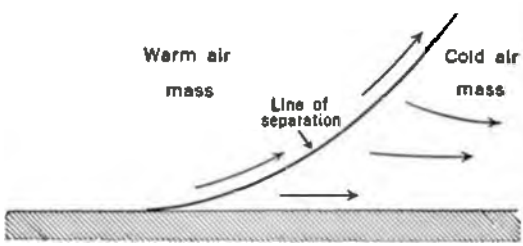


FIG. 9.17. A warm air mass overtaking a cold air mass.

It may be wondered how the foregoing information may be put to practical use. One way, and it is the best, is to get a daily weather map from the local office of the weather bureau. The map shows conditions all over the United States for that one day, giving the specific locations of the various air masses. At the center of each air mass may be found the words " low " or " high." The word " low " refers to air that is relatively warm and moist. Air coming from the north is usually cool and dry and labeled

" high." If a series of successive weather charts are obtained and studied, the progress of the various highs and lows can be followed across the continent. By studying the local weather conditions and then tying them in with the general conditions, as shown by the weather map, it is possible with a little practice to predict times when conditions will be suitable for ultra-high bending.

The preceding paragraph is a sketchy outline of air mass phenomena. More detailed information can be understood only after some knowledge of elementary meteorology has been acquired. It is suggested that some elementary book be read, followed by the series of reference articles given at the end of this chapter. The articles do not represent the final conclusions on this subject, but are merely accounts of several experiments carried out at the ultra-highs. Undoubtedly more information will be forthcoming as the ultra-high frequencies gain in prominence.

Before concluding this chapter, there are several points that may be added, points that may help our understanding of what occurs when radio waves leave the transmitting antenna. The first concerns diffraction and refraction. Although mentioned separately heretofore, there is reason to believe that they may both occur with the same wave. The waves, after having been refracted in the troposphere, do not come back to earth and then end abruptly. The downward deflection of the signal is gradual. This points to a situation where refraction brings the radio beam to some distant point, beyond which the wave progresses a little further because it is bent around the curvature of the earth by diffraction. Thus diffraction acts in two places: once in connection with the direct ray and once with the relatively long distance refracted ray. In each case the signal is heard at points beyond the place where it would ordinarily be received if the diffraction process were absent.

Next, we find that, although bending of radio waves is possible in the lower atmosphere, such bending occurs in appreciable amount only for those rays that make small angles with the ground. It is doubtful if refraction would take place for waves

sent up at angles that exceed 20° or so. For all u.h.f. waves, the most consistent method to use is the line-of-sight method. Under nearly all conditions it will prove reliable.

Summary. Radio waves, in their travels from transmitting to receiving antennas, are usually attenuated by the various objects with which they come in contact. In order to study these effects, it is first desirable to differentiate between the two general types of waves found. One is called the ground wave and generally dies out rapidly, especially at the higher frequencies. It probably finds its greatest application at the broadcast frequencies (550–1500 kc.), thereafter decreasing sharply. Occasionally, police radio systems utilize the ground wave at the F-M frequencies, but only for very short distances.

The sky wave is responsible for the long distance communication in the band of frequencies from 1500 kc. up to 20 mc. This sky wave travels upward at some angle to the ground until it hits the region above the earth known as the ionosphere. This layer consists of great quantities of ionized atoms, due largely to the ultra-violet rays from the sun. The ions and, especially, the free electrons are excited by the passing radio waves and, due to their complicated motions, act in such a manner as to bend or refract the wave back to earth. This process, however, also gives rise to absorption of energy by the vibrating electrons. Hence, the radio waves suffer attenuation. This loss is greatest at the lower, or E layer, thereafter decreasing with height. At the highest layer, the F_2, the waves of highest frequency (about 40 mc.) are refracted, with only slight energy loss.

The second great loss of energy occurs when the sky waves hit the earth on the downward voyage. The process of reflection takes place and generally introduces a loss which is dependent on the type of soil present.

In order to understand the bending of waves of frequencies above 40 mc., it is necessary to look to other phenomena, one of which is the Sporadic E refraction. Due to reasons not yet definitely established, patches of ions occur in the lower E layer which have very high electronic densities. These patches, or

spots, cannot be generally predicted in advance. When they do occur, they reflect and refract frequencies up to 56 mc.

Frequencies higher than those just mentioned have been found to be propagated by three general methods.

1. The line-of-sight method, which gives rise to the strongest and most consistent signals yet observed. This method can be relied on at all times, very little change occurring because of varying weather conditions. All frequencies may use this method.

2. The diffraction method, whereby the radio waves have a tendency to follow the earth for some distance beyond that which can be obtained by the first method. As the frequency increases, however, this effect becomes less and less noticeable.

3. The last method described dealt with air masses and the conditions necessary for refraction to occur in the lower atmosphere, i.e., that portion of the air which we call the troposphere. The conditions are sometimes predictable in advance after careful study of the weather maps. The frequency limit of this method is as yet unknown.

Further reference articles on the refraction of u.h.f. waves by air masses.

1. Ross HULL, "Air Mass Conditions and the Bending of Ultra-High Frequency Waves." *Q.S.T.*, vol. 19, p. 13, June 1935.

2. Ross HULL, "Air Wave Bending of Ultra-High Frequency Waves." *Q.S.T.*, vol. 21, p. 16, May 1937.

3. M. S. WILSON, "Ultra-High Frequencies and the Weather." *Radio*, p. 33, January 1940.

QUESTIONS

CHAPTERS 1-9

CHAPTER 1

1. Name the two defects that prevent ordinary apparatus from being used at the ultra-high frequencies.

2. What is meant when it is stated that the Q of a coil or tuning circuit is low?

3. How long should a short-circuited Lecher wire system be in order to act as a resonant circuit? Why?

4. Draw the circuit of a push-pull, tuned-plate, tuned-grid oscillator using Lecher wires.

5. What is meant by the transit-time effect?

6. What are some of the more common ways that can be used to overcome its effect in ordinary tubes?

7. Describe the revised ideas on electron flow.

8. Can currents flow in the grid or plate circuits if no electrons hit these elements? Explain.

9. How does the Barkhausen and Kurz arrangement of a tube differ from the conventional applications?

10. Describe the operation of the B-K oscillator when there is a constant potential on the elements.

11. Do the same as in question 10, except assume an alternating potential on the grid.

12. Where does the electron get the energy that it gives to the grid?

13. What are the disadvantages of the B-K oscillator?

14. Draw a diagram of a B-K oscillator.

15. Why can a tube of the " acorn " size generate waves of higher frequency than a conventional larger tube?

CHAPTER 2

1. Describe the action of an electron in a uniform magnetic field.

2. How does the magnetron differ from the previously mentioned tubes?

3. Name the two types of magnetrons which are used for the high frequencies. Which can generate the higher frequencies?

4. Define negative resistance and demonstrate by the use of characteristic curves how it can be obtained from some tubes.

5. How does a negative resistance aid a coil and condenser combination to oscillate?

6. Describe the action in the negative-resistance magnetron oscillator.

7. Would this type of oscillator function without the magnetic field? Explain.

8. Why does the negative-resistance magnetron fail at the high frequencies?

9. Explain the action of the transit-time magnetron oscillator.

10. What are some of the difficulties encountered with this latter type of oscillator?

11. Why must the energy-giving electrons be removed from the interelectrode space?

12. What type of tuning unit is used with the transit-time magnetron oscillator?

13. Draw a diagram of this type of oscillator.

14. Do all electrons give up energy to the oscillating circuit? Explain.

15. Where do the electrons obtain the energy that they give to the plates?

CHAPTER 3

1. Describe the tuning unit used with a Klystron oscillator.

2. How is this unit excited?

3. What are the first set of grids in the Klystron called? What do they do?

4. What are the second set of grids called? Are they placed near or far from the first set of grids? Why?

5. Explain what occurs in the drift space.

6. What is the arrangement of the electrons when they arrive at the second set of grids?

7. Where is the d.c. energy in the beam expended?

8. Give the requirements that the catcher grids must satisfy for proper operation.

9. What does the Applegate diagram illustrate?

10. How is the energy fed back from the second pair of grids to the first pair?

11. Describe the various phase shifts incurred in the Klystron, considering its use as an oscillator.

12. Draw a diagram of this type of oscillator.

13. What type of power supply is needed here? Why?

14. Describe the action of the reflex Klystron oscillator.

15. Draw its diagram.

16. How may the frequency of this device be altered? Is it very great?
17. Where is energy removed from this oscillator? How?

CHAPTER 4

1. What are the common uses for a transmission line at the low frequencies?
2. Name the components of a transmission line? How are these usually stated?
3. Define characteristic impedance and explain what happens when a line is terminated in its characteristic impedance.
4. How can the action of an infinite line be simulated?
5. What other factors may influence a signal as it travels down a line? How?
6. What occurs to a signal if a line is not terminated in its characteristic impedance?
7. Draw the voltage, current, and impedance relationships on a quarter-wave line, both open-ended and shorted.
8. Why can a quarter-wave line act as a resonant circuit? Describe, using the two types of terminations.
9. How does a quarter-wave line introduce a phase shift? Explain either graphically or mathematically.
10. Name two other uses for these quarter-wave lines.
11. How does the action of a half-wave line differ from that of a quarter-wave line?
12. Could a half-wave line, either open-ended or short-circuited, be used in place of the quarter-wave line for tuning purposes? Explain.
13. Why is it desirable to use these lines in place of coils and condensers at the ultra-high frequencies?
14. What factors determine whether a line will act as either a capacitance or inductance at any one frequency?
15. For question 14, does the end termination make any difference? If so, how?
16. Draw a diagram of a high-pass filter using appropriate sections of transmission lines.
17. Do the same for a low-pass filter.
18. How do amateurs sometimes use lines to eliminate standing waves on antenna feeders? Illustrate.
19. Which is more desirable, ordinary open-wire transmission lines or coaxial cables? Why?
20. What phase shift is occasioned by the use of a half-wave line? Three-quarter wave line?
21. Draw the voltage and impedance relationships on a half-wave line that is open-ended. Short-circuited.

CHAPTER 5

1. What are the advantages of wave guides over transmission lines at the ultra-high frequencies?

2. What restrictions do wave guides impose on radio waves, restrictions that are not present on waves that travel in free space?

3. What is the importance of polarization of radio waves in wave guides?

4. What is the significance of electric and magnetic lines of force in wave guides? How are they shown?

5. Describe electromagnetic wave travel in guides.

6. What is meant by cut-off frequency and what is its significance with respect to wave guides?

7. Explain the relationship between the notation for these waves and longitudinal and transverse vibrations.

8. Draw an end view of a guide showing the electric field distribution for a $TE_{0,2}$ wave.

9. For the same guide, show the field distribution for a $TE_{0,1}$ wave.

10. For the excitation of TE waves, how should the antenna be placed in a guide? Why?

11. Explain the TM type of antenna.

12. How do the above two antennas differ (a) with respect to placement? (b) with regard to length?

13. How are the coordinates modified for cylindrical wave guides?

14. What factors determine the cut-off frequency of any guide?

15. Name several uses for wave guides.

16. In what respects are wave guides similar to transmission lines and in what respects do they differ?

17. Do the electric field distributions differ for cylindrical wave guides from those shown for rectangular guides? Illustrate.

CHAPTER 6

1. Which are more flexible with regard to frequency changing: cavity resonators or transmission lines? Why?

2. How may we derive a cavity resonator from a transmission line?

3. How can a cavity resonator be formed from a wave guide?

4. What is the shortest length of a resonator for any one frequency? Why can it not be shorter?

5. In a cavity resonator, what would correspond to a large current? A large voltage?

6. How are cavity resonators excited? (Two methods.)

7. Explain the reason for the narrow section of the Klystron cavity resonator.

8. Describe how energy from a resonator may be removed by inductive coupling.

9. Do the same as in question 8 for the case of capacitive coupling.

10. What modifications are necessary in the ordinary definition of Q when it is applied to cavity resonators?

11. How can the above definition of Q help us design a resonator for best results?

12. What type of probe is used for capacitive coupling? Inductive coupling?

13. Describe how an electron or group of electrons can excite a resonator.

CHAPTER 7

1. How are radiation patterns laid out?

2. What is their purpose?

3. Will the same antenna give rise to the same pattern under all conditions? Explain.

4. What is the simplest nondirectional radiator? Is this true for both the horizontal and vertical planes? Illustrate.

5. What is the difference between driven arrays and parasitic arrays?

6. Why are arrays to be preferred over the simple single-wire antenna?

7. Give the characteristics of a parasitic array formed with one driven element, one reflector, and two directors. Draw a diagram of the layout showing correct distances between elements.

8. What is the effect of adding more parasitic wires to an array?

9. Name two systems of units that may be used to indicate the relative gain of a directional array.

10. How does a parabolic reflector function?

11. What precautions must be observed when using parabolic reflector systems?

12. What are the advantages of a corner reflector over the parabolic reflector?

13. Were the directional properties of wave guides increased by the use of horns? Explain.

14. Describe the general effects of flare angle on the directional properties of electromagnetic horns.

15. Do the same for the length. How are the two effects correlated?

16. Draw a sketch of a biconical horn and discuss its directional effects.

17. In the biconical horn, what are the effects of the flare angle and length of the sides? Explain in terms of horizontal and vertical plane radiations.

18. Where might the radiating antenna be placed for use with any of the electromagnetic horns? What adjustments should be made for best transmission?

19. What might be some advantages that multiunit horn arrays could offer over a single horn?

CHAPTER 8

1. Explain how skin effect causes the resistance of any wire to change with frequency.

2. What effect does change in frequency have on inductance? Why?

3. Name some factors that would cause low-frequency instruments to become inaccurate at the ultra-highs.

4. What is meant by leakage conductance and why is it important?

5. Describe how an ordinary wavemeter works. Why does this instrument become unsuitable for u.h.f. work?

6. Explain how transmission lines are utilized for wavelength measurements at the ultra-high frequencies.

7. Draw a diagram for a set-up using a transmission line wavemeter.

8. Name several different voltmeters that are suitable for u.h.f. work.

9. Why is the ordinary a.c. voltmeter limited to relatively low frequencies?

10. Describe the action of a thermocouple and show how this may be applied to u.h.f. voltage measurements.

11. In order to use the low-frequency calibrations, what precautions in thermocouple construction should be observed so that the same calibrations hold at the high frequencies?

12. What is meant by ambient temperature? Explain what effect a change in this temperature would have on a thermocouple meter movement?

13. Draw the circuits of a diode and of a triode vacuum tube voltmeter. Why are not pentodes, as such, used in vacuum-tube voltmeter arrangements?

14. What two effects tend to make a vacuum-tube voltmeter inaccurate at the ultra-high frequencies? Explain each one in detail.

15. What is the common principle upon which many u.h.f. power meters are built? Does this differ from low-frequency meters? Explain.

16. Explain one method that might be employed to calibrate a voltmeter — at least for relative readings.

17. Name several u.h.f. power measuring devices. What precautions must be observed in the construction of the power dissipating element?

18. Explain the operation of special P-M tubes.

19. What property of a transmission line permits its use for the measurement of power?

20. How may vacuum-tubes be utilized for u.h.f. power measurement? What are the requirements for the anode voltages in this case? Why?

CHAPTER 9

1. Describe the convention regarding polarization of radio waves.
2. By what two general methods are radio waves propagated? Which is used for broadcast purposes (550 to 1500 kc.)?
3. Of what use is the ionosphere?
4. Name the three general layers in the ionosphere? At what average heights can they be found?
5. Why is the ionosphere useful for long distance communication? Which layer is responsible for the usefulness of 40 mc. waves? Why?
6. What is the difference between reflection and refraction?
7. Name three conditions upon which ionospheric refraction depends.
8. What is meant by the critical frequency of an ionospheric layer?
9. When does the zone of silence occur?
10. Explain how the ionosphere refracts radio waves.
11. What is the Sporadic E layer? What is its significance?
12. Give an account of the ordinary losses in radio waves brought about by natural causes.
13. Is u.h.f. propagation different from low-frequency radio wave travel? Explain.
14. Name the three methods whereby u.h.f. waves are propagated.
15. Upon what does the most consistent and reliable method depend?
16. How do the prevailing weather conditions determine how far u.h.f. signals may be heard?
17. Which method is least reliable? Why?
18. What is meant by diffraction? How does it affect u.h.f. radio communication?
19. What are the conditions required in the troposphere for the bending of radio waves?
20. What must take place in the ionosphere before it is possible to refract u.h.f. waves?
21. What is the range of frequencies that can be normally refracted by the ionosphere?

References

[1] Ferris, W. R., *Proceedings of Institute of Radio Engineers*, Volume 24, p. 82, Jan., 1936.

[2] Hull, A. W., *Physical Review*, Volume 18, p. 31, July, 1921.

[3] Kilgore, G. R., *Journal of Applied Physics*, Volume 8, p. 660, Oct., 1937.

[4] Linder, E. G., *Proceedings of the Institute of Radio Engineers*, Volume 27, p. 732, Nov., 1939.

[5] Kraus, J. D., *Proceedings of Institute of Radio Engineers*, Volume 28, p. 513, Nov., 1940.

[6] Southworth, G. C. and King, A. P., *Proceedings of Institute of Radio Engineers*, Volume 27, p. 95, Feb., 1939.

[7] Barrow, W. L., and Lewis, F. D., *Proceedings of Institute of Radio Engineers*, Volume 27, p. 41, Jan., 1939.

[8] Barrow, W. L., Chu, L. J., and Jenson, J. J., *Proceedings of Institute of Radio Engineers*, Volume 27, p. 769, Dec., 1939.

[9] Barrow, W. L., and Shulman, Carl, *Proceedings of Institute of Radio Engineers*, Volume 28, p. 130, Mar., 1940.

[10, 11] Nergaard, L. S., *R.C.A. Review*, Volume 3, p. 156, Oct., 1938.

[12] Brown, G. H., *Electronics*, Volume 13, p. 20, Oct., 1940.

INDEX