RADIO ENGINEERING PRINCIPLES

BY

HENRI LAUER, B.S.,

ASSOCIATE MEMBER AMERICAN INSTITUTE OF ELECTRICAL ENGINEERS
FORMERLY LIEUTENANT, SIGNAL CORPS, U. S. A., ASSISTANT IN THE PREPARATION
OF TRAINING LITERATURE ON RADIO THEORY AND EQUIPMENT

AND

HARRY L. BROWN, B.E.E.,

MEMBER AMERICAN INSTITUTE OF ELECTRICAL ENGINEERS; FORMERLY CAPTAIN, SIGNAL
CORPS, U. S. A., IN CHARGE OF THE PREPARATION OF THE TECHNICAL
TRAINING LITERATURE USED IN THE SIGNAL SERVICE

FIRST EDITION
SECOND IMPRESSION

McGRAW-HILL BOOK COMPANY, INC.
NEW YORK: 239 WEST 39TH STREET
LONDON: 6 & 8 BOUVERIE ST., E. C. 4
1920

THE MAPLE PRESS YORK PA

Respectfully dedicated to the Signal Corps, United States Army, and to the many colleges throughout the land working with it, in the hope that the splendid development of radio communication inspired by the necessities of war may continue to the advancement of mankind in peace.

FOREWORD

I am greatly pleased with the manuscript of "Radio Engineering Principles" and recommend it as a work which brings the study of this important subject up to date. It covers fully and clearly without too great use of mathematics, the theory involved in the wonderful developments in the Art of Radio Communication made during the war, except for certain reservations which it is not possible to release at this time.

Major General,
Chief Signal Officer of the Army.

War Department,
Office of the Chief Signal Officer,
Washington, D. C.

PREFACE

This book is written in an effort to meet the need and great demand for a general textbook on radio covering the new and extensive developments in the art made during the war. It is therefore devoted very largely to the study of the characteristics and use of the three-electrode vacuum tube in radio telegraphy and radio telephony, since it is around this device that the present and future of the science seem to center. But to meet the requirements of a general textbook, the principles involved in the older radio apparatus are also treated with sufficient fulness to inform the student on all essential principles of wireless communication.

The authors have presumed their fitness to write such a book because of their peculiar contact during the war with the new developments accomplished in this country and abroad, though not all of these developments, nor indeed the most wonderful of them, are herein published because of the wishes of the military authorities to keep them secret.

In the detail development of the principles involved, for which credit is due the first-named author, the electron theory is made use of frequently as it often gives a clearer conception of what takes place under certain conditions. Mechanical analogies are avoided. Mathematics is resorted to only to indicate the application in the problems of design, or the relations, in concise form, existing among the various quantities of a radio circuit. The description of any specific apparatus is purposely avoided, with the object in mind of devoting the entire space of the book to the principles involved, though the general means of utilizing these principles in practical work are of course given. With the principles understood it is a simple matter to apply them to any specific radio set.

The authors will welcome and greatly appreciate having their attention directed to any errors of omission or commission which may occur in this first edition.

THE AUTHORS

New York City
Nov. 11, 1919.

TABLE OF CONTENTS

Chapter I. Underlying Electrical Theory

Chapter II. Properties of Oscillatory Circuits

CHAPTER VII. THE THREE-ELECTRODE VACUUM TUBE AS A DETECTOR

CHAPTER VIII. THE THREE-ELECTRODE VACUUM TUBE AS AN AMPLIFIER

CHAPTER IX. THE THREE-ELECTRODE VACUUM TUBE AS AN OSCILLATOR

CHAPTER X. RADIO TELEPHONY

CHAPTER XI. DIRECTIONAL RADIO AND LOOP ANTENNÆ

CHAPTER XII. MISCELLANEOUS APPLICATIONS OF RADIO CIRCUITS

LIST OF PLATES

RADIO ENGINEERING PRINCIPLES

CHAPTER I

UNDERLYING ELECTRICAL THEORY

In the explanation of a number of phenomena underlying the theory and operation of radio apparatus, it is almost essential that the student have at least a general knowledge of the so-called electron theory of matter and the interpretation in terms of that theory of the phenomena of electric conduction and induction. This becomes of especial interest for the study of the latest developments of radio apparatus such as the three-electrode vacuum tubes, where, for the first time, use is made of electricity having no material support, of electrons traveling in space empty of all matter. Under such conditions, the functioning of the apparatus is best and most fully understood by the use of the electron theory.

For the above reasons, the first chapter of this book differs from the usual text book in that, assuming the reader to be already familiar with the usual electrical phenomena, it gives a rapid review of certain elementary notions, fundamental to radio, presented in the somewhat different light of the electron theory, and taking into account the effects of energy radiation. All these questions have been treated as simply as possible, for the discussions in this chapter are intended only to give a physical conception of the phenomena.

Electricity and the Ether

The observation of electrical phenomena has led to the assumption that there are two kinds of electricity, called positive and negative, for the reason that their effects are exactly opposite. The existence and identity of electricity can hardly be doubted today, since negative electricity has been obtained and isolated in a state separate from all matter. Positive electricity however has as yet only been observed fixed on material bodies, or associated with negative electricity.

The existence of electricity is evidenced by its effects. Masses or charges of electricity can be made to react upon each other, to attract or repel each other, and the actions of these forces have been observed to take place in a vacuum between isolated charges as well as in material media and between charges supported by material bodies. Since it is impossible to conceive any action whatsoever to be transmitted from one point of space to another without some intervening agent or medium, these observations have led to the assumption of a medium, called the ether, present throughout space, and even, as will be seen later when studying the constitution of matter, between the atoms of material bodies. For convenience in explanation, the ether is thus assumed to be the agent conveying the actions of the forces between various separate bodies or charges.

ELECTRICAL PHENOMENA IN VACUUM

Coulomb's Law.—In order to study the action of two electric charges upon each other, the simplest case will be taken up first; namely, that of *two charges, concentrated at two points, isolated from all material supports, and placed in empty space,* the only medium between them being the ether. Under these conditions, m and m' being the respective values of the charges, that is, the quantities of electricity making up each charge, and r the distance between the two charges, Fig. 1, each charge will exert upon the other a force f along the line mm' and numerically equal to

$$f = \frac{mm'}{r^2} \tag{1}$$

This law, which was discovered experimentally and can be demonstrated mathematically, is known as Coulomb's law.[1] It will be noticed that if both charges are of the same polarity, that is, both positive or both negative, the product mm', and therefore the value of the force f, will be positive, while it will be negative when the charges are of opposite polarities. This corresponds, in the former case, to

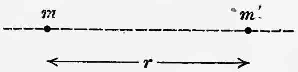


FIG. 1.

[1] For a mathematical demonstration, see S. G. Starling, *Electricity and Magnetism,* p. 117 in the 1912 edition.

a repelling force, which tends to set the charges in motion away from each other; in the latter case, the effect is opposite, there being an attraction between the two charges, which tends to move them toward each other.

Field Intensity.—Coulomb's law may be used for defining the unit of charge or of quantity of electricity. Thus the unit charge is that charge which, when placed at a unit distance from an equal charge, is repelled by the latter with unit force. In the practical system of units, this unit charge is called the coulomb.

The force exerted by a charge m on a unit charge placed at a distance r from it will then be equal to

$$f = \frac{m}{r^2} \tag{2}$$

This force is called the *intensity* of the electric field due to the charge m, at the point of location of the unit charge.

If the charge m is considered isolated in space, the above equation will express the intensity of the electric field due to the charge at any point distant r from it. Thus, the field intensity is the same for all the points of the sphere of radius r and having the charge m at its center. The direction of the force at each point of the surface, is normal to the latter, and is along the radius of the sphere, passing through the point considered.

Potential.—By definition, the work done by the force of the field of a charge m in moving a unit charge from a point A to a point at infinite distance from m is the *potential* at the point A.

From this it follows that the *difference of potential* between two points of the field is equal to the work of the force of the field done in transporting a unit charge from one point to the other. If the two points are at a distance from each other equal to the unit of length, and the work done for moving a unit charge from one point to the other is equal to the unit of work, then the potential difference between the two points is the unit of potential difference. In the practical system of units, this unit is called the volt.

From the above remark that the field intensity is the same for all points of a sphere having the charge m at its center, it may be seen that all the points of that sphere are at the same potential.

Electrostatic Field.—Coulomb's equation expresses the effect of a stationary charge m in space on any other charge, and if the latter is the unit charge, then Coulomb's equation expresses the field of the charge m at any point of space. That is, it expresses the peculiar state of the medium, in this case, of the ether, due to the existence of the charge m, whereby the medium has the ability to exert a force on any other charge which may be present, and produce work by displacing the latter. Thus, by placing a

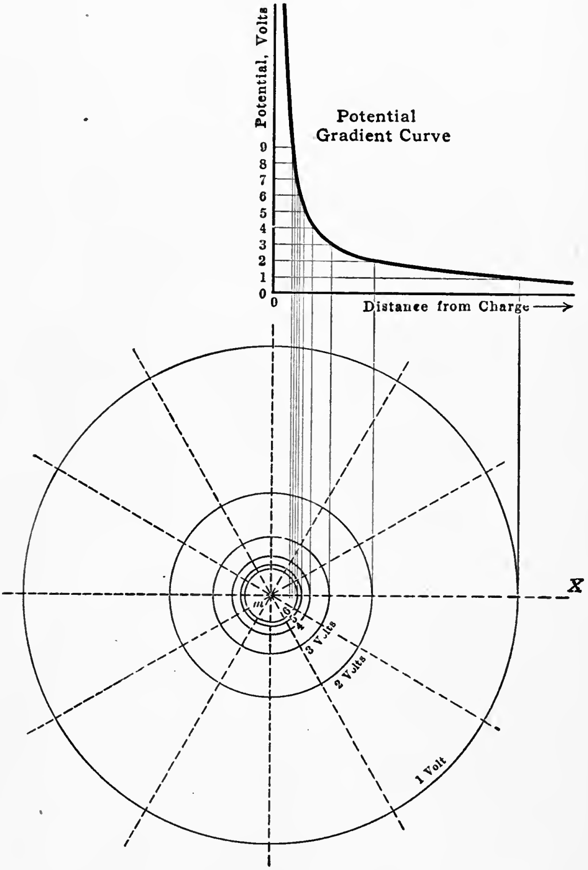


Fig. 2.

charge at some point of space, the surrounding medium is, so to speak, stressed, and a certain amount of *potential energy* is stored and distributed in it, this potential energy being transformed into work whenever some other charge is placed in the field and allowed to move under its action. As explained in a previous paragraph, the amount of potential energy available at any point, that is, the value of the potential at that point, depends on the distance of that point from the charge producing the field, and on the value of the charge.

In order to give an adequate graphical representation of the electrostatic field, it is then necessary to indicate (*a*) the direction of the force at each point, and (*b*) the potential at each point of the field. In practice, these are indicated for a limited number of points only, the values at other points being obtained by interpolation.

The electrostatic field of a single charge *m* is thus represented in Fig. 2. The direction of the force at any point is along the straight line joining that point to the charge *m*. This is shown

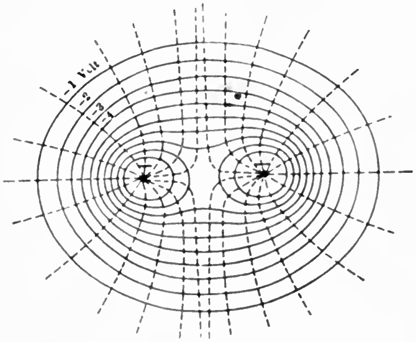


FIG. 3.

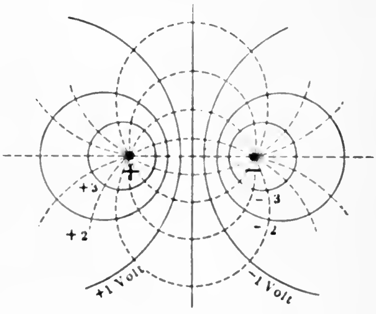


FIG. 4.

for various points by the dotted radial lines. The potential is best shown by means of equipotential surfaces. These are plotted by first obtaining the potential gradient as expressed by Coulomb's equation along a radial line of the field, for instance *mX*. Coulomb's equation of the particular field considered is first represented by a potential-distance curve, from which the distance corresponding to any given potential may then readily be obtained. These distances are the radii of the corresponding equipotential spheres. This has been done here for potentials of 1, 2, 3, . . . volts, and the circles drawn are the intersections of the plane of the paper with the equipotential surfaces. By rotating the entire figure around any one of the radii passing through the charge, these circles describe the corresponding equipotential surfaces.

It will be noticed that the distance between two consecutive spheres is not the same. This is an immediate consequence of Coulomb's law.

Another important conclusion is the fact that the product of the value of potential by the radius of the corresponding equipotential surface is a constant, whichever surface is considered. This means that the charge m, instead of being all concentrated at a central point may be uniformily distributed on an equipotential surface, without disturbing the shape or strength of the external field, and the equation of this field remains the same, provided that the distances r are counted not from the charge, but from its geometrical center. Thus in Figs. 3 and 4 are shown the resultant fields of two charges of the same and of opposite polarities. The direction of the force at each point is tangent to the dotted lines, and the potential is given by the equipotential surfaces.

Variation of the Field with the Charge.—The above considerations were based on the assumption that the charge m creating the field, was fixed in value and fixed in its position in space. It is possible to conceive that the charge m is suddenly changed to some other value m'. This disturbance at the center of the field will then propagate itself outward from that point throughout space until the field at each point has reached the value required by Coulomb's law for the charge m'. However, due to the properties of the ether, this change of the field intensity resulting from the change of the charge, does not occur simultaneously at all points of space. It first takes place at the points of space in the immediate vicinity of the charge, the disturbance in the field at those points being then communicated to immediately adjacent points, and so on outward throughout space in all directions.

This is somewhat analogous to the propagation of a compression wave in an elastic homogeneous medium, such as the propagation of sound in air. It is a familiar fact that when a whistle is started blowing, a man 500 feet away will hear it start about ½ second after the actual beginning of the sound, while a man 1000 feet away will hear it one second later. Thus, until the sound wave has reached some given point, the air at that point will be undisturbed, despite the fact that the sound wave has been started and is moving in its direction.

Due to the similarity of the propagation of the electrical disturbances with that of sound in air or of a ripple or wave on the surface of the water, this has been called an electric wave. Also, a wave or ripple started on the smooth surface of a pond will travel outward at a uniform speed of a few feet per second, determined by the viscosity and elasticity of the liquid. A sound wave set up in air will travel at a uniform speed of about 1100 feet per second, this speed depending also on the elasticity and density of the medium. The speed of sound in air is fully determined by the physical constants of air, and does not depend in any way on the loudness or pitch of the sound. Similarly, an electric disturbance, such as a sudden change in the value of a charge producing an electric field will travel in the ether at a uniform speed, determined by the fixed characteristics of the ether, and independent of the magnitude of the change or disturbance. This speed has been measured, and found to be equal to 300,000 kilometers (186,000 miles) per second. Thus at a point 300,000 kilometers away from the charge the increase or decrease of field intensity will occur one second after a similar change has taken place at the point at which the charge is located. This velocity of propagation of the electric wave in the ether is equal to the speed of light in empty space, which seems to indicate that electric waves are of the same nature as light waves, and are propagated in the same medium.

It should be noticed that an increase in the value of a charge results in an increase in the volume of the equipotential spheres. This can be easily demonstrated from Coulomb's equation. Consider a charge m producing at any point of the sphere of radius r and of center m a force f equal to

$$f = \frac{m}{r^2}.$$

If now the charge is increased to some value m', the field intensity will increase throughout space, and the points at which the force is equal to f will be at a distance r' from the charge such that

$$\frac{m'}{r'^2} = f$$

from which

$$\frac{r}{r'} = \sqrt{\frac{m}{m'}}$$

which gives the ratio of the radii of the spheres of the same potential before and after the increase of the charge. This increase in the volume of any one equipotential sphere will take place for all the spheres of the field in succession, each sphere beginning its expansion as soon as the adjacent inner sphere has been expanded. The propagation of this change, as shown above, takes place with the speed of light.

Another point of importance is that an increase or decrease in the value of the charge simply changes the value of the intensity of the field without changing its direction; that is, the lines of force of the field are always straight lines passing through the center of the charge, whatever the value of the latter may be. Thus, the electrostatic field may be considered as somewhat similar to a compression strain of the ether along the lines of force of the field.

The process of wave propagation of this strain, as explained above, is given the name of *radiation* of electrostatic energy. The reason for this name will be better understood after electromagnetic phenomena have been taken up.

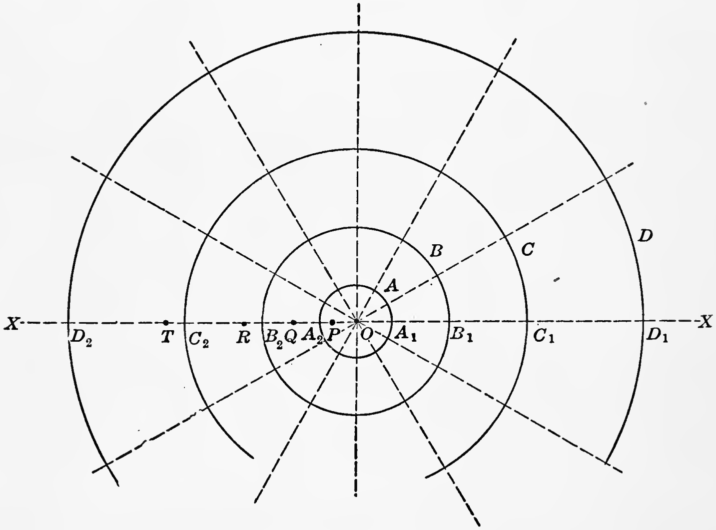


FIG. 5.

Field of a Moving Charge.—Having studied the electric field of a stationary charge fixed at some point of space, it is now possible to take up the transformation of such a field resulting from a motion of the charge.

Consider an electric charge m at rest at a point O of space, Fig. 5. The electrostatic field of this charge is represented by the equipotential spheres A, B, C, D, corresponding to certain values of potential a, b, c, d, while the direction of the field is along straight radial lines, shown in dotted lines on the figure.

Now let the charge m be suddenly set in motion toward the left at a uniform speed s along the straight line XX. For a point on the line XX located in the immediate vicinity of point O and at the right of that point, the removal of the charge m from its original position O and its motion toward the left, will be eqivalent to leaving the charge stationary at O and gradually reducing it from its original value m. This equivalent case was studied in the previous section, and it was shown that the resultant decrease in the field would propagate from point O with the velocity of light V. Thus after a certain fraction of time t_1 the field will start decreasing at point A_1, but until that time, which is the time the wave reaches that point, the field there remains unchanged.

Similarly, for a fixed point located ahead of the charge, the effect of the motion of the charge will be equivalent to leaving the charge m at the point O and increasing its value. This would start a wave of increased field in the direction of the actual motion of the charge, this wave moving with the velocity of light V. Thus, after the time t_1, this wave will have reached the point A_2.

Now, assuming that the speed s of the charge m is smaller than the velocity of light V, the distance traveled by the charge during the time $t_1 = \frac{OA}{V}$ required by the wave to reach the sphere A will be

$$OP = st_1 = s\frac{OA}{V}$$

Thus, at the time t_1 the charge m will be at the point P instead of point O, but all of the field exterior to the sphere A will be in the same state as when the charge was in its original position O.

The moment the wave has reached the equipotential sphere A, the sphere will be set into motion at the same speed as the charge m, so that the relative position of the charge and that sphere will remain the same as long as the charge retains the same speed.

In a similar way, it may be seen that the wave started at O when the charge was set in motion will reach the sphere B after a total period $t_2 = \frac{OB}{V}$ during which the charge has traveled the total distance

$$OQ = st_2 = s\frac{OB}{V}$$

Generally speaking, a point at a distance r from the original position O of the charge will be reached by the wave after a time $t = \frac{r}{V}$ during which the charge has traveled a distance

$$d = st = \frac{sr}{V}.$$

And as soon as the wave reaches that point, the corresponding equipotential sphere will be set in motion at the same speed and in the same direction as the charge. This distance d is the distance between the center of the sphere of radius r and the position of the charge m at the time the wave due to the motion of the charge reaches that sphere.

From this it is easy to represent the electrostatic field of a moving charge by showing the new positions of the equipotential

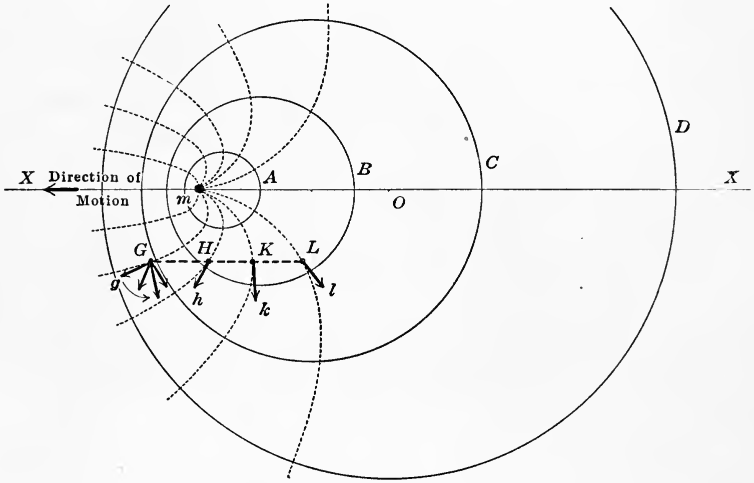


Fig. 6.

spheres with respect to the charge. This has been done in Fig. 6. The direction of the field is obtained by drawing lines of force, which are lines passing through the charge center and normal to all equipotential surfaces at the points of intersection. These lines are shown dotted in Fig. 6.

Electromagnetic Field.—The change in the shape of the electrostatic field of a charge due to its being kept in motion instead of remaining stationary is accompanied by certain phenomena which result from this distortion of the field and from its motion through space, particularly the latter.

Consider a point G, Fig. 6, fixed in space, and located outside the line XX. With the moving charge m in the position shown, the field at point G will have the direction Gg, tangent to the line of force passing at that instant through the point G. As the charge moves along the line XX toward the left, different lines of force of the field will pass through G. Thus, when the charge m has moved along XX over total distances equal to GH, GK, GL, the field at G will successively assume directions parallel to Hh, Kk, Ll. It may then be seen that, as the charge m is made to move along the straight line XX, the field at the fixed point of space G will be along a straight line passing through G, contained in the plane of the paper and gradually rotating around that point. There is thus set up at point G, in addition to the electrostatic force or compression strain due to the presence of the charge m, a torsional or rotational strain, due to the motion of the charge and the resultant rotation of the direction of the field at the point considered. This torsional strain of the ether takes place along a line perpendicular to the direction of motion of the electric charge setting up the field, and perpendicular also to the direction of the electrostatic field at the point considered. This force is called an *electromagnetic force.*

From the above explanation, it may be understood that the magnitude of this electromagnetic strain depends both on the value of the electrostatic field and on the rate at which its direction is turning around the point considered. It therefore depends, for a given point of the field, on the value of the moving charge, on its speed of motion, and on the shape of its trajectory.

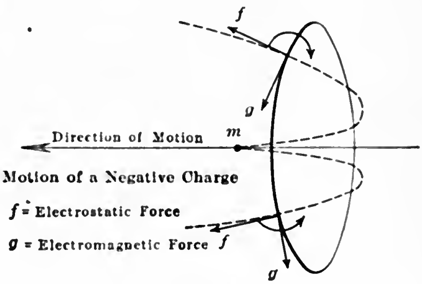


Fig. 7.

Also, the strain will be reversed in direction when the direction of motion of the charge is reversed.

Since the lines of force shown for one plane in Fig. 6, exist identically for all planes containing the line XX, the electrostatic and electromagnetic fields shown for point G will have the same values respectively for all points of the circle described by point G when Fig. 6 is rotated about the line XX. This is represented in Fig. 7, from which it may be seen that the direction of the electromagnetic field at each point of the circle will be tangent to the circle. This circle is therefore a line of force of the electromagnetic field, the direction of the force along this circle being dependent only on the direction of motion of the charge m, and on whether the charge is positive or negative.

Forces Acting during the Motion of a Charge.—As explained in the previous parapraph, when an electric charge m is moved along a straight line path at a constant speed s, the electrostatic field of the charge, which was symmetrical when the charge was at rest, is distorted, and an electromagnetic field which did not exist when the charge was at rest, appears as a result of the motion.

Referring to Fig. 6, it may be seen that as a result of the motion of the charge, the latter assumes an eccentric position with regard to the equipotential spheres A, B, C, etc. The eccentricity, that is, the ratio of the distance of the charge from the center of any one of the equipotential spheres to the radius of the latter is expressed by the equation

$$\frac{d}{r} = \frac{s}{V},$$

where s is the speed of the charge, V the velocity of electric wave propagation, r the radius of the sphere and d the distance between the charge and the center of the sphere considered. Thus, the amount of distortion of the electrostatic field is directly proportional to the speed s.

As a result of this distortion, the potential gradient ahead of the moving charge is steeper than the gradient in back of the charge, and a force is thus exerted by the medium upon the charge, in such a direction as to tend to move the charge toward the center of the equipotential spheres. This direction is opposite to that of the motion of the charge and the magnitude of the force is greater the greater the speed of the charge, being a function of the eccentricity of the charge. The distortion of the field, and therefore the motion of the charge thus requires the continued application to the charge of a constant force in the direction of the motion, equal and opposite to the reactive force of the field. This force, producing the motion, is called the electromotive force, while the reactive force of the field, which opposes the motion is called the counter-electromotive force.

Having now the electromotive and counter-electromotive forces in equilibrium, and the charge moving at a constant speed, if an attempt is made to increase the speed by increasing the electromotive force, the eccentricity of the charge will increase, and therefore also the unbalance between the potential gradients ahead and in back of the charge. Thus, an increase in the electromotive force is immediately followed by an increase of the counter-electromotive force, and the effect is a tendency to prevent any increase in speed. This effect is somewhat analogous to the inertia of a moving body, except that the apparent mass of an electric charge is seen to vary with its velocity, being zero when the charge is at rest.

Similarly, in order to decrease the speed of the charge, the electromotive force needs simply to be reduced. The counter-electromotive force will then be greater than the electromotive force, and will therefore tend to drive the charge backward toward the right, Fig. 6. But such a motion would immediately be followed by a force opposed to the force producing that motion, and this would be directed to the left, thus tending to keep the speed of the charge to its original value.

These forces which tend to oppose any change in the speed of the charge are attributed to the magnetic field of the charge, which, it was shown, was created by the motion of the charge. Applications of this property will be found in later paragraphs.

Summarizing the above, the motion of an electric charge results in a distortion of the electrostatic field, and in the production of an electromagnetic field. The intensity and shape of these fields depend on the magnitude of the charge, its velocity, and the shape of its trajectory.

It should be noted that, when the charge m is traveling at uniform speed along the line XX, the force applied to it (electromotive force) being exactly balanced by the reactive force (counter-electromotive force), the resultant force is zero, so that there is no force acting on the charge, which therefore has no acceleration. For the same reason, where these conditions prevail the electromotive force does not perform any work.

However, when the electromotive or driving force was first applied when the charge was at rest, there existed no reactive force. The driving force, in setting the charge into motion thus accomplished a certain amount of work, expending a corresponding amount of energy, until it was counterbalanced by the reactive force, and ceased to act and perform further work.

This expenditure of energy on the part of the driving force was accompanied or followed by two phenomena. The first is the creation of a reactive force equal and opposite to the driving force, this representing an expenditure of energy on the part of the electrostatic field of the charge, and therefore a corresponding reduction of the total potential energy of the charge. The second was the setting up of an electromagnetic field, that is, the creation of a store of electromagnetic energy, the amount thus stored being exactly equal to the amount of electrostatic energy lost or expended by the charge.

The function of the applied driving force was then, by setting the charge into motion, to transform a certain amount of the electrostatic energy of the charge into an equivalent amount of electromagnetic energy, this transformation requiring a certain amount of work on the part of the driving force. This transformation once established and having exhausted the energy of the source supplying the driving force, the conditions remain as they are, that is, the charge remains in motion at a constant speed.

In order to stop the charge, it is then necessary to apply to it a force equal and opposite to the original driving force. That is, an amount of energy exactly equal to that expended in setting the charge in motion must be expended in the opposite direction in bringing it to rest. And this is accompanied by the disappearance of the electromagnetic energy and its restoration to the electrostatic field of the charge, bringing the latter back to its original value.

Thus, while the setting into motion of the charge requires a certain amount of work and results in the transformation of electrostatic energy into electromagnetic energy, the stopping of the charge requires an equal expenditure of energy, but in the opposite direction, and results in the reverse transformation of electromagnetic energy into electrostatic energy.

Summarizing, the total amount of energy expended by the driving agent for one starting and stopping cycle of the charge, is thus zero. The total amount of energy stored in the two fields (electrostatic and electromagnetic) of the charge remains constant. The transformation of electrostatic into electromagnetic energy, or the reverse transformation, creates a force opposed to that causing the transformation, and therefore requires on the part of that latter force, the performance of a certain amount of positive or negative work and the expenditure of a corresponding amount of energy.

These facts will find their application later in connection with alternating currents and resonance phenomena.

ELECTRICAL PHENOMENA IN MATTER

The various electrical phenomena which have been observed to take place in matter, such as the development of electricity by friction of one body against another, or the phenomena of electromagnetic and electrostatic induction, lead to the assumption that electricity is present in material bodies, and may be brought into evidence by placing these under suitable conditions. When the materials are not under such special conditions, they generally do not display any electrical properties and are said to be in a neutral state. An important consideration is the fact that whenever a quantity of electricity is produced on a material body, a quantity of electricity exactly equal, but of opposite polarity is produced simultaneously on some other body or bodies.

Although the reader is probably familiar with the theory of the atomic structure of matter, it may be well to recall this in a few words.

Atomic Structure of Matter

The basis of modern chemical science is the theory that all matter is made up of a large number of molecules, all similar as to size, weight, etc., for a given substance, which are the smallest fractions obtainable of this particular substance. The molecules of all substances are known to be compounds of a number of so-called "elements" of which there are about 70, consisting of the metals and metalloids. These substances are supposed to be made up of distinct particles, similar to the molecules, called atoms, presumably of a simpler structure than the molecules, and retaining their individual identity within the molecules of compound bodies. Thus for instance, a molecule of water is made up of one atom of oxygen and two atoms of hydrogen.

While the atoms of the various elements may differ greatly in their chemical behavior, they all exhibit certain fundamental properties which are common to all. Such a universal property is the ability of assuming any one of the solid, liquid or gaseous states. Thus, any given element will assume these three physical states successively as its temperature is raised, the only difference between different substances being the actual temperature at which the passage from one state to the other occurs. This has been explained by assuming that, for all solid bodies, the atoms are vibrating or rotating about a position of equilibrium, at a frequency and with an amplitude increasing with the temperature of the body, until, for a temperature high enough these oscillations become so great as to overcome the inter-atomic forces which gave the solid its stable structure, thus allowing the atoms to change their relative positions. The body then reaches the liquid state. Further increase of temperature resulting in greater atomic agitation, brings about collisions between the atoms of increasing strength, which result in an increase of the average distance between the atoms, an expansion of the liquid, and finally, as the collisions become strong enough, in the projection of the atoms out into space, thus forming the gaseous state of the body.

Electron Theory of Matter

While the theory of the atomic structure of matter may give a satisfactory account of chemical phenomena, and a number of physical properties of matter, it cannot be adapted, in the form given above, to the interpretation of the action on a body of forces such as gravitational and electrical forces. As mentioned previously, the electrical properties of matter are explained by assuming the existence of electricity within the various substances. From the above explanation of the atomic structure of matter, it may be seen that there are two possible theories as to where this electricity is stored within the substance of the body; it may be in the intervals between the atoms, or within the atoms themselves. While both of these theories have been worked out, the latter one is now generally adopted, and has been perfected so that it will enable one to satisfactorily explain a considerable number of the properties of matter. Some of these properties which are studied in this book, are the properties of electrical conductors and insulators, the phenomena of electrostatic and electromagnetic induction, ionization of gases, thermionic emission and energy radiation.

The similarity of certain properties of the atoms of the various chemical elements, such as the property of assuming three physical states, solid, liquid or gaseous, and the property of being subject to gravitational force, leads to the conclusion that the structure of these various atoms, although different for the various substances, must be based on a single fundamental plan.

An atom is to be considered as having at its center a mass or charge of positive electricity, around which rotate a number of small masses or charges of negative electricity. The sum total of the negative charges is normally equal to the positive charge, so that the amounts of electricity of opposite polarity exactly neutralize each other, and no exterior electrical force can be observed. All the negative charges are equal to each other, and are called *electrons*. These electrons are supposed to be the smallest possible fraction of negative electricity obtainable. The diameter of one electron is estimated as 1×10^{-13} cm. The hydrogen atom is about 60,000 times larger. The charge of the electron is about 1.59×10^{-19} coulomb. These electrons rotate around the central positive charge along definite orbits and at considerable speed. The entire structure is quite similar to our solar system, with the sun at its center and the planets revolving around it in definite orbits. However, while the various planets are of different sizes and weights, all the electrons are identical. The difference between the atoms of different substances is then in the number of electrons revolving around the central positive charge. This accounts for the difference in the atomic weights and other properties. Thus, while an atom of hydrogen probably has only two electrons, an atom of mercury has several thousand of them.

Another similarity with a solar system is the existence of electrons in the atoms of certain substances such as the metals, which follow very eccentric orbits, parabolas or long ellipses, like the comets of our solar system. These electrons are thus projected at a comparatively great distance from the center of the atom, and may even go so far as to escape the action of one atom to enter the system of another one. Such electrons, which are not bound to one particular atom are called "free electrons." They are projected from one atom to another, traveling in all directions in the space between adjacent atoms. And, while they are at times under the action of the central positive charge of one atom, when they come in its proximity, they are immediately expelled, as if colliding with the atoms. When these free electrons are in the space between the atoms, they are thus only very loosely bound to one definite atom, and may therefore be easily torn from it under a suitable electric field. They will then travel between the atoms, thus creating within the body a current of electricity, which, by definition, is the transfer of a quanity of electricity from one point to another. This will be explained in greater detail in a later paragraph.

It is not considered necessary in a book of this kind to enter into a detailed explanation and computation of the various forces acting on the revolving electrons. Some of these forces are the attracting force of the central positive charge, the repelling forces from the other electrons of the atom, the reactive force on the electron due to it own motion, and a centrifugal force resulting from its apparent mass. Due to the effective inertia of the electrons resulting from the magnetic field set up by their motion, as explained on page 13, the electrons tend to maintain a constant linear velocity, and any attempt to shift the position of the orbit of the electrons with respect to the central positive charge or to alter its radius, is accompanied by a corresponding variation of the forces acting on them. These forces act in such a way that the orbit or ring of electrons will tend to spring back to its original position. The system is thus possessed of a sort of elasticity, which permits it to be distorted, but which also restores it to its former shape when the distorting force is removed.

The above theory will be applied to the explanation of electrical phenomena in matter. These phenomena are very similar to those studied in connection with an electric charge in empty space, the electrical properties of matter being simply those of the electrical charges constituting the atoms, taking into account the various forces on the charges due to the surrounding charges making up the atoms. The behavior of electric charges in matter or material media, is thus simply a modification of the properties of electric charges in ether, and they will be studied in the same order as was followed in the case of an electric charge isolated in empty space.

Dielectric Constant.—In accordance with the electron theory, substances may be divided into two classes; those in which there are no free electrons, called insulators or dielectrics, and those in which there are free electrons, called conductors. In reality, practically all materials contain free electrons, but many contain them in such small numbers per unit volume that in most cases they may be considered as non-existent.

Consider then a stationary electric charge m placed in an insulating medium, such a air, for instance. The only difference between this and the case of Fig. 1 is that, instead of the charge being placed in a space empty of all matter and therefore occupied only by the ether, it is placed in a space containing an insulating material; that is, containing ether and a number of molecules or atoms, or, which amounts to the same thing, a number of sets of small positive and negative charges, each arranged in a structure as described heretofore to form an atom. The effect of the charge m is to establish an electrostatic field in the surrounding ether, as in the case of Fig. 1. Suppose, to give a definite example, that the charge m is negative. Then, the effect of its field will be to repel any negative charge and attract any positive charge which may be present in the field, with a force f given by Coulomb's relation, equation (1).

The effect of the charge m on any atom of the medium will then be to attract the positive central charge of the atom and to repel the negative electrons with a force expressed by the above equation. This will result in a shift of the electron ring or orbit with respect to the central positive charge, which will be of such a magnitude that the internal forces of the atom tending to keep the positive charge and the electron rings centered with respect to each other, will exactly balance the force of the field on it due to the charge m. The entire atom, however, will not be attracted or repelled by the charge, since the forces of the charge on the negative and positive parts of the atom are equal and opposite, these two parts of the atom being equal quantities of positive and negative electricity. The effect of the charge on the insulating medium is thus to produce a displacement or distortion of the atomic structure. This distortion varies in accordance with the distance of the atom from the charge and is proportional to the field intensity. Furthermore, this distortion requiring for its establishment a certain amount of work on the part of the force of the field, results in an absorption of a corresponding amount of energy, which is a portion of the potential energy of the field. The amount of potential energy thus stored in the dielectric material depends then on the value of the charge, on the internal atomic forces limiting the displacement of the atomic structure, and on the number of atoms per unit volume of the dielectric. It thus depends on the nature of the dielectric occupying the space surrounding the charge.

It was shown in the study of the electrostatic field of a charge in ether that the setting up of this field was accompanied by a storage of potential energy in the ether. Since in the case of an insulating medium, a part of that potential energy is used in effecting the atomic distortion of the medium, and is thereby stored in the atoms of the medium, the amount of energy remaining in the ether as potential energy or energy available for further work is, at each point of space, decreased by an amount represented by the ratio $1/k$ which is evidently dependent on the nature of the medium. This decrease of potential energy produces a decrease in the potential at all points of the field, and therefore a decrease in the radius of all equipotential spheres. In order to restore the field to the value it would have in the absence of material medium, it is thus necessary to increase the value of the charge from a value m to a value m_1 such that

$$\frac{m_1}{m} = k$$

The value of k which is characteristic of the dielectric material, is called the *dielectric constant* or *specific inductive capacitance* of the material. Its value varies for different materials, being approximately equal to 1.00059 for air, 2.3 for paraffin, 6.6 for mica, 10 for glass, 80 for water. These values will of course vary with the temperature of the medium which, it was shown, varies the distance between the atoms of the substance, and therefore the number of atoms per unit volume.

From the above explanation of the effect of a material medium on the effective value of an electric charge m, it is seen that Coulomb's law, as modified by the presence of matter in the space occupied by the field of the charge, is expressed by the equation:

$$f = \frac{1}{k}\frac{mm'}{r^2} \qquad (3)$$

This will also express Coulomb's law in a vacuum since the dielectric constant k was taken as unity for space empty of all matter.

Electrostatic Induction.—Having studied the effect of a stationary electric charge on substances containing no free electrons, materials containing both "bound" and "free" electrons will be studied here. It is evident that a material cannot consist solely of free electrons, which are electric charges without material support.

Consider a charge *m*, Fig. 8, which sets up in space an electrostatic field, represented by its equipotential surfaces. To fix ideas, assume the charge *m* to be negative. Let *AB* represent a rod of a material containing free electrons, such as a metal for instance, placed with its axis along one of the lines of force of the field, that is, along a straight line passing through the charge *m*. Under these conditions, the free electrons will be repelled by the negative charge *m*, in accordance with Coulomb's law, and, since they can move between the atoms, as explained on page 17, they will be driven toward the extremity of the rod

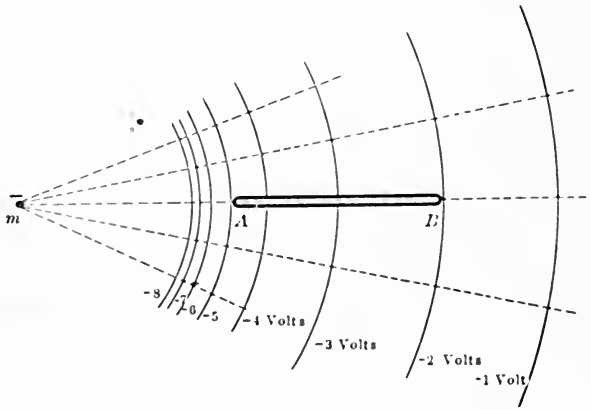


FIG. 8.

marked *B*. This causes in the rod a deficiency in the normal number of free electrons at the end *A* and an excess of electrons at the end *B*. These ends will therefore be charged positively and negatively, respectively, since the sum of the central positive charges of the atoms of the rod at end *A* is now slightly greater than the sum of the negative electrons, while at the other end the reverse is true. The excess of negative electrons at one end is equal to the deficiency at the other end.

This accumulation of a negative charge at the end *B* of the rod will produce in that region an electrostatic field which, adding to that of the charge *m*, will raise the potential of that portion of the field. On the other hand, the positive charge at the end *A* of the rod produces an electrostatic field opposed to that of charge *m* and will therefore lower the field in that region. The result is that the difference of potential which was existing between points *A* and *B* due to the field of charge *m* is neutralized or destroyed by the establishment within the conductor of negative and positive charges which create a field which is superimposed upon that of charge *m*. The free negative electrons in the rod thus move toward the end *B* until a sufficient negative charge has accumulated there to make the potential at that end equal to that at the end *A*. There being then no longer any difference of potential along the rod, the transfer of electrons ceases and a condition of equilibrium is reached. This condition is represented in Fig. 9.

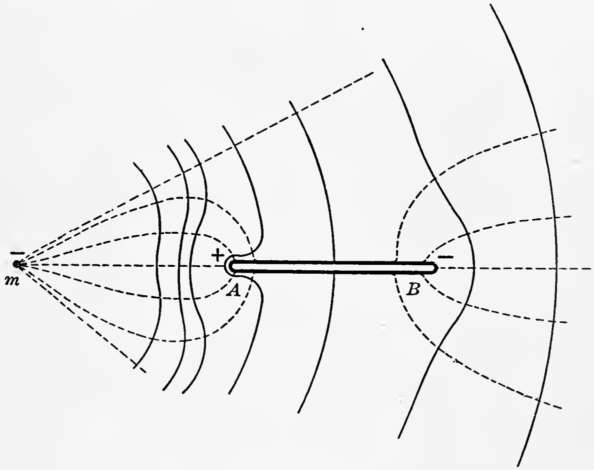


Fig. 9.

It should be noted that during the very short interval of time required for the motion of electrons from one end of the rod to the other, each electron creates during its motion an electromagnetic field, as explained in the case of a moving electric charge. This magnetic field disappears when the condition of equilibrium has been reached. The motion of the electrons along the rod constitutes a current of electricity, which flows whenever a difference of potential is established between two points of the conductor, that is, whenever the conductor cuts different equipotential surfaces. The current flow will stop when these surfaces have been so distorted or shifted that they no longer cut the conductor. If then the value of the charge *m* is continuously varied between two values m_1 and m_2, an electric current will flow in the conductor in alternate directions. This current is said to be induced in the conductor by the variations of the charge *m*.

The motion of the electrons in the conductor requires a certain amount of work on the part of the force of the electrostatic field of the charge m. In other words, some of the potential energy of the field is spent in doing this work, and is thus stored or used in the system, the transformation being accomplished through the temporary motion of the electrons and temporary setting up of a magnetic field around the conductor.

Condenser. Capacitance.—This property of a conductor placed in an electrostatic field whereby the electrons move in such a way as to equalize the field along its surface, is made use of in devices called condensers, for storing electricity in the form of potential energy in a dielectric or insulating medium. Such a condenser consists essentially of two conductors placed at some distance apart and insulated from each other. In most cases, these conductors are given the shape of parallel plates, A and B, Fig. 10. If these plates are connected to the two terminals of a source of continuous potential, such as a direct current generator G, a certain difference of potential v will be established between the two plates. In the case of Fig. 10, plate A will be positive and plate B negative, this being accomplished by the generator, which removes some of the free negative electrons from the metal of plate A and places a like number of electrons on plate B, creating a flow of electricity along the metallic circuit AGB. When the electrostatic field created between the two plates by the presence of the charges on them, which is in a direction opposite to the electromotive force of the generator, is equal to the latter, the transfer of electricity along the circuit ceases and the condenser is then said to be charged.

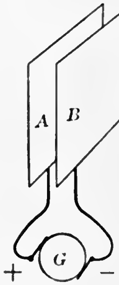


Fig. 10.

It may be seen from Coulomb's law that, for a given condenser, the value of the electrostatic field is directly proportional to the value of the charges on the plates, and therefore also that the charge of a condenser is directly proportional to the difference of potential between the plates, this being the measure of the field. The ratio m/v of the charge of a condenser to the potential difference between its plates has therefore a constant value C for a given condenser. This value C is called the *capacitance* or electrostatic capacity of the condenser. In the practical system of units, the unit of capacitance is the farad, which is the capacitance of a condenser which requires a difference of potential of 1 volt between its plates to store a charge of one coulomb. The subdivisions of this unit which are frequently used in radio work are the microfarad (mfd.) and micro-microfarad (micro-mfd.) which are respectively the millionth part of a farad and of a microfarad.

It has been shown previously that if a charge m produces a field f in vacuum, a charge km will be required to produce the same field in an insulating medium of dielectric constant k. If then $C = m/v$ is the capacitance of a condenser in vacuum, the capacitance of the same condenser in an insulating medium of dielectric constant k will be

$$C' = \frac{km}{v} = kC$$

This gives a method for measuring the dielectric constant k of a medium, consisting in measuring the capacitance of a condenser in vacuum and then of the same condenser having as a dielectric the material under test. The ratio of the latter value to the former gives the value of the constant k.

The capacitance of a parallel plate condenser may be calculated from the formula

$$C = 0.0885 \times \frac{kS}{d \times 10^6}$$

where C is the capacitance in microfarads, S the area in square centimeters of one side of one conducting plate, d the distance between the plates in centimeters and k the dielectric constant of the insulator separating the plates.

The flow or current of electricity along the metallic circuit which was shown to take place while the condenser was being charged, is called the *charging current* of the condenser. This is not the only displacement of electricity taking place in the system at that time during the charging or transient period. The accumulation of opposite electric charges on the two condenser plates by creating an electrostatic field in the medium between and surrounding them, produces in the dielectric substance a similarly gradual distortion of the atoms, as was explained in a previous paragraph. This creates a limited shift or displacement of electric charges in the dielectric medium, called a *displacement current.*

The condenser having been fully charged to the voltage of the generator G, a quantity of electricity Q, equal to the charge of an electron multiplied by the number of electrons transferred, has been transferred from one of its plates to the other. If the condenser is then disconnected from the generator, the unbalance of the electrical condition of the plates will persist, since there is no way of transferring electricity from one plate to the other, they being insulated from each other. The electrostatic field will therefore remain constant, and thus a certain amount of electrical energy will have been stored in the condenser as potential energy. If V is the difference of potential between the plates, it may be seen from Coulomb's law and from the above explanation of the process of charging the condenser, that the amount W of potential or electrostatic energy stored in the condenser is equal to

$$W = \frac{1}{2}QV$$

and since by definition

$$C = \frac{Q}{V}.$$

the above expression becomes

$$W = \frac{1}{2}CV^2 \tag{4}$$

It will be shown later, when studying the discharge of a condenser, how this potential energy may be used to perform work by its transformation into kinetic or electromagnetic energy.

The above expression of the electrostatic energy stored in a condenser may be demonstrated as indicated in the following: From the definition of the capacitance C of a condenser, the charge Q of the condenser corresponding to a difference of potential V across the condenser is equal to

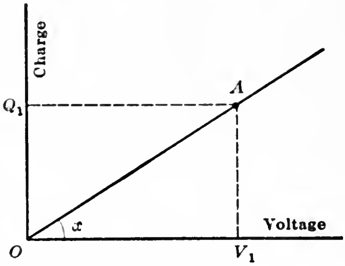


FIG. 11.

$$Q = CV$$

This may be represented graphically as in Fig. 11, the slope of the straight line representing the equation being equal to C. For any given potential V_1, the amount of energy stored in the condenser is then equal to the area of the triangle OV_1A, and is therefore $W = \frac{1}{2}Q_1V_1$ which is the expression given previously.

There are many different forms of condenser construction. Condensers may be divided in fixed and variable types. Those of the first type are generally made up of sets of alternate metal foil and paper, mica films or glass plates, all even and all odd numbered metal foils being commonly connected respectively to form the two plates of the condenser.

Variable condensers are generally made up of a set of fixed, parallel metal plates and a set of movable plates. The latter are separated from the former by air or oil and mounted on a shaft so that they may be rotated and a variable portion of the movable plates inserted between the fixed plates, thus varying the capacitance of the condenser.

Electric Current.—It was shown in the study of electrostatic induction that a flow of electrons, that is, a flow of electric charges, takes place in a metal rod placed in an electrostatic field so as to cut different equipotential spheres. In other words, a flow of electricity takes place in a conductor when a difference of potential or electromotive force is maintained at two points of the conductor. In the case of electrostatic induction studied above, this electric current flow is only temporary, due to the appearance, as a result of the transfer of electricity within the conductor, of electric charges that neutralize the original electromotive force which gave rise to the electric current. If however by some means, the electromotive force is maintained at its original value, such as for instance by continuously removing the neutralizing charges, the electric current will flow continuously instead of only temporarily. This may be done by inserting the conductor AB, Fig. 12, in a complete, closed conducting circuit, at some point of which an electromotive force is created and maintained by some suitable device, such as a generator G. Under these conditions, a potential gradient is established along the entire circuit, and under the effect of the resulting electric force, the free electrons are set in motion at all points of the circuit. There is thus a steady, continuous flow of electrons or electric charges around the circuit, which creates a so-called *electric current*. By definition, the electric current in a

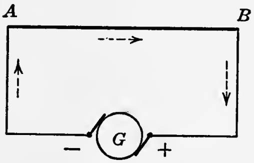


FIG. 12.

circuit is measured by the quantity of electricity passing one point of the circuit during a unit of time. In the practical system of units, the unit of current is called the ampere.

A steady current flowing in a circuit has the same value at all points of the circuit. This being quite obvious, no demonstration of the fact seems necessary here.

The direction of the electric current, before the application of the electron theory, was assumed to take place from the positive to the negative end of a conductor. The direction of the electron flow, the electrons being negative charges, is seen to be from the negative to the positive terminal. This distinction should be remembered when speaking of electric current and electron current.

Inductance.—It was shown in the study of the motion of an electric charge that as a result of motion, a magnetic field is created in the space surrounding the charge, the direction of the force of this field, that is, the lines of force of this field, being along circles concentric with the trajectory of the charge.

An electric current, being the result of the parallel motion of a large number of similar charges, will then give rise to a magnetic field, the lines of force of which are circles surrounding the wire carrying the current. The intensity of this field is the sum of that of the fields of the individual charges or electrons. The intensity of this magnetic field, as measured by its effect on a unit magnetic mass, is evidently *directly proportional to the current* in the wire, which is for the present assumed to be surrounded by a non-magnetic medium. This may easily be demonstrated by the following illustration.

FIG. 13.

Consider a straight conductor *AB*, Fig. 13, carrying a certain current *I*. This current creates in the surrounding space a magnetic field of value Φ. If a second wire *A′B′* carrying a current equal to *I* is placed alongside the wire *AB*, the field surrounding the two wires, being the sum of the two fields, will be equal to 2Φ. If the wires are so close to each other that they may be considered as one single wire, then the effects will be equivalent to having doubled the current through the wire *AB*. And since this relation holds true for any value of current *I*, the relation may be written

$$\frac{\Phi}{I} = L$$

where I is the current in the circuit, Φ the magnetic flux surrounding the circuits, and L a constant, called the *inductance* of the circuit. It may thus be seen that the inductance of the circuit is equal to the flux when a unit current is flowing through the circuit. The unit of inductance is then the inductance of a circuit in which a unit current creates a unit magnetic flux. In the practical system of units, this unit of inductance is called the *henry.*

The inductance therefore depends on the same factors that affect the value of the magnetic field surrounding the wire. These factors, as will be demonstrated below, are the shape of the circuit, and the nature of the medium surrounding the wire.

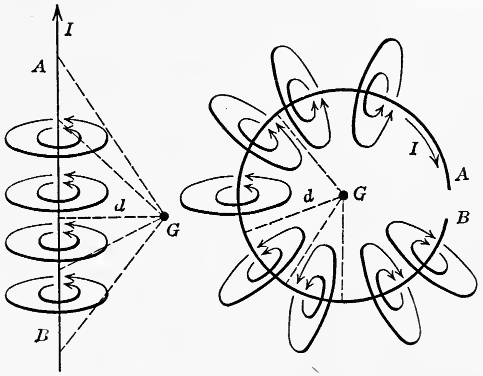


FIG. 14.

That the value of the magnetic field depends on the shape of the circuit may be seen by reference to Fig. 14. Consider the straight wire AB which carries a current I, and a fixed point G at a certain distance d from AB. The magnetic field intensity at that point is the resultant, or vector sum, of the elementary magnetic fields due to the various electrons in motion in the wire. This field thus depends on the distances from point G to the various points of the wire AB. These distances being directly dependent on the shape of the circuit, the magnetic field and therefore the inductance of the circuit depend on its shape. The inductance of the circuit will thus be increased if the distances of all its points to point G are made equal to the minimum distance d. This may be done by bending the conductor into a circle having point G at its center. This is made use of when it is desired to have a circuit of large inductance.

Such a circuit is obtained by shaping the conductor into a coil or solenoid, the added effect of the various turns creating a considerable electromagnetic field within the coil by bringing a great length of the wire within a short space. The inductance of a single layer solenoid coil is given by the formula

$$L = 4\pi nNS$$

where n is the number of turns per centimeter length of coil, N the total number of turns, and S the cross-sectional area of the coil.

The other factor affecting the field, and therefore the inductance, is the nature of the medium surrounding the circuit, and in which the magnetic field is set up. Similarly to the electrostatic field, the magnetic field produces an atomic distortion of the medium, which absorbs some of the magnetic energy stored in the field. This introduces a multiplying factor, called the *permeability* of the medium, which is equivalent to the specific inductive capacitance in the case of the electrostatic field. The inductance of a circuit in a medium of permeability μ is then μL, where L is the inductance of the circuit in a vacuum.[1]

It was explained in the case of a moving electric charge that the electromagnetic field represents a store of electromagnetic energy. This energy was shown to be stored in the medium surrounding the moving charge during the time required by the charge to attain its final speed. In the case of an electric current, the energy is likewise stored in the medium surrounding the circuit during the time required by the current to rise from zero to its final value. Let i be the value of the current in a circuit of inductance L, at a certain instant of the transient period, counted from the time the current was started, and let Φ be the corresponding flux around the circuit. When the current increases by an amount di, the electromagnetic energy stored as a result of this increase is

$$dW = \Phi di$$

And since, from the definition of inductance, $\Phi = Li$, the equation becomes

$$dW = Lidi$$

Integrating over the time t_1 required for the current to reach

[1] More complete information on the calculation of the inductance of a circuit may be found in various publications of the Bureau of Standards, Washington, D. C., as mentioned in its Circular No. 24.

its final value I, the total amount of electromagnetic energy stored is

$$W = \int_{to}^{I} Lidi = \frac{1}{2} LI^2 \qquad (5)$$

Resistance.—From the phenomenon of electric conduction or electric current flow as interpreted in terms of the electron theory, when the conductor AB, Fig. 12, is not under the action of an electromotive force, it is seen that the free electrons while rapidly moving between the atoms do not have any general and common direction. Their motion is at random, and the resultant current is zero. For the present purpose, the free electrons may therefore be considered at rest within the conductor. If now an electromotive force is applied to the conductor, as explained above and shown in Fig. 12, all these free electrons will be set in motion in one direction under the influence of the force of the applied electrostatic field. What has been said in the case of a single electric charge moving in empty space in a straight line path, may be applied to each one of these electrons. Certain conditions, however, prevail in the case of metallic conduction, which somewhat alter the phenomena.

The first of the modifications is due to the fact that, under the electrostatic field resulting from the emf. applied to the conductor AB, there will be, as in the case of insulating materials, a distortion of the atoms of the conductor. This will absorb a part of the potential of the field which is proportional to the specific inductive capacitance of the metal forming the conductor AB.[1] This has been shown to reduce the amount of potential energy of the field available for further work, so that the electrostatic field due to the applied emf. will produce on each free electron of the conductor a force smaller than it would, had the free electron been in empty space. In other words, the same applied emf. will impart to the electrons in the conductor a velocity smaller than it would, were the electrons in a vacuum. This reduction in the velocity of the flow of electrons along the conductor, reduces the amount of electricity flowing past any one point of the latter in the unit of time, and hence the ultimate value of the current.

Another condition to be considered is the fact that the moving electrons enter in frequent collision with the atoms of the metal

[1] See *Physical Review*, August, 1918, p. 130.

which happen to be in their path. This reduces the velocity of the electrons proportionally to the rate of the collisions. It also sets the atoms of the metal into vibration and thereby raises the temperature of the conductor. The latter phenomenon will be taken up later.

As a result of these conditions, the force f on one electron resulting from the electrostatic field due to the applied emf. will set the electron into motion at a velocity s, smaller than that, s', which it would impart to the electron in vacuum. The ratio $\frac{s}{s'}$, of these velocities is a constant, characteristic of the material constituting the conductor and depending on its specific inductive capacitance and therefore on its atomic structure, and on its number of atoms per unit volume.

Another factor affecting the value I of the current flowing in a given conductor under a given applied emf. E is the number of free electrons contained in a unit length of the conductor. For any given conductor it is thus seen that, under constant conditions, the ratio of the applied emf. to the current produced has a constant value, characteristic of the conductor. Thus

$$\frac{E}{I} = R \tag{6}$$

This is the familiar relation known as Ohm's law. The constant R is the *resistance* of the conductor. From the above discussion, it may be seen that this resistance depends on the material making up the conductor and on the dimensions of the conductor. The first one of these factors is determined by the "resistivity" of the material, which is the resistance of a conductor of that material having a unit length and a unit cross-sectional area. If r is the resistivity of the material, the resistance of a conductor of length l and cross-sectional area a is then

$$R = r\,\frac{l}{a}$$

This formula may be immediately derived from the above explanation of electric current conduction. It shows that the resistance of a conductor varies directly with its length and inversely with its cross section. In general, materials containing a large number of free electrons will have a low resistivity.

Forces Producing the Electric Current.—Since the electric current results from the motion of a number of similar charges or electrons, an idea of the phenomenon will be obtained by studying the case of a single electron.

In the case of a single charge moving in vacuum, it was shown that, after a charge has reached a constant speed, the applied electromotive force does nor perform any work in keeping the charge in motion. This corresponds to the case of a steady current in a wire, when the electrostatic and electromagnetic fields and the current have reached their final steady values.

In the case of a charge moving through a conductor, it was shown that the charge collides with the atoms of the material making up the conductor, as indicated by the resistance of the conductor. This creates a retarding force, which tends to stop the motion of the charge. In order to keep the charge in motion an additional force must therefore be applied to the charge, and this force will perform work during the entire duration of the motion of the charge. Furthermore, the energy thus expended and transformed into heat will not be restored to the source of energy when the current is stopped. In this respect it is unlike the energy which was expended at the starting of the current and absorbed by the creation of the magnetic field. There is thus a continuous loss of energy due to the resistance of the conductor.[1] Consider then a conductor for simplicity, of uniform cross section. Let N be the total number of electrons in the conductor, and E the applied emf. required to overcome the resistance of the conductor and keep the electrons in motion, maintaining a current I in the conductor. The force applied to each electron is then

$$e = E/N$$

[1] An experimental proof has been given of the fact that a current, once started in a closed conducting circuit, will continue to flow in the latter without the further application of any emf., provided the circuit has no resistance. This was accomplished by making use of the property of metals to suddenly acquire extremely low resistance at low temperatures (of the order of 1 to 4 deg. absolute). A small coil of lead wire was short circuited and immersed in liquid helium, and a current induced in it by the sudden variation of a magnetic field. The current was then found to flow in the coil for about four days, without the supply of additional energy to the circuit. For full description, see "Communications from the Physical Laboratory of the University of Leyde, by H. Kamerlingh Onnes, Nos. 140-b, 140-c and 141-b."

If s is the speed of one electron, under the force e, then the work done by this force e during one second, that is, the power expended is

$$p = es$$

The total power expended for driving the N electrons at the speed s is then

$$P = E \times Ns = EI$$

And since, from Ohm's law, $E = RI$, the power lost in the circuit is expressed by the relations

$$P = EI = RI^2 = E^2/R$$

The energy lost as heat in the time t is then

$$W = EIt = RI^2t = \frac{E^2t}{R} \qquad (7)$$

Electromagnetic Induction.—From what has been said in connection with the motion of an electric charge, it is seen that whenever the current through a circuit of inductance L is made to vary from its steady value I, there is a change in the amounts of the electrostatic and electromagnetic energy. As a result of the variation of the electromagnetic field it was shown that a force was developed in the system in a direction opposite to that producing the variation of current, and equal to that force. This counter-electromotive force, called the induced counter emf., is therefore proportional to the *rate of change* of the current. Thus, if the current variation is di during the time dt, the variation of electromagnetic flux, and therefore the induced emf. is equal to

$$-L\frac{di}{dt}$$

the minus sign indicating that the force is opposed to that producing the variation.

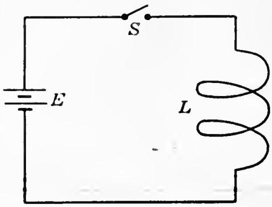


Fig. 15.

Consider now the circuit of Fig. 15, comprising a coil of inductance L and of resistance R, and a battery or generator of constant emf. E. When the switch S is closed, the electromotive force E is impressed across the coil. An electromotive force is thus suddenly applied to the electrons in the wire of the coil, as was the case when studying the motion of a single charge. As already explained, the electromotive force applied to each electron will produce work in accelerating the electrons and bringing them up to their final constant speed. This electromotive force has two components. One of these is the force required to overcome the effect of the resistance of the wire, as explained above in the paragraph on resistance. This component represents a permanent expenditure of energy, and constantly increases with the speed of the electrons, being a maximum when the speed has reached its final steady value. The other component is equivalent to the force applied to a charge moving in vacuum. It effects the distortion of the electrostatic field of the electrons, and the setting up of the magnetic field. As the speed of the electrons increases, the reactive force of their field increases, and finally, when the electrons have reached their final steady speed, the reactive force is equal to and exactly counter-balances this component of the applied electromotive force. This second component of the applied emf. thus becomes zero when the speed of the electrons, and therefore the current, reaches its final steady value. The electromotive force E which was at first applied across the coil as the switch S was closed then reduces to the force E' which is equal to IR after the steady state has been reached. The remainder of the energy of the electrostatic field originally created by the electromotive force E has been transformed into electromagnetic energy by the process explained above.

During the transformation, there was, however, an unbalance between the electromotive and counter-electromotive forces, and therefore also between the driving and the reacting electrostatic fields which produced a resultant electrostatic field. This field acts upon any electric charge or free electron which may be present in the space occupied by it. If then, Fig. 16, a second coil is wound exactly parallel to the first coil considered, insulated from it and so close to it as to almost coincide, this resultant field will act not only upon the free electrons of the first coil, but also on those of the second coil, setting up an emf. in it. And if this coil is closed on some outside circuit, a current will be made to flow in it. This phenomenon is called *electromagnetic induction*, and the current or emf. in the dotted line circuit of Fig. 16 is said to be induced.

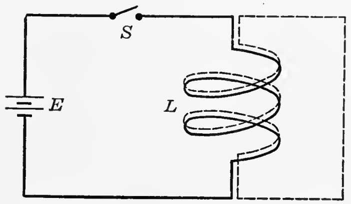


Fig. 16.

As an illustration of this induction phenomenon, consider a current I flowing through a wire AB, Fig. 17, under the effect of the constant emf. of the battery E. Around each point of the wire AB, there is a magnetic field, which may be represented by its lines of force, as shown by the circles. At all points of any one of these circles, the intensity of the magnetic field is the same, each circle corresponding to a certain value of the intensity. If now the emf. impressed across the wire AB is increased, the current through the wire will increase in proportion, and so will the intensity of the magnetic field at each point of the space

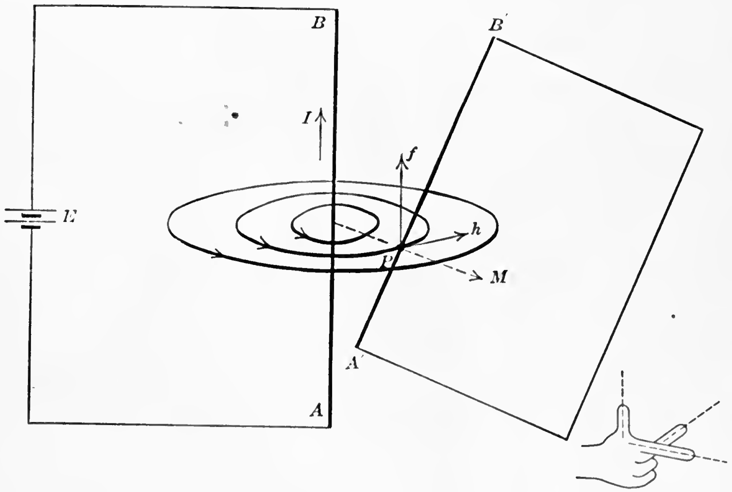


Fig. 17.

surrounding the wire. Thus, after the increase of current will have taken place, all the lines of force of the magnetic field will have increased radii.

Consider now a point P, fixed in space with respect to the wire AB. When the steady current I is flowing in the wire AB, the electromagnetic field at the point P is along the direction h, tangent to the magnetic line of force passing through point P. From the above explanation of electromagnetic induction, and from the remark concerning the change in the radius of the magnetic lines of force, it may be seen that when the current I is increased, the circle which passed through the point P is increased in radius, so that the magnetic force h, without changing direction, is displaced in the radial direction M, perpendicular to h. Also, it was shown that this displacement is accompanied by the setting up of an electrostatic force f, which, being parallel to the direction of the electrostatic field causing the increase in current I, is seen to be perpendicular to M and h. If then an electric charge is placed at P, it will be set in motion along the line f, at the time of the motion of h along M, and in a direction depending on the polarity of the charge. If the current I is decreased instead of being increased, the reverse effects will take place.

This is easily remembered by means of the three-finger rule, whereby the right hand thumb will point in the direction of the force f when the index finger points in the direction of motion of the magnetic lines of force, and the middle finger points in the direction of the magnetic field at the point considered.

If the electric charge which is assumed to be placed at P is a free electron in a wire $A'B'$, it will be set in motion under the effect of the force f. Since all free electrons in the wire $A'B'$ are under similar conditions, there will be a motion of these electrons along the wire. This will accumulate a number of electrons at one end of the wire, producing a negative charge at that end and a positive charge at the other, thus establishing an emf. across the wire. If the latter is connected to a closed circuit, an electric current will flow in it as a result of this emf.

Since the electricity in the wire $A'B'$ must move along the direction $A'B'$, instead of moving directly in the direction of the force f, the actual force producing this motion in the wire depends not only on the magnitude of the force f, but also on the angle $B'Pf$ of that force with the wire $A'B'$. This is equivalent to saying that it depends on the angle α, of the wire with the plane of the magnetic lines of force. It may then be easily demonstrated that the force along $A'B'$ which sets the electrons in motion is proportional to

$$F = f \sin \alpha$$

It is thus seen that the electromotive force induced in the wire $A'B'$ when magnetic lines of force are made to cut the wire is a maximum when the wire is at right angles to the direction of the lines of force and their direction of motion. It is equal to zero when the wire is in the plane of the lines of force and of their direction of motion. In this latter case, the force f would be normal to the wire and therefore could not produce any motion of the electrons along the wire.

Mutual Inductance.—In order to induce an emf. or a current in a circuit *B*, Fig. 18, by varying the current through a circuit *A*, it is not necessary to have the circuits so close to each other as to coincide, as was assumed in Fig. 16. The electromagnetic field of circuit *A* extends out into space. Thus if the varying electromagnetic field of the inducing circuit *A* is made to link with circuit *B*, induction will take place between the two circuits. The induction effect will of course be smaller the greater the distance between the two circuits, since the intensity of the field decreases with increasing distance.

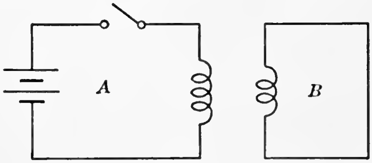


Fig. 18.

If a unit current is made to flow in circuit *A*, then the flux *M* linking with circuit *B* is called the *mutual inductance* of the two circuits. Likewise, if a unit current is made to flow through circuit *B*, an equal flux *M* will link circuit *A*.

From what has been said above, it is seen that the mutual inductance of two circuits depends on their respective inductances, on their relative positions, and on the permeability of the medium.

If at each point, the direction of the wire making up the circuit is perpendicular to the direction of the field at that point, the electrons in the wire will not be set in motion along the wire, and there will be no induced emf. or current in the circuit. In this case it is seen that the mutual inductance *M* is zero.

Electromagnetic Radiation of Energy.—It was explained in the previous paragraphs that electrical energy could be transferred from one circuit to another by varying the electric current in the first circuit from its steady value. It was shown that the induction effect was due, finally, to a motion of the magnetic lines of force of the current in the first circuit, these lines moving away from or toward the inducing circuit, depending on whether the current in it was increasing or decreasing. This increase or decrease in the radius of the lines of force of the magnetic field surrounding a wire carrying a variable current is in some respects similar to the increase or decrease in the radius of the equipotential spheres of an electric charge at rest, but of varying value, which case was studied in an early paragraph.

Thus, the magnetic line of force passing through point P, Fig. 17, will not start expanding or shrinking until all the lines within it have started to expand or shrink. Similarly, no line external to the line of point P will start moving before the line of point P.

The velocity of propagation of a disturbance in the magnetic field is, as in the case of the electrostatic field, equal to the speed of light. Thus, if point P, Fig. 17, is 300,000 kilometers away from the wire AB, the magnetic line of forces at that point will be set into motion one second after the current in the wire has varied from its steady value. The rate of variation of the field, however, is proportional at any point to the rate of variation of the current in the circuit. If then the current in a circuit is alternately increased and decreased, or is alternately reversed, such as is the case with an alternating current, there will be a succession of waves of increasing and decreasing electromagnetic energy traveling from the circuit outward into space. Whenever these waves pass a point of space occupied by an electric circuit, they will induce an emf., and if the circuit is closed, an electric current as was explained previously, thus performing work and effecting the transfer of energy from the wave emitting circuit to the second circuit.

The induction effect may be quite considerable even at great distances, since it does not depend on the absolute value of the field, but simply on the rate of change, and on the magnitude of the change.

Underlying Principles of Radio Communication

The conclusions of the discussion of this chapter may be summarized as follows:

1. It is possible to set up an electric current in a circuit by establishing around this circuit a varying electrostatic or electromagnetic field, or both, having suitable direction.

2. These fields may be produced by varying the current in another circuit, entirely separate from the first.

3. It is thus possible to transfer electrical energy from one circuit to another one entirely separate from the first.

4. The transfer of energy from one circuit to another by this

method depends on the magnitude and the rate of the field variation, and therefore on the magnitude and rate of the current or charge variations producing them. The transfer ceases when the variation stops.

These conclusions are applied to radio communication, where the transfer of electrical energy is effected between the transmitting and the receiving stations. On account of the considerable distance which generally separates the two stations, the electrostatic and electromagnetic induction effects must be very great and of a regular well-defined nature in order to distinguish between the signals and auxiliary and disturbing induction effects heard at the receiving station, which may result from natural or other causes. For this reason, the following conditions are established in all radio communication systems.

1. Alternating currents are set up in the transmitting circuit, of very high frequency and of great intensity or voltage. This insures a high rate of variation of the interlinked electrostatic and electromagnetic fields and a great magnitude of these variations.

2. The transmitting circuit is given a shape suitable for producing fields extending to great distances and generally in the direction of the receiving circuits more than in other directions.

3. The receiving circuit is given such a shape and position as to link it with as large a proportion of the field of the transmitting circuit as possible.

In the following chapters, a detailed study will be made of these several points.

CHAPTER II

PROPERTIES OF OSCILLATORY CIRCUITS

Oscillatory Discharge of a Condenser

Physical Explanation.—Consider a condenser C, Fig. 19, which may be connected by means of a switch to a battery E or to a circuit of inductance L and resistance R. If the condenser is first connected to the battery E by connecting A to A' and B to B', the condenser will in a short time become charged. That is, an electric current will flow in the metallic circuit in the direction $EB'BCAA'E$, accumulating a negative charge on the upper plate of the condenser and a positive charge on the lower plate, as was explained in the previous chapter. This establishes an

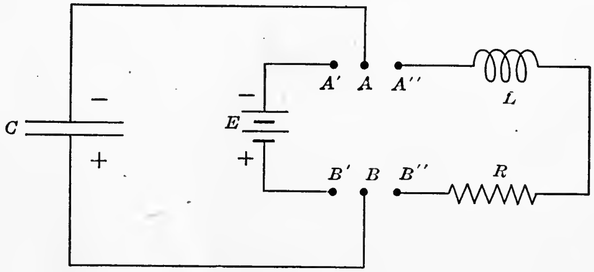


Fig. 19.

electrostatic field about the condenser, and the charging current will flow in the circuit until this field exactly counterbalances the emf. of the battery. The condenser will then be fully charged, and the amount of energy stored in it as potential energy is

$$\tfrac{1}{2}CV^2$$

where C is the capacitance of the condenser and V the potential difference between its plates.

Suppose now that the condenser, after having thus been charged, is connected to the circuit LR, by disconnecting A from A' and B from B' and connecting A to A'' and B to B''. Under the effect of the electrostatic field of the condenser, or, which is equivalent, under the effect of the potential difference V between the condenser plates, an electric current will start flowing in the circuit CLR in the direction of $CBB''RLA''AC$, opposite to the direction of the charging current. This current is the result of the motion of the electrons in the circuit under the effect of the electrostatic field of the condenser. From what has been explained in the previous chapter, it is seen that these electrons will not reach their final speed instantly, due to the reactive force resulting from their motion. This reactive force manifests itself in the form of an electromagnetic field around the circuit, and this was shown to be proportional to the inductance L of the circuit.

The speed of the electrons and therefore the current in the circuit, will thus continuously increase at a rate depending on the inductance of the circuit and the instantaneous impressed emf. And since the condenser is discharging, the energy of the system which was at first entirely stored in the condenser as electrostatic energy, is being gradually transformed into electromagnetic energy. This, however, is not the only transformation taking place, a certain amount of energy being required to overcome the resistance R of the circuit.

Thus the electrostatic energy which was originally stored in the condenser disappears, and becomes transformed partly into electromagnetic energy and partly into heat, the transformation being complete when the condenser is discharged.

When this condition has been reached, the difference of potential across the condenser is zero, and there being no longer any emf. in the circuit, the current flowing through it will stop. This, however, does not occur instantaneously, on account of the electromagnetic field of the circuit, which, it was shown, acts in such a manner as to tend to prevent any change in the velocity of the electrons moving in the circuit. This may be better understood by considering that any change of current in the coil L, by changing the magnetic field in that coil, will induce an emf. in it which is in a direction tending to maintain the current flow.

The electric current will thus continue to flow in the circuit for some time after the condenser is completely discharged, and will therefore charge the condenser again, but this time in a direction opposite to the original one. This current being due to an induced emf. in the coil L, is the result of an expenditure of the electromagnetic energy of the circuit. Thus the electromagnetic energy will gradually transform itself into electrostatic energy which will be stored in the condenser as the latter becomes charged, and into heat as a result of the expenditure of energy required to overcome the resistance R of the circuit. When the entire electromagnetic energy has thus been transformed, the circuit will again have a store of potential energy in the condenser, but the condenser will be charged with a polarity opposite to the original one.

The conditions are therefore similar to the original conditions, except for the polarity of the condenser plates. Also, since the partial transformation of electrostatic or electromagnetic energy into heat was shown to be a permanent one, the amount of electrostatic or potential energy of the system is less than the original store of potential energy.

The condenser being still connected to the circuit LR, it will again start discharging, but in the opposite direction, and the phenomenon will repeat itself with an electric current flowing in the circuit in alternately opposite directions. The discharge of the condenser is therefore said to be oscillatory. It is accomplished by an alternate transformation of electrostatic energy into electromagnetic and of electromagnetic energy into electrostatic, each cycle being accompanied by a loss in the form of heat of a part of the energy of the system. The oscillation will stop when all the energy originally accumulated in the condenser has been thus dissipated as heat in the resistance.

It is then evident that the number of cycles of the oscillation is directly dependent on the value of the resistance of the circuit, being greater the smaller the resistance.

It is possible that the resistance may be so high as to prevent the discharge from being oscillatory. This may be explained as follows: When the condenser first begins to discharge, the difference of potential across its plates is a maximum and the current in the circuit is zero. The discharge of the condenser brings about a gradual increase of the current and decrease of the emf., the IR drop being of course proportional to the resistance of the circuit. If then the resistance is high, the emf. may be reduced so much during the discharge that it cannot bring about further increase of the current through the resistance. And since the emf. across the condenser continues to decrease as long as the current is flowing, the current and emf. will then both decrease together and approach zero as the energy of the original charge is dissipated in heat. There will thus be no store of electromagnetic energy finally set up in the circuit and the condenser will not be recharged. The oscillation then ceases after only a portion of one cycle of discharge and recharge, and the discharge is said to be *aperiodic.*

From the above explanation of the oscillatory discharge of a condenser, it may be shown that the frequency of the alternating current or oscillation is determined by the values of the capacitance and inductance of the circuit. Neglecting the resistance of the circuit, the first phase of the discharge is the transformation of the electrostatic energy of the condenser into electromagnetic energy in the coil, the transformation being accompanied by the establishment of an electric current in the circuit. The time required for the transformation is that required for the current to grow from zero to its maximum value I. The amount of energy originally stored in the condenser of capacitance C when charged to a voltage V was $\frac{1}{2}CV^2$. The amount of energy stored in the coil of inductance L when the condenser is entirely discharged, that is when the current I is flowing, is $\frac{1}{2}LI^2$. Neglecting the resistance, we have then

$$\frac{1}{2}CV^2 = \frac{1}{2}LI^2, \text{ or } V\sqrt{C} = I\sqrt{L}$$

Thus, the smaller the inductance L, the greater the discharge current I. And since, by definition, the current is equal to the amount of electricity passing one point of the circuit per unit of time, we have

$$V\sqrt{C} = I\sqrt{L} = \sqrt{L}\left[\frac{dQ}{dt}\right]_{max}$$

where Q is the original charge of the condenser. Thus the smaller the inductance, the more rapidly will the charge or quantity of electricity accumulated on the condenser flow through the circuit. From the above expression, it is then seen that, for a given condenser, the time required to once discharge the condenser and therefore the time to go through one cycle of the oscillation, and finally the *period* of the alternating current oscillation vary directly as the square root of the inductance. The frequency therefore varies inversely as the square root of the inductance.

It may be shown in a similar way that, for a given inductance L, the period will vary as the square root of the capacitance, and the frequency will therefore vary inversely.

The period and frequency of the free oscillatory discharge of a condenser through an inductance are called the *natural period* and *natural frequency* of the circuit. The value of these expressions in terms of the circuit constants will be given in a later paragraph.

The oscillation current was shown to be of decreasing maximum amplitude at each reversal, due to the continual loss of energy in the resistance of the circuit. The oscillation, in other words, is *damped.* The measure of the damping, or the *decrement* of this oscillation, which depends primarily on the resistance, is defined as the ratio of two successive maxima of the current in the circuit with the current flowing in the same direction. A quantity frequently encountered in radio circuit calculations is the logarithmic decrement, which is the napirian logarithm of the arithmetic decrement just defined. The expression of the decrement in terms of the circuit constants is given in a later paragraph.

Summarizing the above discussion, it is seen that an alternating current may be obtained in a circuit containing inductance and capacitance by first charging the condenser and then letting it discharge through the inductance. The amplitude of this alternating current will constantly decrease due to the resistance of the circuit. The frequency of the alternating current is determined by the inductance and capacitance of the circuit, and may be made as small or as great as desired. In practice it has been possible to obtain by this method oscillations of frequencies as high as several millions of cycles per second.

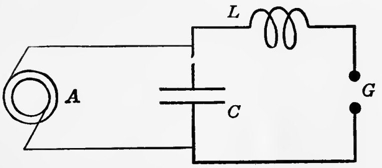


Fig. 20.

On account of the damping, however, the actual duration of an os cillation is in general not more than a fraction of a second. If it is desired to have a succession of oscillations in the circuit, it is then necessary to recharge the condenser after each oscillation has died out. This is done by inserting a spark gap in the oscillatory circuit and charging the condenser by means of a source of alternating or pulsating current, as will be explained in the following paragraphs.

The method is illustrated by the diagram of Fig. 20. The oscillatory circuit comprises the condenser C, inductance coil L, and gap G. The two plates of the condenser are connected to the two terminals of an alternator A generating a sine wave alternating emf. If the gap G is large enough so that it will

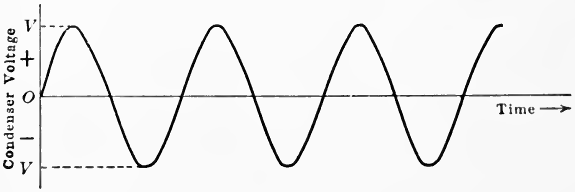


FIG. 21.

withstand the maximum voltage of the alternator without breaking down, the potential difference between the condenser plates will at every instant be the same as the alternator emf. and may be represented by a sine curve as in Fig. 21, from which it is seen that the potential difference varies between the maximum values $+V$ and $-V$ volts.

If the gap G, which is connected in parallel with the condenser across the alternator terminals, is set so that a spark will jump across it when a voltage V', smaller than V, is impressed upon it, the phenomena are of a quite different nature.

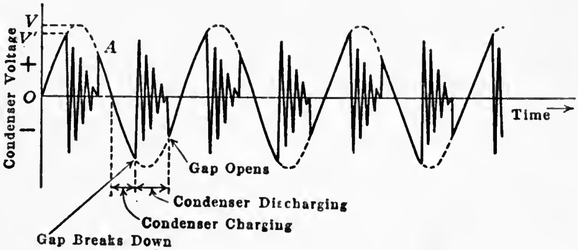


FIG. 22.

Thus in Fig. 22, when the alternator voltage rises from zero to the value V', the condenser voltage rises similarly, and an amount of potential energy equal to $\frac{1}{2}CV'^2$ is stored in the condenser. As the voltage reaches the value V' for which the spark gap G is set, the latter suddenly breaks down; that is, a spark bridges the gap, ionizing the gas between the electrodes and filling the space between them with metal vapors. This suddenly closes the oscillatory circuit *LGC* and the condenser *C* discharges through the inductance *L*. If the resistance of the wires making up that circuit, and that of the spark across the gap *G* is not too great, the discharge of the condenser will be oscillatory, as was explained previously. This creates in the *LGC* circuit a damped alternating current of a frequency determined by the values of *L* and *C*, and produces across the condenser *C* an alternating emf. of the same frequency, as shown in Fig. 22. When the entire amount of potential energy originally stored in the condenser by the alternator has been dissipated as heat in the oscillatory circuit, the oscillation ceases, as was shown previously, and the current stops flowing across the gap *G*, which is again open. The potential across the condenser then again follows the alternator voltage, until the latter reaches the value $-V'$ in the next half cycle, when the phenomenon repeats itself.

It is thus possible to obtain a series of successive groups of oscillations in the oscillatory circuit *LC*. The frequency of the oscillations in each group is entirely determined by the constants of the oscillatory circuit. The frequency of the groups of oscillations is, however, in the above explanation, equal to twice the alternator frequency, there being one oscillation group for each half cycle. It may be seen, however, that if the voltage *A*, Fig. 22, impressed on the condenser after the discharge, is sufficient to break down the spark gap, a new discharge will take place before the end of the first half cycle of the alternator. The frequency of the groups of oscillations is thus dependent on the alternator frequency and on the spark gap adjustment. In general, this adjustment is so made as to give but one discharge for every half cycle of the alternator.

The periodical succession of oscillatory discharges is thus seen to require the periodical charging of the condenser to a certain potential. The alternator was chosen here in the explanation of the phenomena, on account of the great regularity of the cycle of its generated alternating emf. In a great many instances, especially for low power circuits, an induction coil (spark coil) is successfully used instead of the alternator. As is known by the reader, the main difference is then that the induction coil produces a pulsating secondary emf. instead of an alternating emf. This does not affect materially the explanation given above. The regularity of the discharges may not be as good as with the alternator, since it depends on the vibration rate of the induction coil armature which varies somewhat.

It should be noted that the decrement of an oscillatory circuit containing a spark gap is dependent not only on the resistance of the wires making up the circuit, but also on the resistance of the spark which bridges the gap. This latter resistance is, in turn, a function of the current passing through the gap, and increases as the current decreases. It may thus be seen that the total resistance of the oscillatory circuit and therefore the decrement of the oscillation, increases as the oscillation amplitude decreases. The oscillation is thus damped more rapidly in a circuit having a spark gap than in the same circuit without the gap but having an approximately equivalent constant resistance.

Mathematical Explanation of the Oscillatory Discharge of a Condenser.—A complete mathematical explanation of the oscil-

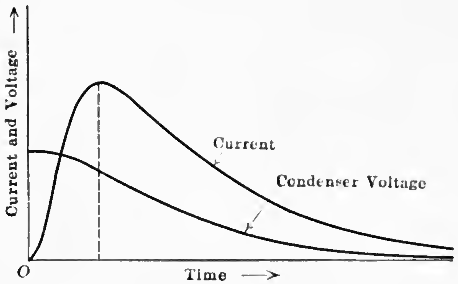


FIG. 23.

latory discharge of a condenser will not be given here. A very good solution of the problem may be found in Fleming's "Principles of Electric Wave Telegraphy and Telephony." The basic idea of the mathematical treatment is as follows:

It was shown that the amount of electrostatic energy variation of the oscillating system of Fig. 19 during a certain interval of time is equal to the sum of the energy lost as heat and the electromagnetic energy variation during the same time. The algebraic sum of these three quantities is therefore equal to zero.

In the time interval dt, the variation of potential energy is

$$d\left(\frac{1}{2}Cv^2\right) = Cvdv$$

The variation of electromagnetic energy is

$$d\left(\frac{1}{2}Li^2\right) = Lidi,$$

and the amount of heat produced is

$$Ri^2dt.$$

The equation may then be written,

$$Cvdv + Lidi + Ri^2dt = 0$$

where C, L and R are the constants of the oscillatory circuit, i the current in the circuit at the instant considered, v the voltage at the condenser, and di and dv, the current and voltage variations during the period of time dt. Since the current i is by definition equal to the quantity of electricity dq set in motion in the time dt, we have

$$i = \frac{dq}{dt}$$

and since

$$q = Cv$$

we have

$$dq = Cdv$$

and

$$i = \frac{dq}{dt} = C\frac{dv}{dt}$$

so that, finally, the equation may be wiitten

$$\frac{d^2v}{dt^2} + \frac{Rdv}{Ldt} + \frac{v}{LC} = 0$$

or

$$v\left(a^2 + \frac{R}{L}a + \frac{1}{LC}\right) = 0$$

An analysis of this equation will show that there are two cases to be considered, according to whether the values of a which will make the quantity between brackets equal to zero are real or imaginary.

If the values of a are real, that is, if

$$\frac{R^2}{4L^2} - \frac{1}{LC} > 0 \text{ or } R^2 > 4\frac{L}{C},$$

the equations of v and i are not periodic functions of time, and the discharge is then said to be *aperiodic*. The current and voltage in the circuit may then be represented by curves similar to those of Fig. 23.

If the values of a are imaginary, that is if

$$R^2 < 4\frac{L}{C},$$

the equations of v and i are periodic and exponential functions of time, and are therefore damped oscillations of frequency and decrement something like that shown in Fig. 24 and given by the expressions

$$\text{frequency} = \frac{1}{2\pi}\sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}} \tag{8}$$

$$\text{decrement} = \pi\sqrt{\frac{R^2C}{L}} \tag{9}$$

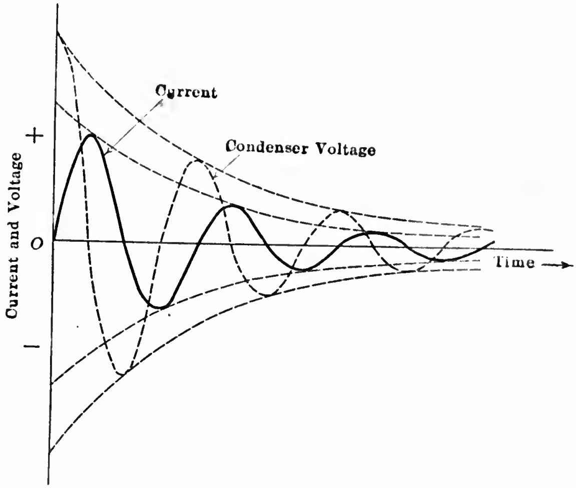


Fig. 24.

In most radio circuits, the resistance R is small and the ratio $\frac{R^2}{4L^2}$ has a negligible value, so that the frequency of the oscillation is generally taken as

$$\text{frequency} = \frac{1}{2\pi\sqrt{LC}} \tag{10}$$

Resonance Phenomena

While induction phenomena are the means of transfer of energy from one circuit to another and therefore the basis of radio communication, resonance phenomena are used to increase the induction effects at great distances, by making the circuits more responsive, that is, of less impedance or inertia, to alternating currents of one definite frequency. These phenomena are studied in the following sections.

Series Resonance.—Consider the circuit of Fig. 25, comprising a condenser C, a resistance R, and an inductance coil L in series with a source of alternating current such as an alternator, which impresses an alternating emf. upon the circuit. An alternating current will then flow in the circuit.

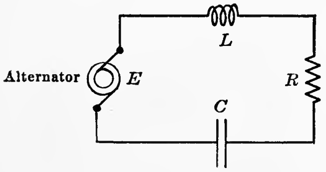


FIG. 25.

In order to obtain the relation between the impressed emf. and the current flowing through the circuit, an equation may be written expressing that the algebraic sum of the emfs. across the inductance coil, the condenser and the resistance, is equal to the impressed emf.

If i is the current at some instant t, and e the impressed emf. at the same instant, then the emf. across the resistance will be equal to Ri. The emf. across the coil L depends on the rate of change of the current at the time t. If dt is an infinitesimal period of time after the instant t, the emf. across the coil will be $L\frac{di}{dt}$. The emf. across the condenser is equal to the ratio of the charge to the capacitance, and is therefore $\frac{Q}{C}$. And since the current in the circuit is the rate of change of the charge, this emf. is seen to be equal to $\frac{\int idt}{C}$.

The equation of the emf. as a function of the current is then:

$$e = Ri + L\frac{di}{dt} + \frac{\int idt}{C}$$

Since the electromotive force e is alternating, it may be expressed as a function of the time by the relation:

$$e = E_{max} \sin(2\pi ft)$$

where f is the frequency of the alternator.

Neglecting the transient terms which are of the form of a rapidly damped oscillation, the relation giving the current in the circuit as a function of the impressed emf. and the circuit constants is then:

$$I = \frac{E}{\sqrt{R^2 + \left(2\pi fL - \frac{1}{2\pi fC}\right)^2}} = \frac{E}{Z}, \qquad (11)$$

the equation relating to effective or maximum values of current and emf. This relation is similar to Ohm's law for direct current, the term Z being substituted for the resistance. This term Z is called the *impedance* of the circuit.

From the above relation it may be seen that the inductive reactance $2\pi fL$ and the capacitive reactance $\frac{1}{2\pi fC}$ vary inversely when the frequency is varied. Thus, if the alternator frequency f is gradually increased, the inductive reactance increases, while the capacitive reactance decreases, the total reactance being at all times equal to the difference of the two. This is represented in Fig. 26 where it is seen that for a certain value of the frequency,

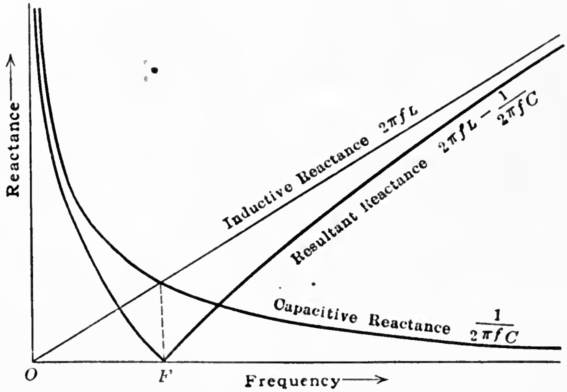


FIG. 26.

the inductive and capacitive reactances are equal, and the resultant reactance is zero. For that frequency:

$$2\pi fL = \frac{1}{2\pi fC}$$

from which

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

This is seen to be equal to the *natural frequency* of the circuit as defined in the previous section.

If then the impressed emf. has a frequency equal to the natural frequency of the circuit, the impedance Z reduces to

$$Z = \sqrt{R^2 + O} = R$$

and the current in the circuit will then be a maximum. When this condition prevails, the circuit is said to be in *resonance* or in *tune* with the impressed emf.

If the frequency of the impressed electromotive force cannot be varied, the oscillatory circuit may be tuned by altering its capacitance or inductance or both in order to make its natural frequency coincide with that of the alternator. The effect of tuning may be shown in a *resonance curve*, as in Fig. 27, where the current flowing in the circuit is plotted against the frequency of the alternator, the maximum of the alternating emf. impressed on the circuit being kept at constant value. At resonance frequency, the current is a maximum, and equal to $I = \frac{E}{R}$. This maximum is therefore greater the lower the resistance of the circuit. Figure 27 shows resonance curves for three circuits, differing only in the value of resistance, curve A corresponding to the circuit of lowest resistance. It is easily seen by plotting these curves in per cent of maximum current that the circuit of lowest resistance will have the sharpest tuning. This means that it will present much less impedance to the flow of currents of its natural frequency than to currents of other frequencies. That is, since the resonance curve is flatter or the peak less pronounced the greater the resistance, the per cent difference between the current flow at resonance frequency F and that at a frequency $F + f$ or $F - f$ slightly different, will be smaller the greater the resistance of the circuit.

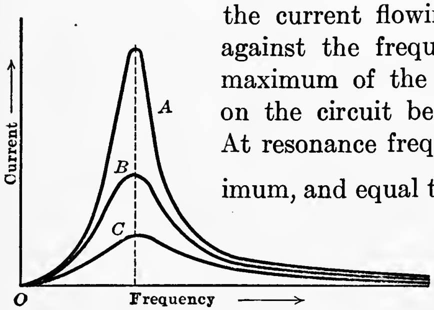


Fig. 27.

Parallel Resonance.—The case of Fig. 28 is different from that of series resonance in that the inductance coil and condenser are connected in parallel across the source of alternating current, the alternating emf. e of frequency f being impressed simultaneously across the coil and the condenser.

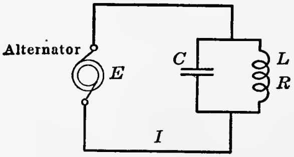


Fig. 28.

Under the alternating emf. e impressed on the oscillatory circuit by the alternator and equal to

$$e = E_{max} \sin (2\pi ft)$$

a current i will be supplied by the alternator, which will divide between the condenser and inductance coil branches of the circuit in inverse proportion to their respective impedances. Thus, neglecting the resistance in series with the condenser, and referring to effective values, the total current in the alternator circuit is

$$I = E\left(\frac{1}{\sqrt{R^2 + (2\pi fL)^2}} - 2\pi fC\right) \qquad (12)$$

This equation may be derived from the general equation (11), as applied to each branch of the oscillatory circuit.

Temporarily neglecting the resistance R of the coil, the equation reduces to

$$I = E\left(\frac{1}{2\pi fL} - 2\pi fC\right) \qquad (13)$$

If then the frequency f of the alternator is gradually increased, the factor $\frac{1}{2\pi fL}$ will continually decrease, while the factor $2\pi fC$ will continually increase. Their difference will therefore become zero for a certain value of the frequency given by the relation

$$\frac{1}{2\pi fL} = 2\pi fC$$

and thus equal to

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

which is the natural frequency of the circuit. From equation (13) it is then seen that the current furnished by the alternator is zero at that frequency. This is shown by the curve of Fig. 29.

If now the resistance of the coil is not neglected, the current supplied by the alternator at resonance frequency will be equal to

$$I = \frac{E}{R}.$$

It should be noted that, although the alternator does not supply any current at resonance, or only a very small current, it still applies an alternating emf. e across each branch L and C of the oscillatory circuit. In each of these branches there is therefore a current flow, but the current flows in the two branches in opposite directions, so that the sum is zero in the external circuit.

The physical explanation is that when for the first time the alternator emf. applied to the circuit rises from zero to its maximum value, it charges the condenser to that maximum value. The condenser then discharges through the inductance as explained previously, at the natural frequency of the circuit. But if this is also the frequency of the alternator, then the emf. impressed on the condenser by the alternator is always in phase with the alternating emf. at the condenser resulting from the oscillation. If, however, the oscillatory discharge took place in the absence of the alternator, the condenser emf. would constantly decrease, due to the resistance losses in the circuit. With the alternator always maintaining the alternating emf. at the condenser with a constant maximum value, it is seen that the function of the alternator is synchronously *to supply the resistance losses in the oscillatory circuit*, so that the current supplied by the alternator is, as found, equal to $I = \frac{E}{R}$.

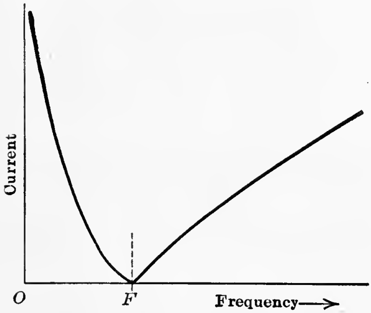


Fig. 29.

In the case studied here, the alternator thus sustains an *undamped* oscillation in the oscillatory circuit. The current then supplied by the alternator depends only on the resistance of the circuit. The current in the circuit itself, however, depends on the value of inductance and capacitance and on the emf. impressed by the alternator across the condenser. In this case, like in the case of series resonance, the current flowing in the oscillatory circuit itself is a maximum when the circuit is tuned to resonance with the frequency of the source of alternating current.

Coupling

In order to set up an alternating current in a closed circuit, it is not necessary to connect the circuit directly to an alternator or other source of alternating current. Such a current may be set up by placing the circuit in an alternating or pulsating magnetic or electrostatic field, so that it will intersect some of the lines of force of the field. Two circuits acting upon each other under these conditions are said to be *coupled.* In the following discussion, the two coupled circuits will be called the primary and secondary circuits respectively, the primary being the circuit producing the fields, and the secondary the circuit in which the current or emf. is induced.

Various Kinds of Coupling.—Two circuits may be coupled to each other in a number of different ways. From the above definition of coupling, it is simply necessary that the two circuits have a common magnetic or electrostatic field, or both. In the first case, the coupling is said to be inductive or electromagnetic, while in the second case it is capacitive or electrostatic.

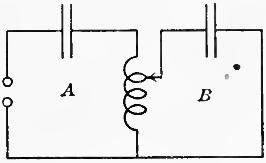


Fig. 30.

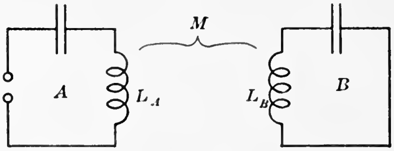


Fig. 31.

Examples of electromagnetic coupling are given in Figs. 30 and 31. In the first case, the primary circuit A and the secondary B have an inductance coil in common. Circuit A, which is the generating circuit, may be any kind of alternating current generating circuit, but is here shown as an oscillatory circuit having a spark gap and excited as previously described. The amount of coupling may be changed by altering the number of turns of the common inductance coil included as a part of the secondary circuit B. The coil acts in the same way as the ordinary auto-transformer.

In the second case (Fig. 31), the two circuits are entirely separate but they have a certain amount of mutual inductance M, due to the coupling of a part of their total inductance. In this case, the degree of coupling may be altered by changing the relative position of the two circuits. The system acts like a transformer. but due to its being generally used for high frequency oscillations, and with its coils often at quite large distances apart, the device is called an "oscillation transformer." In computations relating to circuits coupled in this manner, it is often useful to consider them as coupled directly, as in Fig. 30, the common coil having an inductance equal to the actual mutual inductance M of the circuits, and each circuit having a total inductance equal to its actual total inductance.

An example of capacitive or electrostatic coupling is given in Fig. 32, where the condenser of circuit A forms a part of the capacitance of circuit B. A case of electrostatic and electromagnetic coupling is given in Fig. 33. The number of possible arrangements is great and only the general classes of coupling have therefore been indicated here.

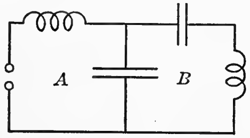


FIG. 32.

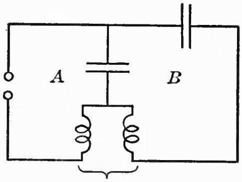


FIG. 33.

Coefficient of Coupling.—While the mutual inductance, capacitance or reactance may be an indication of the degree of coupling, the factor generally used in the computation of ratio circuits is the *coefficient of coupling*, which in the case of two coupled tuned circuits is defined by the expression:

$$k = \frac{X_M}{\sqrt{X_A X_B}} \tag{14}$$

where X_M is the mutual reactance of the two circuits, and X_A and X_B are the inductive or capacitive reactances of the primary and secondary circuits respectively.

Applying this definition to two inductively coupled tuned circuits of mutual inductance M and of respective inductances L_A and L_B, the coefficient of coupling is:

$$k = \frac{M}{\sqrt{L_A L_B}}$$

It should be noted that the greater the value of M, the greater the value of k for two given circuits, the maximum value of k being 1, when $M = \sqrt{L_A L_B}$. This is the case of the circuits shown in Fig. 30, where $M = L_A = L_B$. The coupling is said to be *close* or *loose* according to whether k is large, that is near unity, or small.

For two circuits capacitively coupled,

$$k = \sqrt{\frac{C_A C_B}{(C_A + C_M)(C_B + C_M)}}$$

where C_A and C_B are the total capacitances in each circuit and C_M is the mutual or common capacitance.

RESONANCE PHENOMENA IN COUPLED OSCILLATORY CIRCUITS

Circuits Loosely Coupled.—In taking up the case of two loosely coupled circuits, it is assumed that the primary circuit can induce a current in the secondary, but that this secondary current cannot react upon the primary and induce an emf. or a current in it. Consider a circuit *A*, Fig. 34, containing a fixed inductance and capacitance and a spark gap, and in which oscillations may be set up as explained in the first part of this chapter. Consider also a second oscillatory circuit *B*, loosely coupled to circuit *A* and having a fixed inductance and a variable capacitance, such as that provided by a variable air condenser. With this arrangement, the frequency of the oscillations in circuit *A* will have a definite value, fixed by the circuit constants, while the natural frequency of circuit *B* is adjustable by means of the variable condenser.

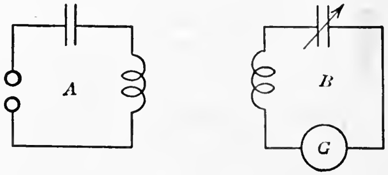


FIG. 34.

The effect of the oscillatory alternating current flowing in circuit *A* at each spark across the gap *G* is then to induce in circuit *B* an alternating emf. of decreasing amplitude, and therefore to produce in the circuit an alternating current of similarly decreasing value. As was explained above, if the natural frequency of circuit *B* were the same as that of the oscillatory current in circuit *A*, then the total impedance of *B* would be a minimum equal to the resistance of the circuit, and the emf. induced by *A* in *B* would produce in the latter the maximum current which would be limited only by the resistance of circuit *B*. This condition may be accomplished by *tuning* circuit *B* to resonance with circuit *A*. This is done by adjusting the variable condenser of circuit *B* until the natural frequency is equal to that of circuit *A*. If an ammeter or galvanometer is inserted in circuit *B*, it will indicate a maximum current when the two circuits are in tune.

There are certain remarks of interest to be made in connection with the behavior of the circuits under the above described conditions. When circuit *B* is not in resonance with circuit *A*, and especially if the oscillatory current in circuit *A* is of a high frequency, its reactance is quite considerable, so that the current induced in it is very small. Circuit *B* then does not oscillate, and the current in it is simply the induced current from A, and it follows the latter exactly in its decreasing amplitude. In other words, the current induced in B is said to be "forced," and it has the same decrement and characteristics as the current in circuit A.

Now if the two circuits are in tune, circuit B will oscillate, as was explained in a previous paragraph on the resonance of a single circuit. For then, after the emf. induced in B by A has charged the condenser in circuit B, the energy supplied by circuit A to circuit B will be used in overcoming the resistance losses in that circuit, and further charge the condenser synchronously, while the condenser of circuit B will discharge in the manner previously described. Conditions in circuit B are then the same as though an alternator were inserted in series with it and generating an emf. equal at every instant to the emf. actually induced in circuit B by circuit A. This case was studied previously, except that, since the oscillation in A is damped, a continually smaller amount of energy is supplied by circuit A to circuit B, and the oscillation in the latter will be damped.

If the decrement of circuit B is *smaller* than that of circuit A, the oscillation in circuit B will be damped less than that in circuit A, and may still continue after the oscillation in the latter has died out. If on the other hand the decrement of circuit B is *equal to* or *greater* than that of circuit A, or even if the circuit B has so high a resistance as to be *aperiodic*, then the oscillation induced in it will be of the nature of a forced oscillation, and will have the same decrement and characteristics as the oscillation in circuit A.

Another effect of the resistance of circuit B on the current which may be induced in it by circuit A is that the maximum current occurring in circuit B when the latter is brought to the resonance point is smaller the greater the resistance. This results in a flattening of the resonance curve, as shown in Fig. 27, and therefore decreases the sharpness of the resonance. This indicates that the greater the decrement of circuit B, the more the oscillation in it will be "forced" instead of "free," and the less the characteristics and constants of circuit B will affect the amplitude of the current flowing through it. And if circuit B has so high a resistance as to be aperiodic, the reactance of the circuit will be such a small fraction of its impedance, that the resonance curve will be quite flat, and tuning will not be sharp.

It should be noted that when circuit A is coupled to another circuit B in which it sets up an electric current, energy is transferred from the former to the latter circuit. It follows that the energy losses in circuit A are not only those due to resistance and spark gap losses as in the case of the single oscillatory circuit considered at the beginning of this chapter, but also those due to energy transfer or radiation. This results in a more rapid damping of the oscillation in circuit A, and therefore in an increase of the decrement of the circuit. The decrement of a circuit is thus seen to have several components, such as a resistance, a spark gap (if any), and a radiation component. Other factors or components may be present in cases of high condenser voltages, due to energy losses resulting from brush discharge, corona, and leakage currents through a poor condenser dielectric or poor insulation of the circuit.

Circuits Closely Coupled.—If the two circuits A and B of Fig. 34 are closely coupled, the phenomena outlined under the preceding heading will be modified, for then the current in circuit B resulting from the emf. induced in it by the current in circuit A will in turn induce an emf. in circuit A, which, as a consequence of the considerations of Chapter I, will be in such a direction as to tend to reduce or oppose the current in circuit A. Only the case generally encountered in radio will be considered here, namely, that of two circuits of sufficiently low resistance to be oscillatory. The case of a closely coupled aperiodic secondary circuit is much the same as the case of loosely coupled circuits, since the secondary circuit B, having no free oscillation of its own, cannot produce any other effect on the oscillation in circuit A than to increase the decrement of that oscillation, as explained above.

Consider then, the two oscillatory circuits A and B as described in the previous section, but closely instead of loosely coupled. Both circuits will be assumed to have the same natural frequency. An oscillation is started in circuit A by first charging its condenser until the spark gap breaks down, when the condenser discharges at the natural frequency of the circuit. This oscillatory discharge, represented by the upper curve of Fig. 35, will be damped due to resistance losses in the circuit, and radiation losses. The latter, as was already explained, are due to a continual transfer of energy from circuit A to circuit B. This corresponds to a continual gain of energy on the part of circuit B, in which the induced alternating current increases in amplitude. This is shown in the lower curve of Fig. 35.

At a certain time t_1, all the energy originally stored in circuit *A* will have been dissipated in the resistance of the circuit or transferred to circuit *B*. The latter circuit at the same instant is in full oscillation, having continually received energy from circuit

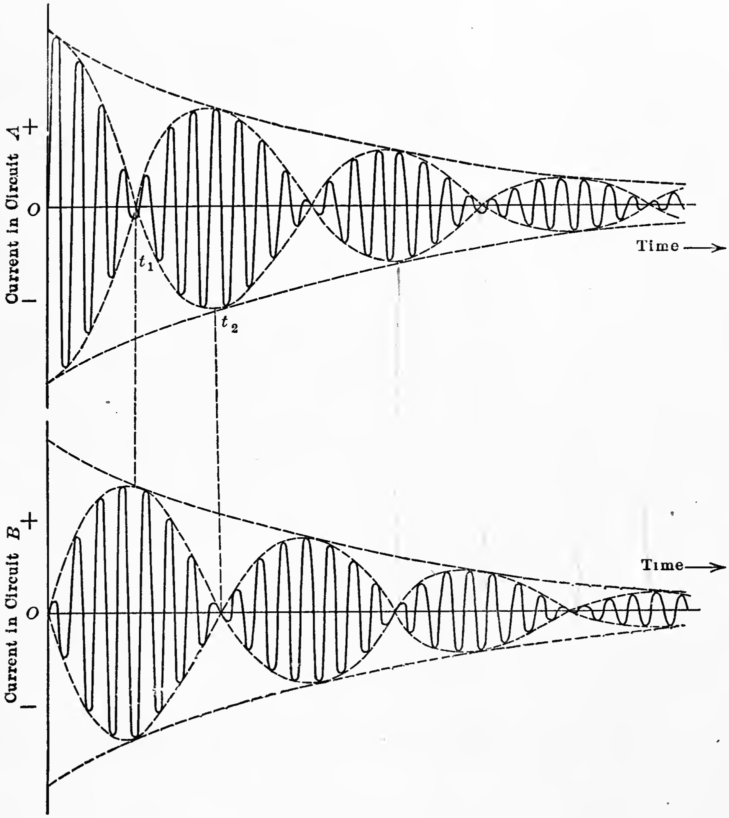


Fig. 35.

A up to this time t_1. This supply of energy having now disappeared due to the total damping out of the oscillation in circuit *A*, the oscillation in circuit *B* will start decreasing in its turn, on account of the resistance losses in circuit *B*. Also, due to the extremely short time required for the entire process, the conducting gases and vapors in the spark gap of circuit *A* have not had time to escape, or cool down, or de-ionize so that the gap is still in a conducting s king A virtually a closed circuit. The alternating current in circuit B will therefore react upon circuit A and induce an emf. which will again set up an ting current in circuit A. There will thus be a transfer of e rom circuit B to circuit A during the entire duration of the ation in circuit B. This will increase the decrement of the o ation in circuit B, until at the time t_2 this oscillation will have entirely died out, while at the same instant circuit A will again be oscillating with an amplitude less than that of its original oscillations by an amount represented by the losses in resistance of the two circuits and the double transfer. The alternate transfer of energy from one circuit to the other thus repeats itself a number of times until all of the energy has been dissipated in the resistance of the two circuits and in the other losses.

If the purpose is to set up an oscillatory current in circuit B, without the alternate transfer of energy between the circuits, it may be seen from the preceding paragraph that this is not possible unless some means is employed to prevent the reaction. This may be accomplished as follows:

If at the time t_1 when all the energy of circuit A has been transferred to circuit B, the spark gap in circuit A is made non-conducting, circuit A will be open and incapable of absorbing any energy back from circuit B. If this is done, then, the result is shown in Fig. 36, as contrasted with Fig. 35. The oscillation set up in circuit A dies out at the time t_1 after having transferred to circuit B that part of its energy which is not lost in its resistance. At the time t_1, the gap in circuit A is made non-conducting by one of the methods indicated later. At this time, circuit B is in full oscillation, and since it can set up no current in circuit A, which is now open, the oscillation in circuit B will decrease at a rate determined solely by the decrement of circuit B, all the energy which circuit B had received from circuit A being used in circuit B itself without any transfer back to the first circuit.

With this method of setting up oscillations in the secondary circuit or, as frequently stated, of exciting the secondary circuit, the primary circuit generally oscillates for only a very short length of time as compared with the duration of the secondary oscillation. The action of the primary current may be compared to a hammer blow which, although lasting a very short time, may set a gong into vibration for a considerable length of time. For this reason, this method of exciting the secondary circuit is sometimes called "impulse excitation."

With this method of excitation, it is seen that the secondary circuit is the main oscillatory circuit, and does not include any spark gap. It is generally a low resistance circuit. The spark gap is of a special "quenched" type, and is located in a primary

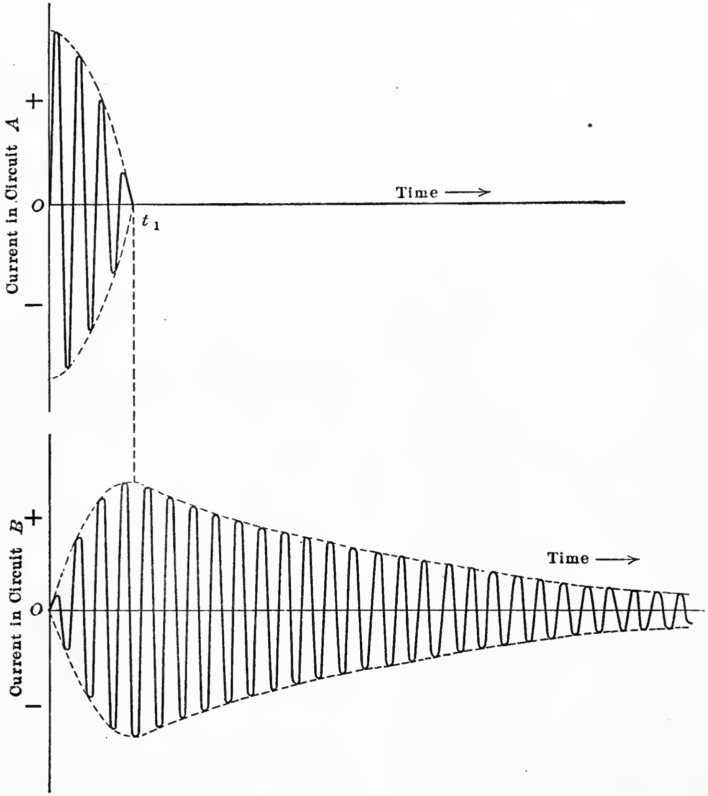


FIG. 36.

or auxiliary oscillatory circuit. This eliminates spark gap losses from the secondary circuit, in which the oscillations are therefore less damped than in the case of a single oscillatory circuit having an open spark gap connected in series, as was considered at the beginning of this chapter.

Spark Gaps and Quenching.—Among the various methods used for making the spark gap non-conducting after the first complete transfer of energy from the primary to the secondary, those described below are probably the most important. Before entering into the detail description of these gaps, it is well to have an idea of how a spark jumps across a gap.

Consider the spark gap of Fig. 37, consisting of two brass balls A and B separated by a short gap and connected to some source of electrical energy establishing a difference of potential between them. The gas surrounding the two balls and filling the gap, which may be air if the gap is open, is made up of molecules or atoms, as explained in Chapter I, and contains also a few free electrons. Under the action of the electromotive force applied across AB, these free electrons are set in motion through the gas, being repelled by the negatively charged ball and attracted by the positively charged ball. In their motion through the gas, the electrons collide with the atoms, and if the emf. is great enough the speed of the electrons may be so high that they will disrupt or break up the atoms by their collisions, knocking off the atoms, so to speak, some of the electrons which normally rotate around the central positive charge of the atom. Each atom which is thus broken up, having lost one or more of its electrons, will have a resultant positive charge, since its central positive charge is now greater than the sum of the negative charges left in rotation about it. The atom in this state is said to be an "ion," and the gas is said to be "ionized." These positively charged ions will travel in a direction opposite to that of the free electrons, thus increasing the electric current in the gap, while at the same time the number of free electrons in the gap keeps increasing as they are torn off the atoms by collision. If the applied emf. is sufficiently large, this electric current taking place between the balls of the spark gap will manifest itself in the form of a visible spark between the balls. This is accompanied by a comparatively sudden and rapid disruption of the molecules of the gas and a rise in temperature of the gas and electrodes. The temperature rise may be great enough to vaporize some of the metal of the electrodes and this vapor, filling the gap, decreases the resistance of the path of the spark and thus increases the current through the gap.

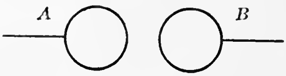


Fig. 37.

Factors affecting the resistance of the spark gap are therefore the nature and pressure of the gas filling the gap, the nature of the metal of the spark gap electrodes, and the shape of these electrodes, inasmuch as this shape affects the potential gradient in the gap and also the temperature of the spark by the rate of dissipation of the heat through the mass of the electrodes. In order to rapidly extinguish or quench the spark, as is required for

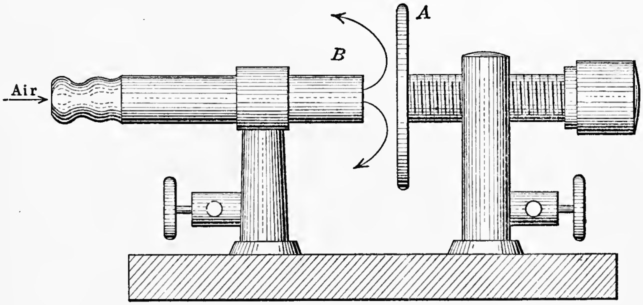
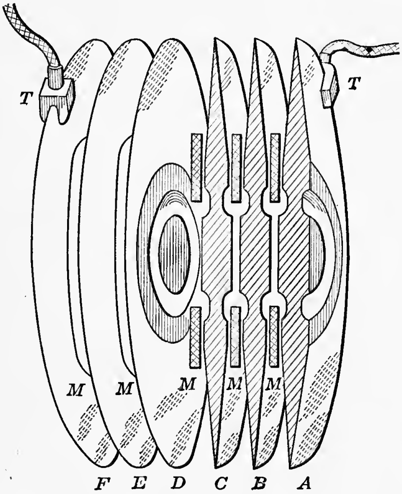


Fig. 38.

impulse excitation, the following methods may be used: rapid de-ionization of the gas; prevention of the formation of metal vapors; cooling of the spark and of the electrodes; increase of the gap length after the spark has been started.



Fig' 39.

One type of gap effects a rapid de-ionization and cooling by continually changing the air within the gap. It is shown schematically in Fig. 38. One electrode of the gap is a metal disc *A*, while the other electrode is a metal tube *B*, which is connected to a blower. Air is blown through this tube against the metal disc and is thus continually renewed. The sparking distance may be changed by means of the screw adjustment of the disc, which may be brought near to or farther from the metal tube.

Another design which is very extensively used on account of its simplicity and the absence of any external blower or motor is shown in Fig. 39. The purpose of this gap is to rapidly cool the spark. This is attained by breaking up the spark into a number of smaller sparks of short length and all in series. The gap comprises a number of silver discs *A*, *B*, *C*, etc., equipped with cooling flanges and mounted parallel and very close to each other but separated by mica, fiber or rubber washers *M*. The central part of the discs is finely polished and forms the sparking surface. According to the breakdown voltage desired, a number of these small gaps are used, connection being made to the extreme discs by means of clips *T*. As the breakdown voltage is reached, all the gaps break down, the total length of the gaps being approximately equivalent to that of the single gap of Fig. 37 when set for the same voltage. The difference from this latter design is however that, with the gap of Fig. 39, the spark is cooled through the radiation of heat by the metal discs, and is thus rapidly extinguished or quenched.

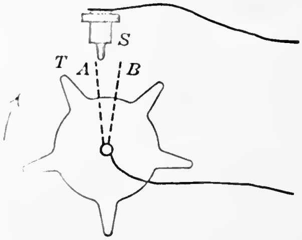


FIG. 40.

A third model of gap, shown in Fig. 40, is known as the rotary gap and is used principally for high power radio sets though frequently for low power sets, when the primary oscillatory circuit is energized by an alternator in a manner described in an early paragraph of this chapter. One electrode of this gap is a stationary metal stud *S*, generally made of some refractory metal to enable it to withstand the heat of the constant sparking. The other electrode is a toothed metal disc or wheel which is rotated close to the fixed electrode. A circuit illustrating the use of this gap is given in Fig. 41. Such a circuit would not be used in practice, as will be explained in a future chapter, but it gives an idea of the operation. This is the same circuit as shown in Fig. 20, except that the ordinary open gap is replaced by the rotary gap. The latter is rotated, for best operation, at a speed so related to the speed of the alternator charging the condenser, that the product of the number of teeth by the revolutions per second of the disc is equal to twice the frequency of the alternator.

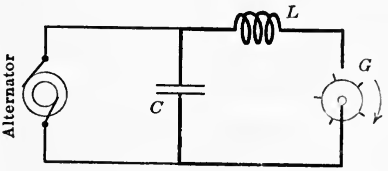


FIG. 41.

Thus,

$$2 \times \text{alternator frequency} = \frac{(\text{number of teeth}) \times (\text{revolutions per minute})}{60}.$$

If this condition is fulfilled, then one tooth of the rotary electrode will pass in front of the stationary electrode at every half cycle of the alternating supply voltage, and there will be one discharge per half-cycle, provided the gap is properly adjusted. This adjustment should be such that the spark will start at a point *A*, Fig. 42, of the alternator voltage cycle, and be extinguished at some point *B*, the oscillatory discharge of the circuit *LCG* taking place in the time *ab*. (See also Fig. 22.) This may be obtained by a suitable angular and radial adjustment of the stationary electrode with respect to the rotary disc, the tooth *T*, Fig. 40, being at position *A* when the alternator voltage is *aA*, Fig. 42. With the gap properly set, the spark starts when the tooth comes in the position *A*, just a little before the tooth is exactly in front of the stud *S*. This makes allowance for the time lag of the spark required by the process of breakdown of the air in the gap as explained above. As the disc revolves farther, the tooth is removed from the stud *S*, increasing the length and resistance of the spark which is finally broken at about position *B*.

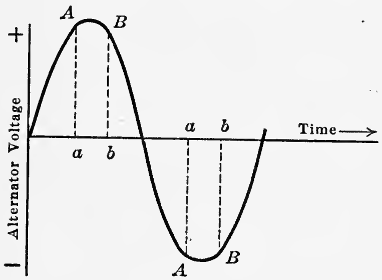


FIG. 42.

With a 500-cycle alternator and a 10-tooth disc, the speed of the disc for one spark for each half-cycle would be

$$\frac{500 \times 2}{10} \times 60 = 6000 \text{ r.p.m.}$$

The rotary gap is generally mounted on the alternator shaft and the number of teeth required determined according to the number of poles of the machine, which facilitates the angular setting of the gap.

The gap may be operated non-synchronously, but this is not as good practice, for the spark will occur at different voltages for successive half-cycles of the supply voltage, thus producing an irregular operation of the system.

Resonance Curves for Coupled Circuits.—The above study of resonance phenomena in two coupled circuits was made with the assumption that the two circuits were in tune with each other, that is, that they had the same natural frequency—realized when the product of inductance by capacitance is the same in each circuit, neglecting the resistance.

A study will be made here of the variation of the current in a circuit containing inductance and capacitance when excited by another oscillatory circuit, the natural frequency of the exciting circuit being kept constant, while that of the excited circuit is varied by varying its inductance or capacitance. The excited or secondary circuit *B*, Fig. 43, will also be assumed to comprise a current indicating instrument *G*, which may be a hot wire galvanometer, hot wire ammeter, thermocouple, etc.

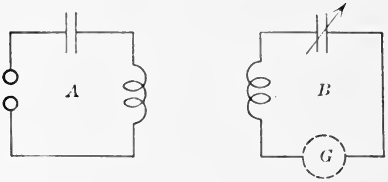


Fig. 43.

As was already pointed out, the effect on circuit *B* of the alternating emf. induced in it by the oscillatory field due to the current in circuit *A* is equivalent to that of an alternator inserted in the circuit *B* and generating an emf. equal at all times to the induced emf. Under those conditions, and with the circuits *loosely* coupled, a resonance curve will be obtained by plotting the current in circuit *B* against its natural frequency as varied by adjusting the variable condenser, similar to that obtained in the case of series resonance of an oscillatory circuit, previously considered. This curve is shown as curve 1, Fig. 44, exhibiting a sharp maximum when circuit *B* has a natural frequency *F* equal to the frequency of the oscillating field exciting it, that is, equal to the natural frequency of circuit *A*.

Now if the coupling of the two circuits is increased, for instance by placing circuit *B*, Fig. 43, nearer to circuit *A*, the field of circuit *A* linking with circuit *B* will be stronger and it may be expected that the current induced in circuit *B* will be greater. However, since the two circuits are more *closely* coupled, the secondary circuit *B* reacts upon circuit *A*. That is, the current induced in circuit *B* will be in such a direction as to set up around the circuits a field opposed to the inducing field of circuit A, and this effect will be stronger, the greater the current flowing in circuit B. The effective field is then equal to the difference between the inducing field of circuit A and the reacting field of circuit B. As the natural frequency of circuit B is gradually increased from zero toward the resonance frequency F, by means of the variable condenser, for instance, the current in circuit B will then increase, as shown by curve 2, Fig. 44. This will also increase the reacting or opposing field, correspondingly reducing the effective or inducing field. And a point will be reached below resonance

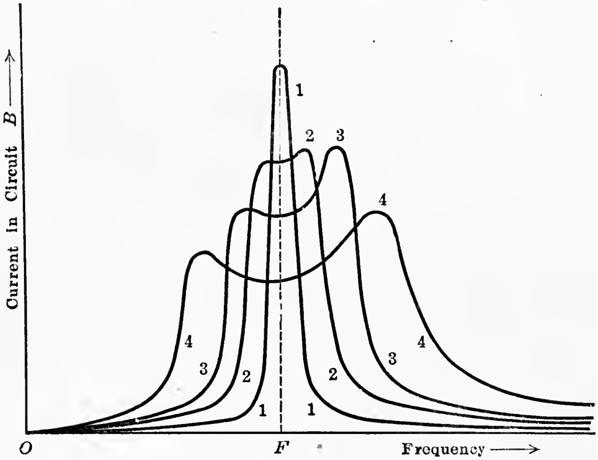


Fig. 44.

frequency where the current in circuit B will cease to grow. The effect of increased coupling of the two circuits is thus to produce in circuit B a flat topped resonance curve, thus decreasing the sharpness of resonance. In fact, when the decrement of circuit B is less than that of circuit A, the reacting field may then become so great as to produce a decrease in the induced current. When the natural frequency of circuit B is then increased beyond the resonance frequency F, the same phenomena occur in the opposite sequence, the current in B will then rise to a second maximum and finally decrease, producing a double humped resonance curve. This effect increases with the coupling, as shown by curves 3 and 4 which correspond to gradually closer or tighter couplings of the circuits.

In the case of closely coupled circuits, the current induced in the secondary circuit B is thus seen to pass through *two* maxima, corresponding to two natural frequencies of circuit B, one being below, and the other above the frequency of the exciting current. The looser the coupling, the nearer will these frequencies come to the true resonance frequency.

Critical Coupling.—Again consider the two circuits of Fig. 43, which this time will be assumed to be both tuned to the same frequency, that is, so that:

$$\sqrt{\frac{1}{L_A C_A} - \frac{R_A^2}{4L_A^2}} = \sqrt{\frac{1}{L_B C_B} - \frac{R_B^2}{4L_B^2}}$$

To fix ideas, suppose these circuits to be inductively coupled. The degree of coupling will then vary in the same manner as the mutual inductance M. If the circuits are at infinite distance from each other, the field of circuit A linking with circuit B will be zero, and the current induced in this latter ciicuit will also be zero. As circuit B is brought closer to circuit A, that is, as the

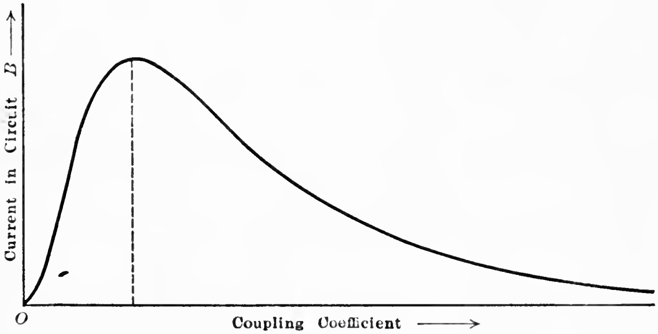


FIG. 45.

mutual inductance M, and thereforе the coupling, is increased, the current induced in B will correspondingly increase. A point will be reached where the reaction of this induced current upon circuit A will be such that the current in B will cease to increase, and, since this reaction grows stronger as the circuits are coupled closer, the current in circuit B will decrease. This is illustrated in Fig. 45, where it is seen that the current induced in circuit B passes through a maximum corresponding to a certain critical value of coupling. This value depends on the mutual inductance and also on the resistance of the circuits, since a given emf.

induced in B will produce in that circuit a current determined in part by the resistance of the circuit, which then also partly governs the reaction of circuit B on circuit A. It is thus seen that the amount of energy transferred from circuit A to circuit B passes through a maximum as the coupling is increased.

Another way to explain this action is to consider the emf. across the condenser of circuit A. The oscillatory current flowing under the effect of this emf. varies inversely with the impedance, which in turn is equal to the sum of the impedance of circuit A and the mutual impedance of the two circuits. For the case here considered, the critical coupling coefficient is obtained for

$$M = 2\pi f\sqrt{R_A R_B}$$

Mathematical Theory of Resonance Phenomena.[1]—A simple mathematical computation of the two frequencies corresponding to the current maxima of the resonance curve of two coupled circuits may be obtained by considering their equivalent directly coupled circuit. Thus, consider the circuits A and B of Fig. 31 and let L_A and L_B be the total inductance in each circuit, and M the mutual inductance. The system is equivalent to that of Fig. 46 where the total inductance of each circuit is respectively the same as for the separate circuit, but where a part M of the total inductance of each is common to both. Thus:

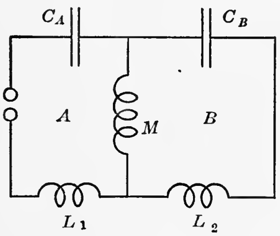


Fig. 46.

$$L_A = L_1 + M$$
$$L_B = L_2 + M$$

Considering then the circuit of Fig. 46, and neglecting all resistance, the oscillatory emf. E acting in circuit A will set up in circuits A and B the currents I_A and I_B given by the relation:

$$E = \left(2\pi f L_1 - \frac{1}{2\pi f C_A}\right) I_A + 2\pi f M \left(I_A - I_B\right)$$

which may be obtained by the application of the general relation previously established, that

$$E = \frac{I}{Z}$$

[1] See Bureau of Standards Circular No. 74, page 52.

In order to eliminate I_B, it may be stated that the emf. across the inductance M is at every instant equal to that across the series combination C_B L_2. Thus:

$$2\pi fM\left(I_A - I_B\right) = \left(2\pi fL_2 - \frac{1}{2\pi fC_B}\right) I_B$$

from which,

$$I_B = \frac{2\pi fM\left(2\pi fL_2 - \dfrac{1}{2\pi fC_B}\right)}{2\pi f(L_2 + M) - \dfrac{1}{2\pi fC_B}} I_A$$

and finally, referring back to the above expression for E,

$$E = \left[2\pi fL_1 - \frac{1}{2\pi fC_A} + \frac{2\pi fM\left(2\pi fL_2 - \dfrac{1}{2\pi fC_B}\right)}{2\pi f(L_2 + M) - \dfrac{1}{2\pi fC_B}}\right] I_A$$

The value of frequency for which the current is a maximum is then that which will make the reactance, as expressed in brackets, equal to zero. By solving with respect to the frequency f, two values are found, and if the natural frequencies of circuits A and B are supposed to be equal, that is, if

$$\frac{1}{2\pi\sqrt{L_A C_A}} = \frac{1}{2\pi\sqrt{L_B C_B}} = F$$

then these two values of frequency giving maximum current will be:

$$f = \frac{F}{\sqrt{1 \pm k}}$$

where k is the coefficient of coupling of the circuits, that is,

$$k = \frac{M}{\sqrt{L_A L_B}}$$

From this result it may be seen that for zero coupling, there is only one maximum of current, for the frequency $f = F$ which is the natural frequency of the two circuits. For maximum coupling, when $M = L_A = L_B = \sqrt{L_A L_B}$, the coefficient k is equal to unity and the two frequencies are

$$f = \infty$$
$$f = \frac{F}{\sqrt{2}}$$

so that in practice, there will be found one resonance frequency in the case of maximum coupling, but this will be different from the natural frequency F of the two circuits.

Wavemeters.—An application of the resonance phenomena treated above is embodied in the "wavemeter." This is a device which may be used for measuring the frequency of an oscillatory or alternating current, the natural frequency of a circuit, the capacitance of a condenser, the inductance of a coil, the decrement of a circuit or an oscillation, and the radio wave length which is a quantity defined in the next chapter. It may also be used for generating oscillations of a known frequency.

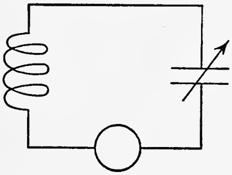


Fig. 47.

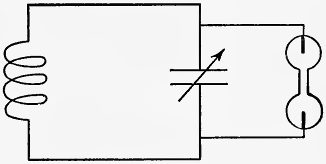


Fig. 48.

The wavemeter is simply a closed low resistance oscillatory circuit, the natural frequency of which may be varied by means of a variable condenser or inductance coil. It is calibrated so that its natural frequency or its constants are known for each position of the variable condenser or inductance. When used as a measuring instrument, the wavemeter circuit also comprises some sensitive current or voltage measuring device. Thus, Fig. 47, a hot wire ammeter may be connected in series with the inductance coil and condenser, or a thermocouple connected to a sensitive galvanometer, or else, Fig. 48, a neon tube or Geissler tube may be connected in shunt with the condenser.

Frequently such current measuring devices as ammeters and thermocouples are of relatively high resistance. In order to keep down the resistance of the wavemeter oscillatory circuit and thereby increase the sensitivity of the wavemeter by reducing the energy losses in it, the current measuring instrument is not inserted directly in the circuit, but is coupled to it, Fig. 49.

If the wavemeter is not used for actual measurements, but simply as an indicating instrument, the sensitive device may be a small low resistance bulb, which will glow when the current in the circuit is a maximum. Another indicating device is the telephone receiver, which is frequently used in conjunction with a rectifying detector when working with damped, or modulated, or discontinuous oscillations. The use of the telephone receiver is fully explained in a later chapter on radio reception.

The use of the wavemeter for measuring the frequency of an oscillation has already been explained, and reference is made to Fig. 43 above. It is simply necessary to loosely couple the wavemeter circuit to the circuit under test, and vary the wavemeter constants until a maximum current is observed through the medium of the indicating device employed. The wavemeter circuit will then be tuned to the frequency of the oscillation, and this frequency may readily be obtained from the wavemeter settings and calibration. The "*wave length*" of the oscillation, which, as is explained in the next chapter, is equal to the ratio of the speed of light (300,000 kilometers per second) to the frequency of the oscillation, is then obtained at the same time, many wavemeters being calibrated directly in wave lengths.

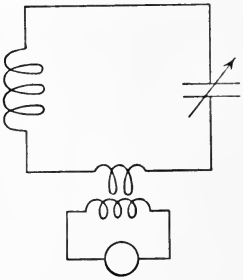


Fig. 49.

The wavemeter may be used to measure the decrement of an oscillation, and if this oscillation is generated by a circuit oscillating freely, such as would be the case if it were excited by the impulse method explained before, the decrement measured will also be that of the oscillatory circuit. The measurement is effected by tuning the wavemeter to resonance as just explained, the reading I_r of the ammeter being recorded, and also the corresponding capacitance C_r of the wavemeter variable condenser. This condenser setting is then varied until a current equal to $\frac{I_r}{2}$ is observed, when the capacitance C of the wavemeter condenser is recorded. The decrement D of the oscillation is then given by the relation:

$$D + d = \pi \frac{\pm (C_r - C)}{C}$$

where d is the decrement of the wavemeter circuit, which, expressed as a function of the wavemeter constants, is equal to

$$\pi R \sqrt{\frac{C}{L}}$$

Wavemeters may be especially constructed and calibrated to read decrement directly, and are then called "decremeters." The description and complete theory of this instrument is given in the Bureau of Standards Scientific Paper No. 235.[1]

[1] It is very simple to make any ordinary wavemeter into a direct reading decremeter by merely using a scale calculated in the proper way. Such a scale is given in Fig. *A*. It can be used on any variable condenser with semicircular plates regardless of the kind of capacitance scale on the con-

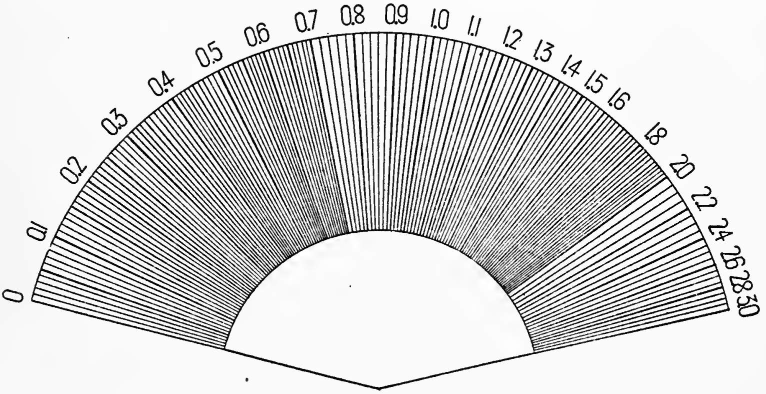


Fig. *A*.

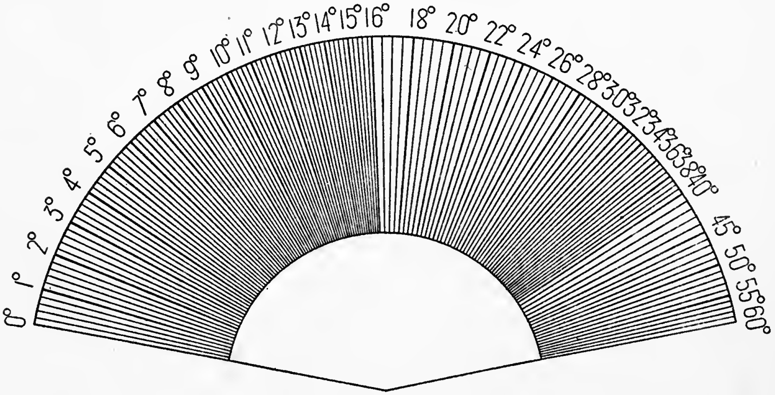


Fig. *B*.

denser or even if the condenser has no scale whatever on it. The scale here given is simply cut out, trimmed at such a radius as to fit the dial and then affixed to the condenser with its zero point approximately in coincidence with the graduation which corresponds to maximum capacitance. This usually puts it on the unused half of the dial opposite the capacitance scale. The scale may be used either with moving condenser and stationary

The measurement of decrement may be done in a similar manner with a wavemeter having a variable inductance instead of capacitance. This is however not quite as satisfactory, for a change in the inductance generally alters the degree of coupling of the wavemeter to the circuit generating the oscillations, and the change in the current, as indicated by the ammeter in the wavemeter circuit, is then due to this change in coupling in addition to the change in the natural frequency of the wavemeter circuit, thus making the result less reliable.

The capacitance of a condenser may be determined by means of a wavemeter by connecting the condenser to the two terminals of an inductance coil of known inductance L and of a resistance sufficiently low to be negligible. The oscillatory circuit thus formed is then excited, and the frequency f of the free oscillations taking place in it is measured with the wavemeter as explained above. The unkown capacitance is then calculated from the relation:

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

Similarly, the inductance of a coil may be determined by connecting the coil to a condenser of known capacitance and proceeding as above.

Finally, the wavemeter may be used to generate oscillations of a standard or known frequency. This is done by shunting the wavemeter condenser by a buzzer and battery, Fig. 50. The wavemeter circuit is set to the desired frequency, and the buzzer is set into operation. When the buzzer vibrator closes the contact A, the battery is

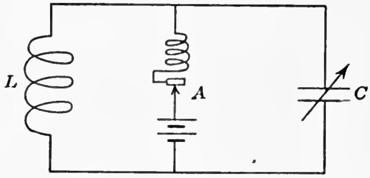


Fig. 50.

dial or with moving dial. A reading of decrement is made by first observing the current-square at resonance, then reading the scale at a setting on each side of resonance for which the current-square is one-half its value at resonance. The difference between the two readings on the scale is the value of the decrement. If it is desired to convert the wavemeter into an instrument for the direct measurement of phase difference or power loss, the scale in Fig. B is used in identically the same way. The difference between the two settings on this scale is the phase difference in degrees. These schemes for converting wavemeters were devised by Dr. J. H. Dellinger of the Bureau of Standards.

connected across the condenser through the buzzer winding and thus charges the condenser. But the buzzer circuit being also connected across and completed through the wavemeter inductance L, the vibrator is attracted toward the buzzer coil, and the contact A opened. The wavemeter condenser C, which was charged in the previous moment, then discharges through the coil L at the natural frequency of the wavemeter circuit. As the buzzer vibrator springs back in place, the condenser recharges, and the process repeats itself, there being one oscillatory discharge for every vibration of the buzzer.

The various uses of the wavemeter will be frequently referred to in the course of this book.

CHAPTER III

ANTENNA SYSTEMS AND RADIATION

Radiation Theory of Radio Communication

Closed and Open Oscillators.—Consider the oscillatory circuits of Fig, 51 containing a condenser *C* connected to an inductance coil *L*. Oscillations may be set up in this circuit, as explained in the previous chapter, for instance, by impulse excitation. When a high frequency alternating current is thus made to flow in the circuit, similarly alternating interlinked electrostatic and electromagnetic fields are set up around it. These fields, theoretically, will extend throughout space, and will therefore induce an alternating emf. in any circuits which may be present and in a suitable position. Such a circuit as Fig. 51, however, cannot be used for long distance radio transmission, as may be seen from the following:[1]

Fig. 51.

Consider a point *P*, Fig. 51, at a comparatively great distance from the oscillatory circuit, say a distance of several miles. The alternating current flowing in the circuit *LC* will flow alternately clockwise and counter-clockwise. That is, the current in the section *AB* of that circuit will alternately flow upward and downward, creating an alternating field at point *P*. But section *A'B'* of the circuit will at every instant carry a current equal to that in the section *AB*, but in the opposite direction, creating at point *P* a field neutralizing that of the section *AB*. As a result of this neutralizing action of one-half of the oscillatory circuit upon the other half, there will be no appreciable radiation from the circuit, exept when the point *P* is so near to the circuit

[1] This is true except when the c cuit is especially designed and used as a loop antenna (see Chapter XI).

that its distance from the various sections of the circuit can no longer be considered as equal.

Since a "closed" oscillatory circuit generally speaking is not suitable for transmitting energy at great distances, use is made of a so-called "open" oscillatory circuit, the principle of which may be illustrated by the circuit of Fig. 52. Consider a straight metal wire $A'B'$, cut in its center by a small sphere gap AB. From the definition of a condenser, this system is seen to be a condenser, and may be made to store a certain amount of electrostatic energy. Thus, connecting A and B to a source of continuous potential, will charge this condenser by removing some free electrons from one wire and placing them on the other. In the case of Fig. 52, the upper wire AA' will be positive and lose some electrons, while the lower wire BB' will gain a like number of

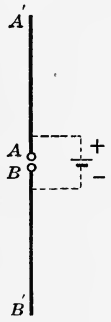


Fig. 52.

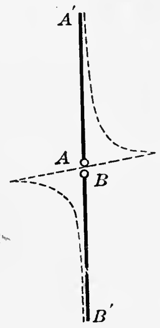


Fig. 53.

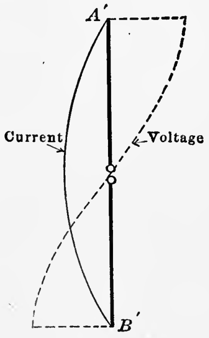


Fig. 54.

electrons, and become negatively charged. This will establish along the wire $A'B'$ a potential gradient, as shown in Fig. 53, where the deficiency or excess of electrons is plotted at each point of the wire $A'B'$ perpendicularly to the axis of the wire. This therefore creates an electrostatic field about the wire, and represents a store of electrostatic energy.

If the potential difference between the two sections AA' and BB' is large enough, a spark will bridge the gap and the electrons accumulated on the lower section BB' of the wire will start moving upward under the action of the electrostatic field into the upper section AA' in a tendency to reëstablish the original electrostatic equilibrium. This will create an electric current flow in the wire $A'B'$, with a corresponding and continual decrease of the electrostatic energy, and the creation of an electromagnetic field and store of electromagnetic energy, as explained in Chapter I. When the entire amount of electrostatic energy has been transformed into electromagnetic energy, with a partial loss due to the resistance of the circuit, the difference of potential across AB will be zero, but, due to the presence of the electromagnetic field, and exactly as in the case previously considered of an ordinary oscillatory circuit, the current will continue to flow in the direction $B'A'$ until the entire amount of electromagnetic energy has disappeared. This will accumulate a number of electrons on AA', and remove a like number from BB', charging the system with a polarity opposite to the original one. Also, as may be understood from the process, and as will be better explained later, the greatest accumulation and deficiency of electrons, during the latter part of the discharge, will be at the extreme ends of the wire, while the current flow will be a maximum at the center. This is represented by the curves in Fig. 54. After this first discharge, the oscillator will begin discharging in the opposite direction, and so on, until the entire original store of energy has been dissipated.

It is thus seen that since the system $A'ABB'$ has inductance and capacitance, an oscillatory current may be set up in it by the sudden discharge of an original store of electrostatic energy, exactly as in the case of the oscillatory circuits considered in the previous chapter.

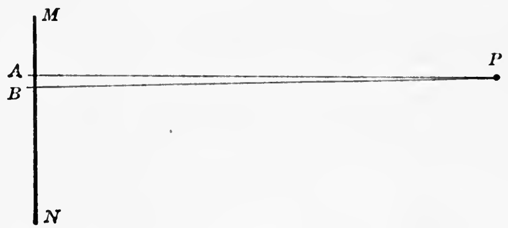


FIG. 55.

Consider then an open oscillatory circuit MN, Fig. 55, of the kind just described. When oscillations are set up in this circuit, a current will flow in it, alternately upward and downward, and will at every instant, be in the same direction at all points of the wire MN. The alternating field set up at a distant point P by this current in a section AB of the circuit MN, will then no longer be in opposition with the field due to some other section of the circuit, as was the case in Fig. 51, and it is thus seen that energy radiation will take place much more effectively.

In practice, ranges of many miles are required, and the radiating oscillatory circuit or *antenna* is generally made of large dimensions, as explained at the end of Chapter I. The wires are supported by insulators and wooden or metal masts or towers, and given varied shapes, as studied in a later section.

This chapter is devoted to the study of the more important properties of such open or radiating oscillatory circuits, also generally known as "antennæ." They are most widely used in radio for effecting the radiation into space through considerable distances of the high frequency oscillatory electric energy generated usually by means of closed, non-radiating oscillatory circuits. Similar antennæ are also used for the reception of the radiated energy. This will be studied in a later paragraph.

Wave Propagation of Radiated Energy.—It was pointed out in Chapter I that any disturbance in an electrostatic or electro-

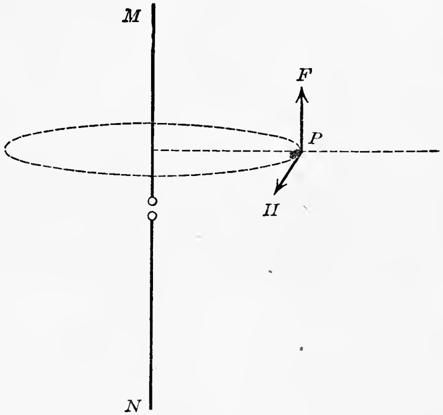


Fig. 56.

magnetic field is propagated throughout space at a definite constant speed in the form of a "wave." This is of especial interest in the study of the process of the transfer of electrical energy from one radiating circuit to another and is therefore taken up here in somewhat greater detail.

Assume an oscillating current to be set up in an open oscillatory circuit *MN*, Fig. 56, of the type described above. For the present discussion, it may also be assumed that the oscillation is undamped, and therefore of the nature of an ordinary high frequency alternating current.

Referring to what was said in connection with Fig. 17, it will be recalled that at each point P of space, there is an electrostatic force F and an electromagnetic force H, the values of which depend on the distance of P from the circuit, and on the strength of the electrostatic and electromagnetic fields of the oscillatory circuit MN, and therefore on the current flowing in that circuit. If the current in MN is alternately reversed, the forces F and H at all points P of space will also be reversed. And since the electrostatic and electromagnetic fields are each a maximum when the other is zero, as was explained for the oscillatory discharge of a condenser, the same will be true of the forces F and H. The direction of the magnetic force H is tangent to a circle passing through point P, having its center at the point of intersection of MN with the perpendicular to MN drawn through point P, and its plane normal to MN. The direction of the electrostatic force F is normal to the plane of the circle just fixed.

Now when an alternating current is made to flow in MN, the field around the wire will reverse periodically. This reversal will however not occur simultaneously at all points of space, but will travel from the wire MN outward into space in all directions with a definite speed, which is that of luminous radiations or light. Thus, Fig. 57, in the first half-cycle, when the alternating current rises from a minimum to a maximum, the magnetic field at a point P_1 in the immediate vicinity of the wire will rise similarly and simultaneously. At point P_2, the field will also rise, but it will start doing so a certain fraction of time after the corresponding change in field at point P_1, the time of lag being that required by the disturbance wave to travel from P_1 to P_2. Similarly, the field at points P_3, P_4, P_5, etc., will increase, but it will start increasing only when the wave traveling with the speed of light has reached them in turn.

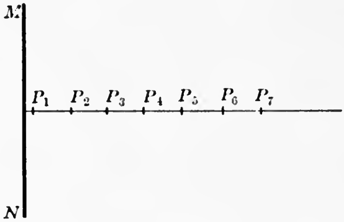


Fig. 57.

Now suppose that the alternating current in MN has reached its maximum value and started decreasing when the wave of increasing field has reached point P_4. This wave will go on traveling outward while a wave of decreasing field will start out from MN the moment the current in it begins to decrease. The field at point P_1, after having reached a maximum, simultaneously with the current in MN, will now begin to dercrease with the current, at a time when the field at P_5 or P_6 has not yet reached its maximum and is still increasing. The wave of decreasing field will now travel outward from MN at the same speed as the wave of increasing speed, and will reach all points in succession. It may thus be seen that *at each point of space,* the field reversals follow the current reversals in the wire, at a rate or frequency equal to that of the alternating current in the wire. But due to the finite speed of propagation of electrical disturbances in ether, these reversals of the field do not take place simultaneously at all points of space. With an alternating current flowing in MN, there will then be a succession of waves alternately increasing and decreasing fields traveling from MN outward.

From this it may readily be imagined that points P_3 and P_5, for instance, might be at just such a distance apart that the field variations at P_5 would lag one complete cycle behind the variations at point P_3. This would then result in a synchronism of the field variations at the two points. The distance between two such points, located on the same line of travel of the wave, is called the *wave length* of the radiation or oscillation sent out by the circuit MN. It is evidently equal to the velocity of propagation of the electric disturbance, divided by the frequency of the oscillatory current setting up the field. Thus,

$$\text{Wave length} = \frac{\text{velocity of light}}{\text{frequency}} = \frac{300{,}000{,}000}{f} = \lambda_m \qquad (15)$$

where λ_m is the wave length in meters. It is thus seen that the higher the frequency, the shorter the wave length. The wave lengths used in radio communication range all the way from several thousand meters to 100 or even 75 meters. Wave lengths as short as a few millimeters have been obtained in laboratory work by the use of especially constructed oscillatory circuits.

Another expression of wave length may be obtained from the formula:

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

whence

$$\lambda = 2\pi V\sqrt{LC} \qquad (16)$$

where λ is the wave length, V the velocity of light, and L and C the inductance and capacitance of the oscillatory circuit. This wave length, being that of the wave emitted by a circuit oscillating at its natural frequency, is called the "natural wave length" of the circuit.

A graphical representation of the field may help in obtaining a clear conception of the phenomenon of wave propagation explained above. Fig. 58 shows the intensity and direction of the field at a certain instant along one of the lines of propagation of the wave. Two points at a distance apart equal to one wave length are shown, to illustrate the synchronism of the fields at

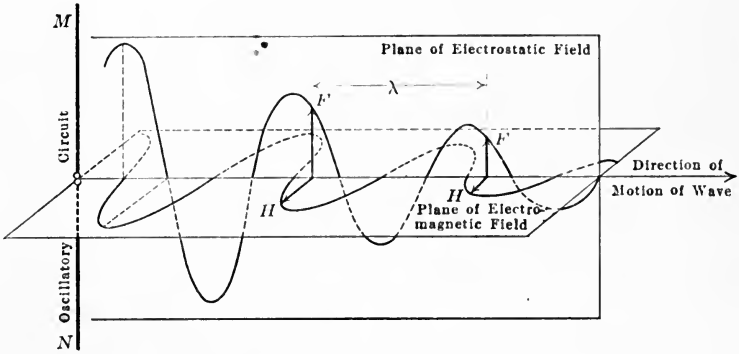


FIG. 58.

these points. As was pointed out before, when either the magnetic or electrostatic field is a maximum at a certain point, the other field at that point is a minimum at that same instant. The figure also shows that while the fields vary at all points in a similar way, the maximum amplitude of the fields decreases with increasing distance from the oscillatory circuit, as is shown in the section of this chapter on transmission formulæ.

Another way of representing the wave propagation is to represent the electrostatic field around the wire MN when this wire is carrying an oscillatory current. This is done in Fig. 59 which shows the intersections of equipotential surfaces of zero potential with the plane of the paper. These surfaces move from the wire outward at a uniform velocity, as was explained in Chapter I. They are seen to be of a cylindrical-toroidal shape, with their common axis along MN. The lines of force of the magnetic field are circles having their centers on MN and their planes perpendicular to MN.

Antenna Resistance.—It was shown in Chapter I that electrical energy was dissipated or lost in a circuit on account of its resistance. In an antenna or radiating oscillatory circuit, electrical energy is dissipated in three different ways.

First, energy is dissipated in the *ohmic resistance* of the wire making up the circuit, the loss being equal to the product of the current squared by the resistance (I^2R).

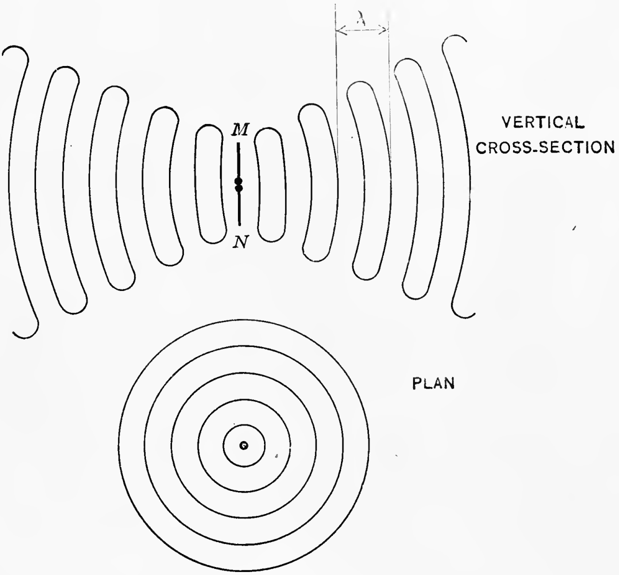


Fig. 59.

Second, the circuit loses energy by *radiation*, this energy being stored in the electric waves traveling outward into space. This radiated energy may be made to perform work by the setting up of currents in conducting materials or circuits placed in the path of the wave. The amount of energy radiated is proportional to the square of the frequency, and therefore inversely proportional to the square of the wave length. It is also proportional to the square of the current in the antenna, and is therefore of the form I^2A. By its similarity with the resistance loss, the factor A has been called the *radiation resistance* of the antenna.

Third, energy is lost to the surrounding objects. On account of the large dimensions usually given to an antenna, the surrounding dielectric is not perfect since it comprises in the vicinity of the circuit, the masts or towers and other more or less conducting materials such trees, houses, metal structures, etc. The alternating fields surrounding the circuits thus perform work, and energy is lost by the circuit and absorbed by the medium in which currents are induced by the motion of such free electrons as may be present. This dielectric absorption loss is approximately proportional to the wave length.

The *effective resistance of an antenna* is thus the sum of the three resistance components indicated above. The resistance

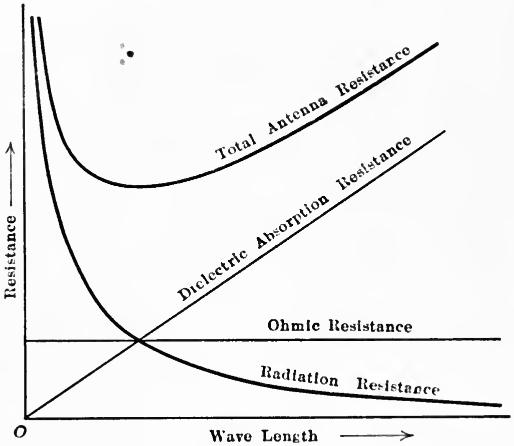


FIG. 60.

of an antenna will therefore vary with the wave length of the oscillations taking place in it. This has been represented in Fig. 60, from which it may be seen that the effective or total resistance of the antenna has a minimum value for a certain critical wave length, that being the wave length at which the sum of the several resistance components is a minmum.

Antenna Constants and Types.—In actual practice, the radiating circuit or antenna of a radio set does not have the shape shown in Fig. 52. Only one of the two conductors AA', BB', is used, the other being generally replaced, except for the connecting lead from the radio set box, by the ground on which the antenna is erected. In order to obtain a good contact with the ground, various methods which are indicated in a later paragraph are used. A simple antenna circuit then takes the form of a single wire suspended vertically and grounded at its lower end, Fig. 61. By "antenna" is henceforth meant the radiating system which comprises the aerial wire, the ground or counterpoise and the connecting leads.

Now consider the segment *AB* of the antenna, Fig. 61, to the exclusion of the rest of the antenna. This segment is at a certain height above the conducting ground and is separated from it by a layer of air which is a dielectric material. This segment and

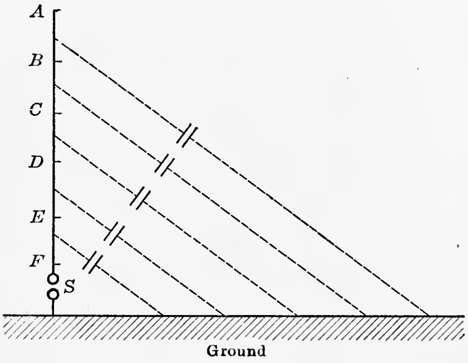


FIG. 61.

the ground therefore form a condenser having a certain amount of capacitance. Similarly, the segments *BC*, *CD*, *DE*, etc., and the ground form condensers. Since the capacitance of these condensers is inversely proportional to the thickness of the dielectric, it will be greatest for the lower sections of the antenna aerial. In other words, the capacitance of the antenna is not uniformly distributed along the wire, but is greater for the lower sections of the wire. On the other hand, since the inductance of the antenna is equal to the ratio of the magnetic flux to the current in the antenna, it is seen to be approximately the same for all sections of the wire.

If a high frequency alternating current is made to flow in the antenna by setting up an oscillation in it, the frequency of this current is equal to $\frac{1}{2\pi\sqrt{LC}}$, which is the natural frequency of the circuit, L and C being the total inductance and capacitance of the antenna. However, due to the unequal distribution of the capacitance along the wire, it is seen that the impedance of the circuit, namely,

$$Z = \sqrt{R^2 + \left(2\pi fL - \frac{1}{2\pi fC}\right)^2}$$

is smaller for the lower portions of the antenna. And since

$$I = \frac{E}{Z},$$

it follows that the total energy of the system $W = EI$, will divide itself so as to give a maximum current at the base of the antenna and a maximum voltage at the free ends. The current and voltage distribution are then approximately represented by the curves of Fig. 62. In the case of the oscillatory circuit of Fig. 52, the maximum current of course occurs at the center, while the maximum voltage is at the free extremities. This was already pointed out in connection with Fig. 54.

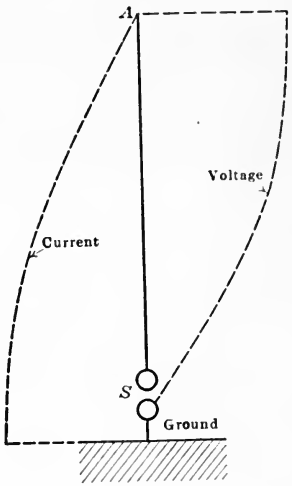


FIG. 62.

It may be shown mathematically, and it has been observed by wavemeter measurements, that a circuit of the kind illustrated by Fig. 52 will oscillate freely at a wave length equal to twice the over-all length $A'B'$ of the wire. The antenna of Fig. 61, being essentially one-half of the antenna of Fig. 52, oscillates at one-quarter wave length. That is, its natural wave length is roughly equal to four times the height of the vertical wire.

In order to change the wave length of a vertical wire antenna it is then simply necessary to vary the length of the wire. Another method is to insert an inductance coil L, Fig. 63, in series with the antenna. This increases the total inductance of the circuit, and therefore also the wave length, as may be seen by reference to the wave length formula given on page 82. In order to shorten the wave length, a condenser C, Fig. 64 may be inserted in the antenna, the total capacitance of the oscillatory circuit being thereby reduced. This is due to the fact that this condenser is in *series* with the capacitance of the antenna system.

Another method of changing the wave length of an antenna is to alter its constants by changing its shape. Thus in Fig. 65, by suspending horizontally a length AE of the originally vertical wire, a so-called *inverted "L"* antenna is obtained. This antenna will have only slightly greater inductance than the vertical antenna of Fig. 61. But all the sections of the wire above point E are

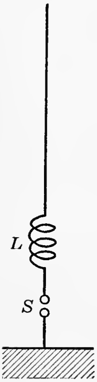


Fig. 63.

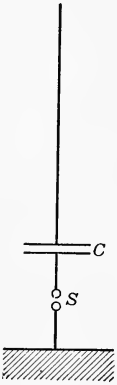


Fig. 64.

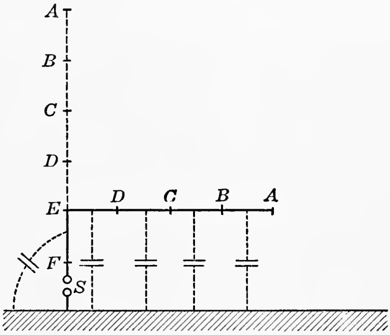


Fig. 65.

thus brought closer to the ground and the capacitance of each is therefore increased and the capacitance of the whole antenna is greater, and therefore also the wave length. Thus, while the wave length of an inverted "L" antenna increases with the length of the wire making up the antenna, the simple rule of the quarter wave length given for the vertical antenna does not hold true in this instance.

The antenna shown in Fig. 66 differs from that of Fig. 65 in that the vertical portion is connected to the middle of the horizontal portion, instead of the end. This antenna is called a *"T" antenna*. With the same dimensions as the inverted "L" antenna of Fig. 65, this "T" antenna has about the same capacitance, as may be seen from the fact that the various parts of the wire have not changed positions with regard to the ground. It has however smaller inductance, as may readily be understood from the following. During one half-cycle the oscillatory current is flowing say

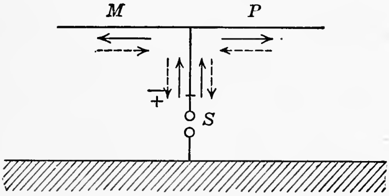


Fig. 66.

upward in the vertical portion of the "T" antenna and the current flows at the same time to the right in the horizontal portion *P* and to left in the portion *M*, these two currents being also of equal intensity. The electromagnetic fields of these currents in the horizontal sections will therefore practically neutralize each other, and the total magnetic flux will be less for the same current than in the case of the "L" antenna.

A similar scheme is used when it is desired to have an antenna of large capacitance. The constructon is illustrated in Fig. 67 and is called an *umbrella* type antenna. The total inductance

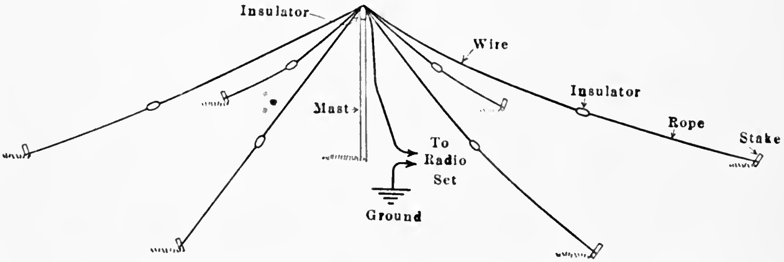


Fig. 67.

is only a little greater than that of the vertical portion of the antenna, while most of the capacitance is due to the condenser effect of the flat top and the ground counterpoise, and is quite large on account of the large area of ground covered by the antenna.

It is in general a rather difficult matter to calculate accurately the constants of an antenna from its dimensions, because of numerous variable factors affecting the constants, such as the surroundings, the presence of trees, houses, etc., the moisture and pressure of the air, etc. Formulæ permitting the calculation of antenna constants may be found in Bureau of Standards Circular No. 74.

Receiving Antennæ.—Reasons similar to those which made the use of "open" oscillatory circuits desirable for emitting or radiating energy have led to the use of open oscillators for receiving energy. In a closed oscillatory circuit, such as that of Fig. 51 above, the waves emitted by a distant oscillatory circuit *P* would induce a current generally too small to be observed because the emf. induced in a section *AB* of the circuit would be practically equal and opposite to that induced in a section *A'B'*.

With an open oscillating circuit, Fig. 55, this will not occur, and induction will take place effectively. Also, as was previously stated, on account of the great distance generally separating the transmitting and receiving circuits, the latter should be linked by as large a portion of the fields of the former as possible. This can be accomplished only by making the receiving circuit of large dimensions, and making it then form an open oscillator. In other words, the receiving circuit should be not only electrically identical with the transmitting circuit, (obtained by tuning) but should also be of similar shape and dimensions. This latter feature is, however, not essential, but may be regarded as desirable and a general guide in the installation of the circuits.

Directional Characteristics.—The energy radiating qualities of an antenna depend on the shape of the fields of the antenna, that is, on the strength of these fields in the various directions around the antenna. It is already known (Chapter I) that the shape of the circuit directly and fundamentally affects the shape of the field. This will be studied here in somewhat greater detail in the case of antenna circuits.

Consider a vertical wire antenna, shown in plan view by the point A, Fig. 68, and assume that a number of observers equipped with receiving circuits, all identical, are scattered about the

Fig. 68.

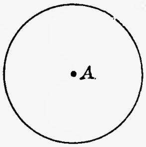
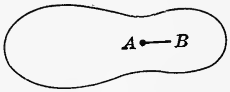


Fig. 69.



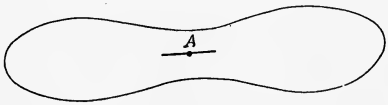


Fig. 70.

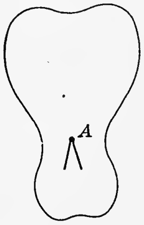


Fig. 71.

antenna. Assume also that each one of these receiving circuits has some device which permits of measuring the current induced in it when the antenna circuit A is oscillating. Now if the observers move toward or away from the antenna A until they all obtain the same current reading in their receiving circuit, they will finally find themselves on a circle having A as its center. This shows that a vertical wire antenna radiates with equal strength in all directions.

A similar test repeated with an inverted "L" antenna results in a radiation curve illustrated in Fig. 69, where *A* is the grounded end and *B* the free end of the antenna. This shows that an inverted "L" antenna radiates energy with greater intensity in the direction of its grounded end. When using an antenna of this type for transmitting, it is therefore important to orient it to point toward the receiving station, that is, with the grounded end nearest that station.

A "T"-shaped antenna, being essentially made up of two inverted "L" antennæ having a common vertical portion, has a curve similar to that shown in Fig. 70. A "V"-shaped antenna, consisting of a double "L" antenna with the horizontal portions in the form of a "V" and with a common vertical portion, has a radiation curve as shown in Fig. 71.

These directional properties are important when it is desired to communicate with a definite station of known location, or when signals are to be prevented from traveling in a certain direction.

The directional properties of receiving antennæ are the same as those of transmitting antennæ. Two stations using inverted "L" antennæ should then, for best results, be set up as in Fig. 72. As a general rule, it may be said that the maximum direc-

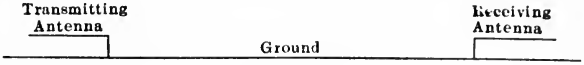


Fig. 72.

tional effect is in the plane containing the antenna aerial and lead-in wires, and in the direction of the lead-in end of the antenna. If, as in the case of the "V" antenna, this plane is not well defined, then the directional effect is in the plane containing the lead-in wire and the geometrical center of gravity of the aerial.

Factors Affecting the Transmission Range.—Consider a transmitting and receiving circuit at a certain distance apart, and suppose that the transmitting circuit is successively excited at various frequencies. Suppose also that in every instance the two circuits are tuned to the operating frequency. If the two circuits were separated by a perfect dielectric medium, that is, by a medium in which there are no free electrons or ions whatsoever, the inductive effect, and therefore the current induced in the receiving circuit, would constantly increase with the frequency of the oscillations set up in the transmitting circuit, since this induced current is directly proportional to the rate of variation of the current in the inducing or transmitting circuit.

However, the dielectric medium separating the two circuits is not perfect, due to the presence in the space separating the circuits, of trees, houses, etc., which are poor insulating materials or even good conductors and therefore contain free electrons. These are set into motion by the waves emitted by the transmitting circuit and absorb part of the energy of the waves, as already explained. Nor is the air itself a perfect dielectric. Especially in daytime, the sun's rays ionize the air and make it more or less conducting. This is explained by the theory that the luminous radiations, being of the nature of electromagnetic waves, produce molecular distortions of the air at such a rate as to break up some of the molecules and produce ions or free electrons.

The energy of the waves emitted by the transmitting station is thus partly absorbed by the media between the two stations, and this absorption increases with the frequency. This was pointed out in connection with the treatment of antenna resistance. It is then seen that, as the frequency of the transmitted waves is gradually increased, the received current will first increase, due to greater inductive effect, but will, beyond a certain critical frequency, decrease again due to excessive absorption. For this reason, long wave lengths (low frequencies) are generally used for long distance transmission, while short waves are used for short distance work. The use of the latter will be shown to be very important in connection with loop antennæ.

Radio Transmission Formulæ.—As pointed out in the previous paragraphs, the amplitude of the electromagnetic and electrostatic forces vary from point to point, decreasing as the point considered is farther away from the transmitting circuit. From the remark of Chapter I that a charge may be considered as distributed over any one of its equipotential spheres, and also from Coulomb's law, it is known that the energy of the fields per unit volume of the medium, decreases inversely as the square of the distance from the transmitting circuit. And since the energy is proportional to the square of the current or voltage, as was shown in Chapter I in the expressions for the energy of an electrostatic or electromagnetic field, it follows that the electrostatic and electromagnetic forces at each point vary inversely as the distance of that point from the transmitting circuit.

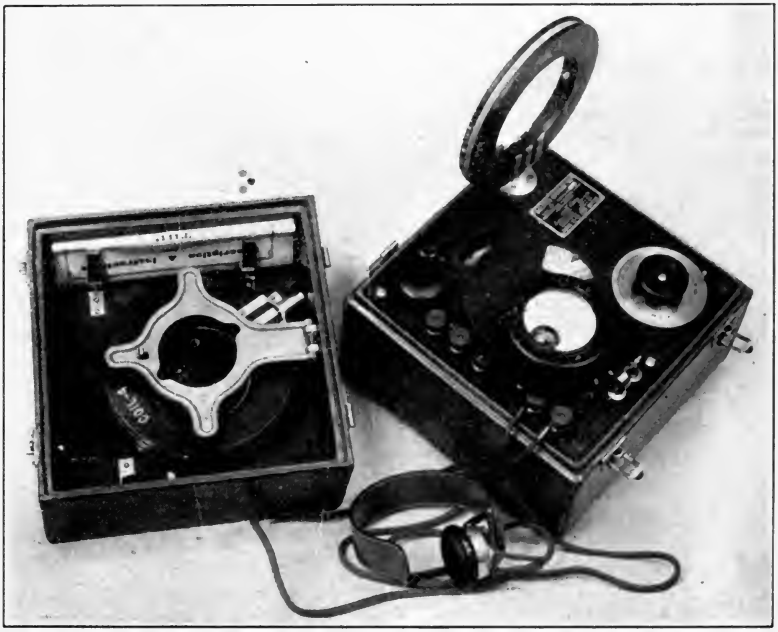


Plate 1.—Bureau of Standards Kolster decremeter and wavemeter type D. Wave length range by means of four interchangeable coils, 70 meters to 3000 meters; decrement range 0 to 0.3.

(*Facing page* 92.)

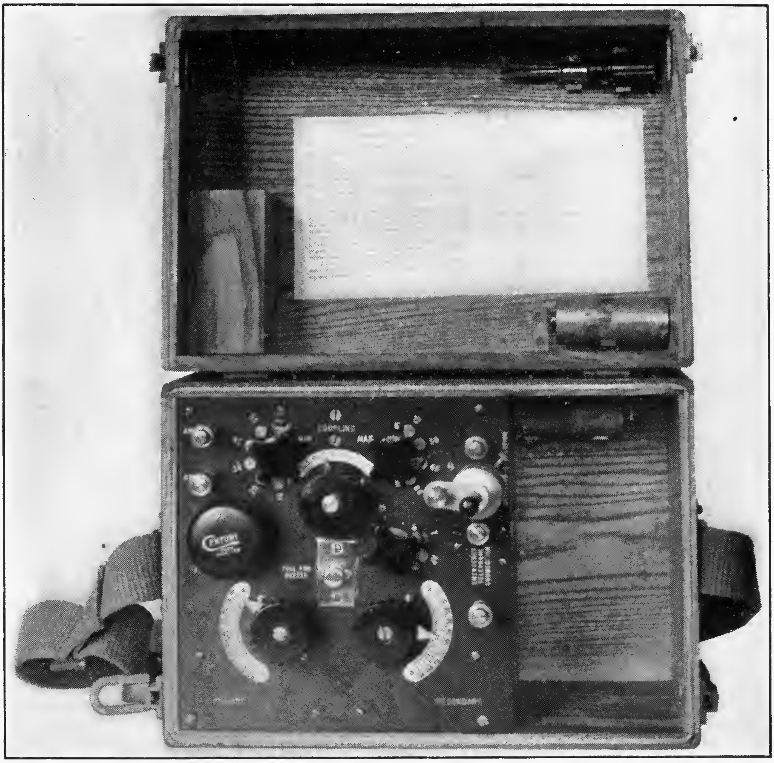


Plate 2.—Simple crystal detector receiving set having tuned primary and secondary circuits. Signal Corps set type SCR–54–A.

Actual measurements have been made of the current received in an antenna, the distance of which from the transmitting antenna was varied.[1] The theoretical law was thereby confirmed, and hence the received current may be expressed by the relation:

$$I_r = \frac{188\ h_s h_r I_s}{R_r \lambda} \times \frac{1}{d} \times \epsilon^{-\frac{\alpha d}{\sqrt{\lambda}}}$$

where I is the current, h the height of the antenna, λ the transmitted wave length, R the resistance of the receiving antenna, d the distance between the sending and receiving circuits, and the subscripts r and s referring to the receiving and sending circuits respectively. The factor α has a variable value and depends on the amount of dielectric absorption. This was shown to vary with the sun exposure and various other factors which were not considered in Chapter I and were only determined experimentally. Similar formulæ hold when using loop antennæ, reference being made to a later chapter.

Actual Construction of Antennæ.—The actual construction of an antenna comprises the choice of a suitable location, the orientation of the antenna in the proper direction, the erection of masts or supports, insulation of the wires, and the making of a suitable ground connection.

The choice of proper location is of great importance, especially in the case of small antennæ, say of a height of less than 60 ft. The presence of trees or conducting structures in the vicinity of the antenna should be avoided if possible and particularly in the direction of transmission or reception. Due regard should be given to the directional characteristics of the antenna, so that maximum range may be secured in the direction of transmission.

The masts supporting the antenna are generally of wood or of metal. In the latter case, the metal masts or towers are frequently erected on insulating bases such as marble blocks, with a view to insulating them from the ground and reducing the energy absorption resulting from currents induced in them by the antenna current.

The antenna must be well insulated from the supporting masts. In general, in a flat top ("L," "T" or "V") antenna, the hori-

[1] See Bulletin of the Bureau of Standards, Washington, 1911, Reprint 159, Vol. 7, No. 3.

M. C. Tissot, *The Electrician*, London, 1906, Vol. 56, p. 848.

W. Duddell & J. E. Taylor, *Journal of the Institution of Electrical Engineers*, London, 1905, Vol. 35, p. 321.

zontal portions are much longer than the vertical portions. The latter is, in fact, merely a *lead-in* wire, to connect the aerial or upper part of the antenna to the apparatus and ground. From what has been said above regarding the current and voltage distribution along the antenna wire, it is seen that the free end of the aerial must be insulated with the greatest care, this being the point of maximum potential. On the other hand, the bottom end of the lead-in wire, and the ground connection must be of low resistance, and are therefore preferably made of heavy copper wire. In Fig. 73 is illustrated schematically the installation of

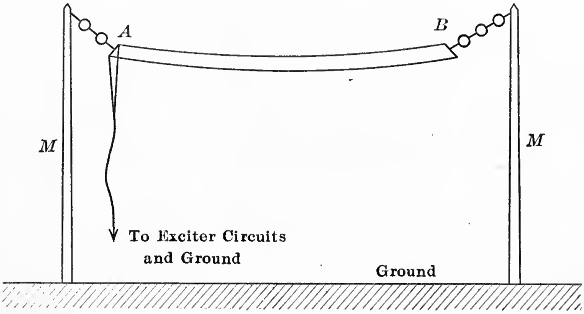


FIG. 73.

an inverted "L" antenna. The flat or upper part is made up of two wires, which decrease the ohmic resistance, and increase the capacitance, since the antenna may be considered as a large condenser, one plate of which is the aerial, and the other the ground. The free end *B* of the aerial is seen to be insulated by means of strain insulators, and for a greater potential than the lead-in end *A*.

The ground connection is of considerable importance and should be of as low resistance as possible. This is necessary since it was explained that the ground forms one-half of the oscillatory circuit, and replaces the one section *BB'* (Fig. 52) of the oscillator. It is of course possible instead of using the ground to use a complete oscillator, as shown in Fig. 74, by stretching an insulated wire or set of wires underneath the aerial. The oscillatory circuit then comprises the inductance of the lead-in wires and wires *P* and *Q*, and the capacitance of the condenser *PQ*. But unless the ground is dry, sandy or rocky, the oscillatory current in the antenna will induce considerable currents in the conducting ground, which will then absorb an appreciable part of the radiated energy.

In order to reduce or suppress this energy loss, the lower part Q, called the *counterpoise* is buried in the ground instead of being insulated from it, and the ground itself is made the counterpoise,

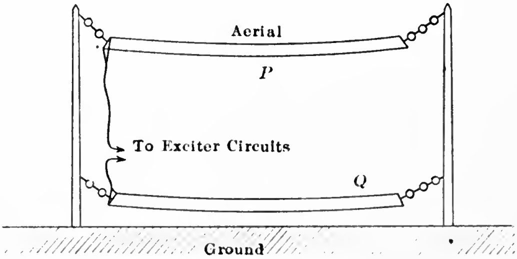


Fig. 74.

Fig. 75, so that the currents in it are not induced and wasted but are actually part of the oscillatory current of the antenna. It is of course necessary that the ground connections be well made. The ground should be of low resistance, and to this end, a large

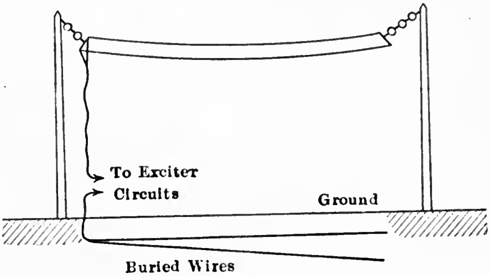


Fig. 75.

number of interconnected bare wires are buried in moistened ground beneath the aerial. So-called ground mats, made of metal gauze or netting are quite effective for this purpose.

Conduction Theory of Radio Communication

The above theory which explains the energy transfer from the transmitting circuit to the receiving circuit exclusively through wave radiation in the ether surrounding the circuits, is of a very general character, and assumes that the two circuits are isolated in space. That is, it neglects the influence of the earth, and is therefore fully applicable only in cases where this influence is negligible. Such cases are those of airplane radio communication, of the transfer of energy between closed oscillatory circuits such as wavemeters, loop circuits and the like, and radio communication between open oscillatory circuits, between which, even if grounded, the greater part of the energy transfer takes place through radiation due to the peculiar design of the circuits.

However, circuits have been developed which make use of another principle, and which appear to insure as good communication as the circuits previously used, the functioning of which was based solely on the wave radiation theory. In order to fully understand the subject of radio, it is therefore necessary to know something of the *conduction theory of radio communication.* It takes into account the electrical inter-connection of the radio transmitting and receiving circuits resulting from the fact that both are grounded.

Consider again the oscillatory circuit of Fig. 53. From the process of the discharge phenomenon, it may be understood that the oscillation consists in the surging back and forth along the wire of a charge of electricity resulting from the electrostatic unbalance originally created. That is, a mass or agglomeration of electrons will oscillate back and forth along the wire between its extremities $A'B'$. During one half-cycle, this mass will move from the gap AB to A', stop, reverse and move back to AB. In the second half-cycle, it will move from AB to B' and back to AB. During each half-cycle, the motion of this mass will have resulted in the radiation of a certain amount of energy in the form of a wave. But this radiated energy represents only a part of the total energy of the moving charge.

The same phenomenon will take place if one of the two halves of the oscillator is given a different shape. Thus, if the section BB' is connected to the ground, Fig. 76, the entire earth will form one-half of the oscillating circuit. The phenomenon is then explained as follows: Consider the oscillating charge at the instant it is at the free end A' of the aerial and stops to reverse its direction and start on its downward surge. At that instant, referring back to the explanation of the oscillatory discharge of an open oscillator, the entire energy of the system is stored in the electrostatic field. As the charge is moving from A' downward, this electrostatic energy is transformed into electromagnetic energy, so the velocity of the charge along the wire increases, until a maximum velocity is reached when the energy transformation is complete. At that instant, the charge is passing the electrostatic center of gravity of the system, which is very nearly at the bottom end B of the aerial. At that moment, the $A'A$-BB' condenser is discharged, and the wire AA' is at the same potential as the ground. But, as explained previously, the current in the circuit continues in the same direction due to the store of electromagnetic energy. That is, a current will flow from the base B of the aerial into the ground, and the electrons making up the oscillating charge will spread over the surface of the ground. They will thus travel in all directions away from the antenna, with decreasing speed as the electromagnetic energy of the system decreases and becomes transformed into electrostatic energy again. When this transformation is complete, the electrons stop, and, due to the electrostatic field resulting from the unbalanced charge of the system, they will flow back into the aerial, the process repeating itself until all the energy has been radiated or lost as heat.

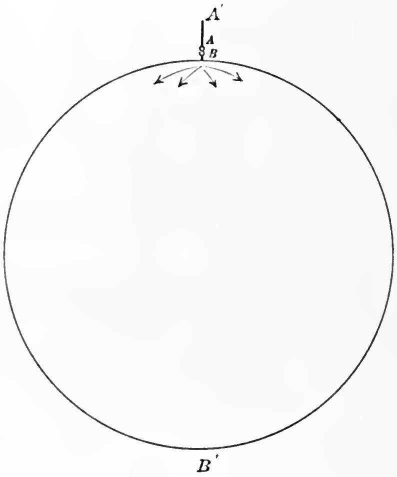


Fig. 76.

Now, in each half-cycle during which the charge oscillates from B to A' and back to B, the main energy loss is due to radiation in the form of a wave emitted perpendicularly to the direction of motion of the charge, that is, horizontally. In the half-cycles during which the charge travels along the surface of the earth, each electron describes a motion back and forth along a meridian line drawn from the foot B of the aerial. A certain amount of energy is thus radiated normally to the earth's surface, that is, vertically. But the electrons of the charge also act upon the free electrons normally scattered over the surface of the earth. The sudden variation of the charge of the globe at point B produces a variation in the electrostatic field which is propagated with the velocity of light. And thus, at all points of the earth's surface, electrons will move to and fro, along meridian lines drawn from point B, and at the same frequency as the antenna oscillations. There is thus an oscillatory current set up at each point of the earth's surface. From this explanation it may be seen that there will be standing waves along the earth's surface. In other words, the current in the huge oscillator made up of the globe and the aerial $A'B$, will oscillate at a harmonic frequency, and not at the fundamental or natural frequency. The oscillation frequency is determined, as explained above, by the constants of the aerial BA'.

Before considering the methods whereby these earth currents can be used and received, it may be well to study briefly the law of propagation of the energy thus sent out from the antenna $A'B$.

First, consider the energy radiated by the aerial in the form of waves. The theoretical law of decrease of energy is the inverse square of the distance law mentioned above. The theoretical field is shown in Fig. 59. In the case considered here, the aerial no longer sets up the waves in empty space and from an isolated position, but in the earth's atmosphere. This consists of a layer of air surrounding the globe and enclosed on one side by the earth's conducting surface and on the other side, by rarefied gases, which, being strongly ionized by the sun's radiation are more and more conducting the higher they are and the nearer to the upper layer of the atmosphere. The waves set up by the grounded

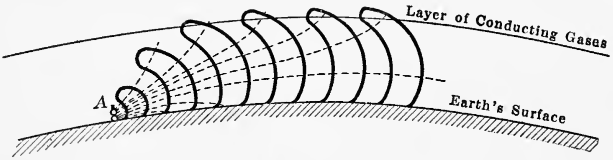


Fig. 77.

oscillator A, Fig. 77, thus propagate in a layer of insulating air enclosed by two conducting layers—the rarefied ionized air and the earth. The field oscillations will then set up alternating electric currents in these two layers, expending at least a part of the energy of the radiated waves. As a consequence, the equipotential surfaces representing the field will no longer have the shape shown in Fig. 59, but will be bent backward, as if trailing behind the parts of the surfaces moving in the insulating layer of air, and as shown by the solid lines of Fig. 77. And since a wave disturbance travels always in a direction normal to its wave front, it may be understood how the energy per unit volume of the insulating medium decreases rapidly as the wave travels away from the oscillator. This is illustrated by the lines of force, shown as dotted lines in Fig. 77, and which are seen to spread very rapidly. This absorption effect is expressed by the exponential factor in the transmission formula given on page 93. The variation of the total electric energy per unit volume of the medium for points at various distances from the transmitting circuit may thus be approximately represented by the curve A, Fig. 78.

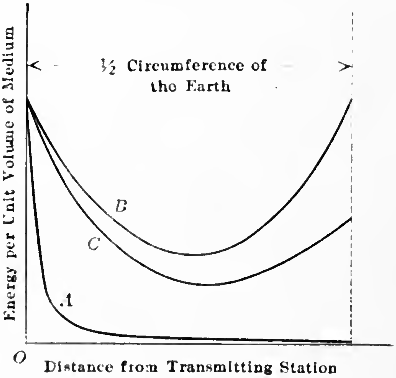


FIG. 78.

Now consider the variation of the earth currents with the distance. For simplicity, assume that the frequency of the oscillations is such that there will be no intermediate nodes of current along the earth's surface, or else consider only points where the current is in synchronism. Neglecting the irregularities of the earth's conductivity at various points, it is evident that all points at an equal distance from the transmitting circuit will be under identical electrical conditions. Thus, A being the transmitting oscillator, Fig. 79, the current at all points of circle B will be the same. Also, under the above conditions, the *total* current passing through the points of the circumference B is equal to the total current passing through any circumference parallel to it. The current or energy per unit volume of the earth's surface is then equal to the total antenna current or energy divided by the circumference at that point, the circumference being in a plane perpendicular to the direction of propagation of the current. If d is the distance of the point considered from the transmitting antenna, R the radius of the

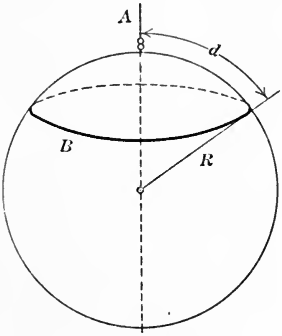


FIG. 79.

earth and I the total current, then the current at any point of the globe is equal to

$$i = \frac{I}{2\pi R \sin \frac{d}{R}}$$

as may be found by a simple computation. This is represented by curve B, Fig. 78. This curve does not take into account, however, the radiation losses of the oscillation along the earth's surface, which were pointed out above. The actual curve would, therefore, be somewhat like that represented by curve C, Fig. 78.

It may seem somewhat startling to indicate that, for points at a distance greater than one-fourth the circumference of the earth, the signal strength increases with the distance. The truth of this has not been determined, in the knowledge of the authors, principally for the reason that the circuits which have been used are mostly sensitive to the radiated waves. With circuits of suitable design, however, it has been possible to receive signals from distant stations, which could not be received with apparatus using a circuit responsive mainly to space radiation. Thus the conclusion reached is seemingly substantiated.

This leads to the method of receiving signals which makes use of the earth currents rather than space radiations. The reception of signals by means of open oscillatory circuits of the type described previously in this book can be explained by the theory of earth currents. In fact, it is believed by some authorities, and it appears probable from the curves of Fig. 78, that these currents are largely responsible for long distance communication. Thus, consider a straight vertical wire antenna, as in Fig. 63 or 64. The oscillatory current taking place in the ground in a horizontal direction, produces at the foot of the grounded receiving antenna, a to and fro motion of an electric charge, and there establishes between the bottom and upper extremities of the receiving antenna an alternating or pulsating emf., under the influence of which an alternating current will flow in the receiving circuit. This current will be a maximum if the antenna circuit is tuned to the frequency of the earth currents, that is, to the frequency of the transmitting

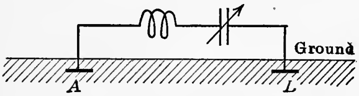


Fig. 80.

antenna. The action, as may be seen, is mostly one of electrostatic induction.

More effective circuits are those making use of electrostatic and electromagnetic induction and direct conduction phenomena. Thus, Fig. 80, a tuned circuit grounded at both ends, will shunt the current which would normally flow in the ground between points *A* and *B*. The current which is thus made to flow in the circuit will merely supply the resistance losses, while resonance currents of considerable strength will flow in the circuit. Similarly, Fig. 81, the circuit may be entirely insulated,

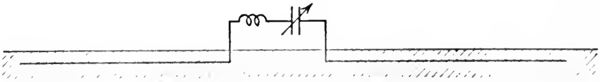


Fig. 81.

the action being then an inductive one. It is simply necessary to bury a length of wire a few feet under the earth's surface. Finally, if the circuit of Fig. 81 is grounded at the ends, induction and conduction effects will take place. It is evident that such receiving circuits will operate best if they are in the direction of propagation of the current, that is, pointing toward the transmitting circuit. Mr. J. H. Rogers in coöperation with the Navy Department has thus obtained satisfactory results over distances of several thousand miles, using buried receiving circuits a few hundred feet long.[1]

[1] The above conduction theory of radio is given as a matter of interest. It has been advanced by several scientists, but being difficult to isolate the respective effects of pure wave radiation and conduction experimentally, the theory has remained open to question.

CHAPTER IV

DAMPED WAVE RADIO TELEGRAPHY

Recalling previous discussion, it will be remembered that radio telegraphy may be accomplished by setting up high frequency alternating currents in a radiating circuit such as an open oscillator or antenna, to induce currents in a similar tuned circuit at a distant receiving station. Systems of radio telegraphy making use of the oscillatory discharge of a condenser for the production of high frequency currents have been called *damped wave* radio telegraph systems, for the reason that the oscillations are not true alternating currents, but are damped. The use of a spark gap of some sort in the discharge circuit has aslo led to the common use of the name "spark set" for a damped wave radio transmitting set.

Damped Wave Radio Transmitting Circuits

From what was said in Chapters II and III, it may be understood how the circuit of Fig. 82 constitutes a damped wave transmitting set. With the key K closed and the spark gap AB suitably adjusted, an oscillatory discharge will take place in the circuit $A'B'$ every half-cycle of the alternator, as explained in the case of Fig. 22. Thus, a succession of groups of damped oscillations will take place in the open oscillator, and a series of damped wave trains will be radiated into space, as long as the key is closed. The frequency of the oscillations will be equal to the natural frequency of the circuit $A'B'$, while the frequency of the groups of damped waves, that is, the number of wave trains per second, will be equal to twice the alternator frequency, there being one discharge per half-cycle.

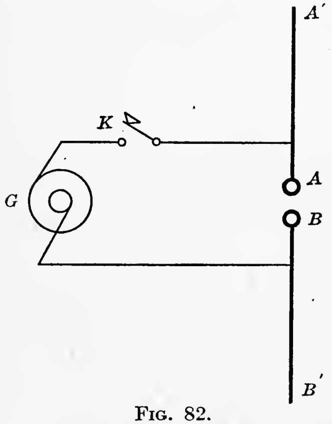


Fig. 82.

As was explained in the previous chapter, the actual radiating circuits used in practice take the form of an *antenna*, so that the circuit would be as shown in Fig. 83. Also, on account of the rather small capacitance of such an antenna, it is clear that it is necessary to charge it before every oscillation to as high a potential as possible, in order to have a large store of electrostatic energy to produce the oscillatory discharge. If this is not done, the oscillatory current will be of small amplitude, with a consequent reduction of the range of transmission. For this reason, the aerial and ground are not connected directly to the alternator *G*. A transformer *T* is used between the antenna and the alternator to step-up the voltage, and it is thus possible to obtain several thousand volts difference between the aerial and the ground.

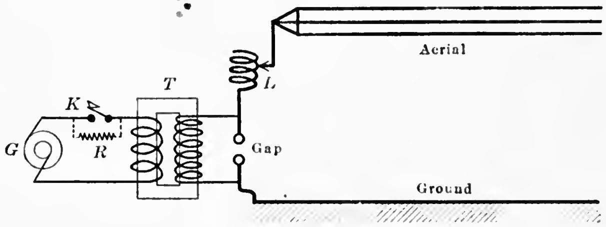


Fig. 83.

The key is usually inserted in the low voltage circuit, in series with the transformer primary winding, as shown in Fig. 83. When the key is closed, an alternating emf. is impressed on the oscillatory circuit, through the medium of the transformer, and wave trains are sent out, one every half-cycle. By closing the key for a shorter or a longer period of time, a short or a long series of wave trains is sent out, corresponding to the dots and dashes of the Morse alphabet code.

A disadvantage of using the key in this manner is that the alternator is thus made to work at either no load when the key is open or at full load when it is closed, this sudden and large change in load resulting in variations of alternator speed and poor operation. The key may therefore be shunted by a resistance *R*, which is great enough to reduce the current through the transformer primary and hence the secondary voltage, when the key is open, to a value just below the break-down voltage of the gap. The variation of load on the alternator is then much less and the radio signals sent are more readily legible.

Another method is to place the key in series with the field winding of the alternator. The alternator then is excited only while the key is closed. This method however has the disadvantage that the inductance of the field winding prevents a rapid building up of the alternator field and thus necessitates a slow sending speed.

As was pointed out before, if it is desired to send out a wave length different from the natural wave length of the antenna, a coil L may be inserted in series with it, or else a condenser, and so adjusted as to give the desired wave length.

The above method of sending out signals, whereby the antenna circuit is *directly excited,* has certain disadvantages. The presence of the spark gap in the antenna circuit introduces in the latter a high resistance which causes a rapid damping out of the oscillations. In other words, the oscillations have a high decrement. This results in a flat resonance curve, as was pointed out in Chapter II, which prevents sharp tuning at the receiving station. This consideration will be studied later.

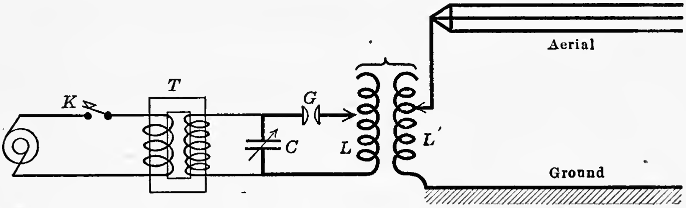


FIG. 84.

In order to prevent such a high decrement of the oscillations, and to avoid a large amount of the energy supplied by the alternator from being wasted in the spark resistance instead of radiated, the so-called *indirect excitation* method is employed. This is illustrated in Fig. 84 and reference is here made to what has been said in Chapter II on impulse excitation of coupled circuits. The method consists in coupling the antenna circuit to an intermediate oscillatory circuit containing the spark gap. The operation of this system is then as follows. Oscillations are excited in the closed oscillatory circuit CLG, when the key is closed, in the same manner as described above for direct excitation. These oscillations are of the natural frequency of the circuit. To this circuit is coupled the antenna circuit, in which oscillations are then set up by induction. This process was fully treated in the section of Chapter II relating to tuned coupled circuits. As was then shown, oscillations of maximum amplitude will be obtained in the antenna circuit, if the latter is tuned to the frequency of the closed oscillatory circuit and coupled to it with the critical degree of coupling giving maximum power. Furthermore, in order to prevent the energy transferred from the closed oscillatory circuit to the antenna circuit from reacting back upon the closed circuit, it is necessary to use a quenched gap at *G*. Resonance of the two oscillatory circuits is obtained by equipping them with variable condensers and coils permitting a variation of their natural frequencies. An adjustable coupling may be obtained by varying the relative positions of the coupling coils, which may be done by rotating one coil so as to change the relative direction of its longitudinal axis, or else by sliding one coil into the other more or less. The operation of the key is the same as in the case with direct excitation.

The operation of the circuit of Fig. 84 may be illustrated by the curves of Fig. 85. When the key is closed, the alternator emf. is impressed upon the transformer *T*, stepped up, and applied across the condenser *C* and gap *G*. This emf. is shown by the upper curve, Fig. 85. In every half-cycle, as this impressed voltage reaches the value *a*, the gap *G* which is suitably adjusted, breaks down, and an oscillatory current takes place in the circuit *LGC*, Fig. 84. This oscillation is, however, rapidly damped, as shown by the second curve of Fig. 85, principally due to the fact that its energy is being transferred to the antenna circuit coupled to it. The induced oscillatory current in the antenna is represented by the lower curve of the figure.

As in the case of coupled "closed" circuits, this induced oscillatory current grows as the energy is being transferred from the closed circuit to the antenna circuit. When this transfer is complete, the antenna oscillates at maximum amplitude, but does not react on the closed oscillatory circuit since this is opened by the quenching of the spark across gap *G*. The oscillations in the antenna, thus cut off from the influence of the closed oscillatory circuit, are permitted to decrease gradually, the decrease being due principally to the fact that the antenna circuit loses energy by radiation. And since the resistance of the antenna circuit may be made small, there being no spark gap in it, it is seen that the decrement of the wave is consequently small and the tuning of the receiving circuit correspondingly sharp.

Another advantage of the use of an intermediate oscillatory circuit between the antenna circuit and the power supply is that the amount of energy stored in the system before every discharge is $\frac{1}{2} CV^2$. For a given supply voltage, that is, for a given voltage at the transformer secondary terminals, the energy of the oscillation is then directly proportional to the capacitance of the oscillatory circuit. The capacitance of the antenna circuit

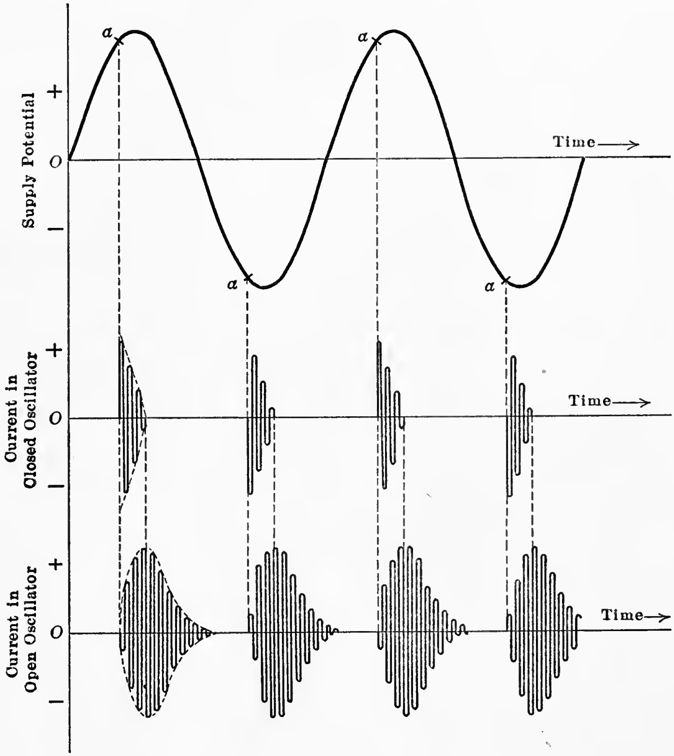


FIG. 85.

being small, the energy stored in it is small as compared with that stored in the closed oscillatory circuit which comprises a condenser of any desired capacitance. If then the directly excited circuit of Fig. 83 is used, a very high supply potential will be required to radiate the same amount of power as would be radiated by the indirectly excited circuit of Fig. 84, using the same antenna and a moderately high potential.

It may be of interest to note that in the circuit of Fig. 84, the condenser C is directly connected across the transformer secondary terminals. The operation of the radio circuit would evidently be the same if the spark gap and the condenser were interchanged, as shown in Fig. 86. From the standpoint of good operation, however, this is not as satisfactory, for as the gap breaks down, the arrangement of Fig. 86 produces a short circuit across the transformer which may either damage the transformer winding or give rise to a "power arc" across the gap, burning up the

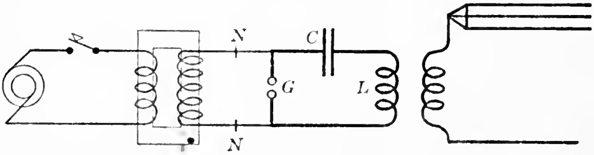


Fig. 86.

electrodes. If this circuit is used, it is generally advisable to insert choke coils at the points N. This will prevent high frequency surges from reaching the transformer winding. These choke coils are not required in the case of Fig. 84, since any high frequency currents taking place in the oscillatory circuit will pass through the condenser instead of through the transformer winding, on account of the lower reactance of the former.

Another property of the circuit of Fig. 84 is that the condenser C forms, with the inductance of the transformer, an oscillatory

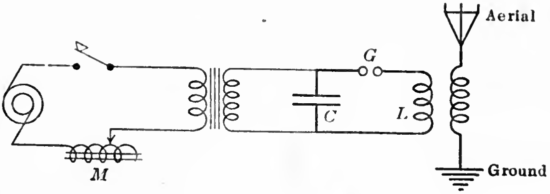


Fig. 87.

circuit connected across the alternator terminals. And if this circuit is tuned to the alternator frequency, conditions will be obtained equivalent to those of series resonance, as studied in conjunction with Fig. 25. This tuning of the power supply circuits can obviously not be done by means of the condenser C, since a change in its capacitance would at the same time alter the constants of the radio circuit LGC. For this reason, a coil M

of adjustable inductance is connected in series with the transformer primary and so adjusted as to bring about resonance of the power circuits to the frequency of the alternator, Fig. 87. As a result of this resonance, a greater voltage will be impressed across the condenser C, and the store of electrostatic energy, $\frac{1}{2} CV^2$, used in producing the oscillatory discharge in the radio

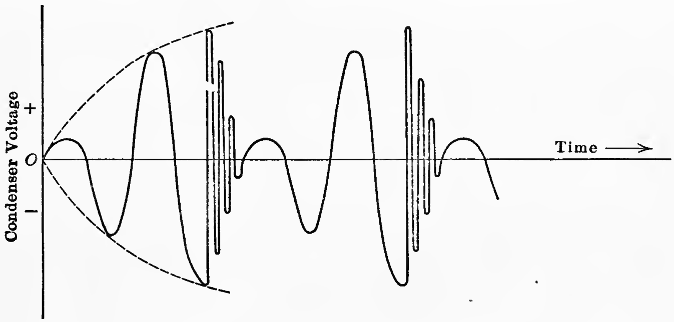


Fig. 88.

circuit will be consequently increased. It should be noted, however, that the number of oscillatory discharges or number of wave trains per second is thereby decreased and is no longer equal to twice the alternator frequency. The high potential resulting from the power circuit resonance and to which the spark gap *G* must be adjusted, is not reached immediately on account of a transient oscillation which exists in the circuit after each discharge. This transient term, which was merely mentioned in the mathematical treatment of resonance, results in operating conditions best expressed by the diagram of Fig. 88.

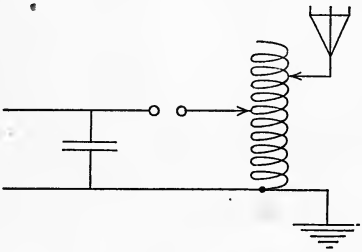


Fig. 89.

Because of this decrease in the number of sparks per second, this method of tuning the power circuit is little used in practice, and the most common circuit is that of Fig. 84. Of course, the oscillation transformer coupling the antenna and closed oscillator circuits may be an auto-transformer, as shown in Fig. 89, or any other coupling device.

Typical Method of Operation.—An idea of the method of operating a damped wave transmitting set may be obtained from the following paragraphs.

A wavemeter is first set to the wave length it is desired to use. The antenna circuit is opened, or, if this is not possible, it is coupled as loosely as possible. The closed oscillatory circuit is then excited by running the alternator and keeping the key closed. This circuit is tuned by adjusting its condenser and inductance coil until a maximum response is obtained in the wavemeter, with the latter fairly loosely coupled to it—placed say at a distance of two or three yards. The closed oscillatory circuit is then oscillating at the desired wave length.

The coupling between the antenna circuit and the closed oscillatory circuit is then given some medium value, and the antenna circuit is tuned to resonance with the closed oscillatory circuit. This condition is reached when an ammeter or other current indicating device connected in series in the antenna circuit gives a maximum indication. The coupling should not be so tight or close as to give a double peaked resonance curve, but should give a good, sharp resonance indication at a single well defined frequency.

The spark gap of the closed oscillatory circuit is then adjusted to give the maximum current in the antenna. It is evident that the gap adjustment, by its affect on the break-down voltage, determines the amount of energy stored in the condenser by the alternator during the charging period. With a set properly adjusted, the ratio of the amount of power radiated by the antenna to the power generated by the alternator, expressed in per cent of efficiency, ranges from about 3 to 15 per cent.

Damped Wave Radio Receiving Circuits

The reception of signals sent out by one of the methods just described consists essentially in having the varying fields due to the oscillations of the transmitting circuit induce currents in a suitable receiving circuit, and then observing these currents by means of some suitable device. In practice, not only induction phenomena are made use of, but also resonance phenomena, in order to increase the effects of the oscillations intercepted by the receiving circuit. That is, the receiving circuit is tuned to the transmitting circuit. This is quite essential on account of the extremely small percentage of the total energy radiated by the transmitting circuit which reaches the receiving circuit. This may be appreciated by considering a transmitting station T and a receiving station R separated by a distance d. The amount of power radiated by T is then, at a distance d, scattered over the entire surface of a sphere of radius d and center T. The receiving station R, although perhaps having a fairly large aerial, will cover only a very small fraction of that sphere, and the amount of energy received is consequently extremely small. Of course, the use of directive antennæ for sending, changes this energy distribution about T somewhat.

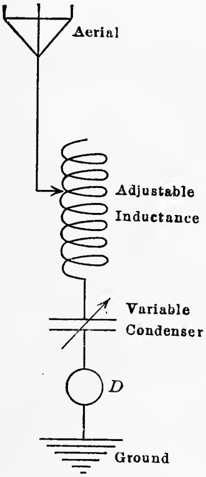


FIG. 90.

As was explained before, the receiving circuit should preferably be of large dimensions, in order to link with as great a part of the field of the transmitting circuit as possible. A simple receiving circuit consists then, Fig. 90, of an antenna circuit tuned to the frequency of the transmitted high frequency oscillations by means of an adjustable inductance and capacitance, and comprising some sensitive device D, which will respond to or evidence the emf. or current induced in the receiving circuit.

The action of the transmitting circuit on the receiving circuit is exactly the same as that studied in the case of two tuned circuits loosely coupled (see Chapter II). The main problem here becomes that of obtaining a device D which will respond to the extremely feeble currents flowing in the receiving antenna. An ammeter is of course out of the question. It would not respond to such small currents, and besides would have too slow an action to accurately follow the dots and dashes sent out at the normal speed of say 20 or 30 words a minute. A device which may perhaps be considered, of the nature of an ammeter, is the "bolometer." This device is however more frequently used for measurements of received current than for actual reception of signals and ordinary communication work. It is therefore not described here, reference being made to J. A. Fleming's book on "Electric Wave Telegraphy and Telephony."

The almost universally adopted method now is to receive the signals by sound, using the ordinary telephone receiver as the device sensitive to the small received currents. The advantage of this device is its combined sensitivity and ruggedness. Consider then a telephone receiver connected at *D*, Fig. 90. Each oscillation train sent out by the transmitting circuit, upon reaching the tuned receiving circuit will induce currents in the latter of the same frequency and general characteristics as those in the transmitting circuit. This frequency, which is the natural frequency of the radio transmitting circuits, was shown to be generally high, say at least 30,000 or 40,000 cycles per second. The telephone receiver will not respond mechanically to these frequencies. That is, its diaphragm has a period of its own, and a certain inertia which prevent it from vibrating at such a high rate. Besides, even if it would vibrate at that frequency, the human ear would not be affected, since the highest audible frequencies are between 16,000 and 20,000 cycles per second. These physical limitations of the receiver diaphragm and the human ear have made it necessary to use some corrective apparatus in conjunction with the telephone receiver to suitably modify the currents induced in the receiving circuit. This device is called a *detector* and its action is explained below.

Referring back to Fig. 85, the lower curve shows the oscillatory current produced in the transmitting antenna when the key of the transmitting circuit is closed. There is one group of high frequency oscillations every half-cycle of the alternator. If the alternator has a frequency of 500 cycles per second, there are thus 1000 groups of oscillations per second. Each one of these sets up an oscillatory electromagnetic and electrostatic field around the transmitting antenna, which is propagated throughout space with the speed of light, in the form of waves or trains of waves, as was previously explained. In traveling, these waves will pass over and around the receiving antenna, setting up similar oscillatory fields and inducing emf's. in the receiving circuits, as represented by the upper curve of Fig. 91. The function of the detector placed in series with the telephone receiver *D*, Fig. 90, is then to *rectify* the current flowing in the antenna under the effect of this induced emf. That is, the detector is a device having unidirectional conductivity of the circuit, so that current may flow in one direction and not (or only feebly) in the other. This will give a unidirectional current in the receiving antenna instead of a symmetrical alternating current oscillation, as shown in the second curve of Fig. 91. The effect on the telephone receiver diaphragm will then be cumulative for each wave train, as may be understood from the following.

The first impulse *a*, Fig. 91, by passing through the telephone winding will produce an attraction of the diaphragm. On account of the high frequency of the oscillation, the second impulse *b* will occur before the diaphragm has had time to spring back in place, and will therefore deflect it further. So will the

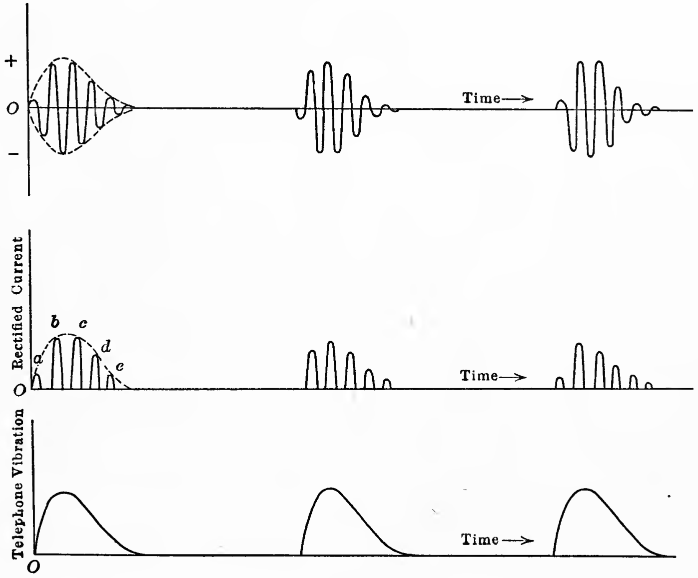


Fig. 91.

following impulses *c*, *d*, *e*, etc., the telephone diaphragm thus being deflected but once for every wave train, as shown in the lower curve of Fig. 91. It will therefore vibrate at the frequency of the wave trains, that is, at twice the frequency of the transmitting alternator. This frequency, which in the case considered is 1000 per second, is well within the range of audibility, and will produce a sound in the telephone. The wave train frequency, that is, the number of wave trains per second is for this reason called the *audio frequency* of the transmitting set, while the frequency of the oscillations within each wave train is called the *radio frequency*.

A detector frequently used and operating as explained above is the so-called *crystal detector*. It consists essentially of a crystal

A, Fig. 92, of iron pyrite, galena, molybdenum, bornite, or carborundum, in contact with a fine wire *W*. It the wire is in contact with a suitable spot of the crystal surface, the system will have unindirectional conductivity. That is, it will offer low resistance to the current flowing in one direction, say from the wire to crystal, but high resistance to the current flowing in the opposite direction. This may be represented graphically by the characteristic curve of such a crystal, Fig. 93, where the emf., positive or negative, applied to the detector, is plotted horizontally, and the resultant current through the detector is plotted vertically.

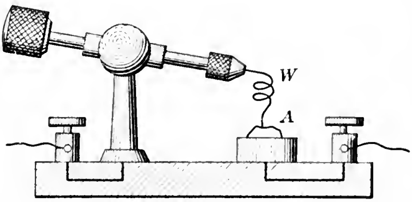


Fig. 92.

If an oscillation produces approximately equal and opposite variations of emf. across the detector, as an incoming wave does, it is seen from the asymmetrical shape of the curve that the current flowing during the negative half of the cycle is insignificant and negligible as compared with that flowing during the positive half-cycle. Rectification of the current so far as its audible effect on the receiver is concerned is thus achieved.

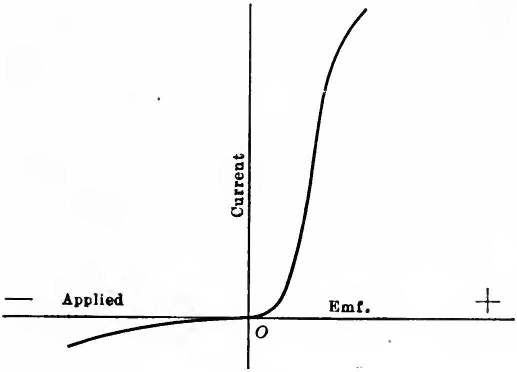


Fig. 93.

There are many other forms of rectifying detectors, such as the electrolytic detector and the vacuum tube. The former is merely mentioned here while the latter is studied in detail in Chapter VII.

Another form of detector which differs from the preceding in that it does not rectify the current, is the magnetic detector. Its operation is based on the principle that the magnetism of a piece of iron is destroyed when the latter is placed in a high frequency alternating magnetic field. The apparatus in its most common form is illustrated in Fig. 94. It consists of an iron wire supported on two pulleys P, which turn at a uniform speed. The wire passes in front of two permanent magnets placed in opposition, so that a section of the wire will be first magnetized in one direction and then in the other, going through a complete hysteretic cycle when passing in front of the two magnets. Around the wire are wound two coils, one of which is connected in series with the antenna circuit, the other with a telephone receiver. When there are no incoming oscillations, the hysteretic cycle of the iron is constant. When an oscillation train suddenly takes place in the antenna, the magnetic hysteresis in this iron is suddenly destroyed, and the resulting decrease in the flux causes a sound in the telephone receiver by inducing a current in the telephone circuit. Thus, for every wave train, there is a temporary decrease in magnetism, lasting as long as the wave train. The flux in the telephone coil thus varies periodically as the wave train or audio frequency, and a sound is produced in the telephone.

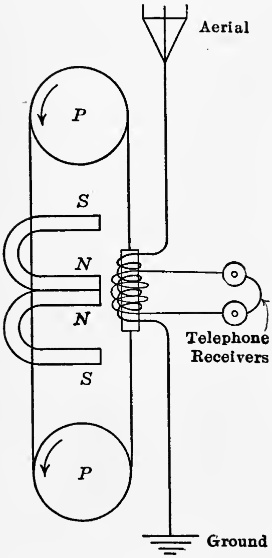


Fig. 94.

Referring to what has been said previously, it is of advantage to keep the resistance of the receiving antenna circuit a minimum. This assists in making the resonance phenomena in that circuit as effective as possible, and permits sharp tuning. The circuit of Fig. 90 is therefore seldom used in practice, since the introduction of the detector in the antenna circuit adds a comparatively large resistance which correspondingly flattens the resonance curve of the circuit. In general, the detector is then placed in a secondary tuned circuit coupled in some manner to the tuned antenna circuit, the latter then responding to the maximum degree to the incoming oscillations. Such a coupled receiving circuit is shown in Fig. 95. The antenna circuit may be tuned by means of the variable coil L_1 and condenser C_1, while the coupled or secondary circuit may be tuned to resonance with the primary or antenna circuit by means of the coil L_2 and condenser C_2. This secondary circuit is shunted by the detector D in series with the telephone receiver T. The coupling may of course be made as in Fig. 96 by means of a common coil. The primary and secondary circuits considered independently are seen to be the same as in Fig. 95. A feature of the circuit of Fig. 95, not possessed by that of Fig. 96, is the possibility of changing the relative positions of the coils L_1 and L_2, thus altering the coupling without changing the actual inductance of either circuit. Also, various combinations of capacitive and inductive couplings are of course possible, but are not shown here.

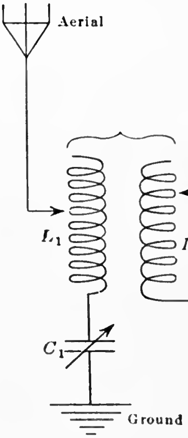


FIG. 95.

The sound in the telephone receivers is frequently improved by shunting them with a condenser S, Fig. 96, of suitable capacitance. The operation is then as follows: Consider one half-cycle b, Fig. 91, of the high frequency rectified current. During the period of its duration, the current flows through the detector D, Fig. 96, and telephone receiver windings T, and charges the condenser S. During the next half-cycle of the oscillation, no current flows through the detector D. The condenser S, how-

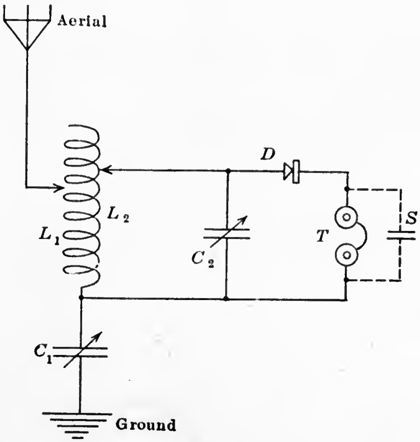


FIG. 96.

ever, being short-circuited by the telephone receivers, will discharge through the latter, and the resulting current in the receivers will be in the same direction as the current of the impulse *b*. The current flowing in the telephone receiver during one wave train will then no longer be as represented by the middle curve of Fig. 91, but will be like that shown in Fig. 97. There

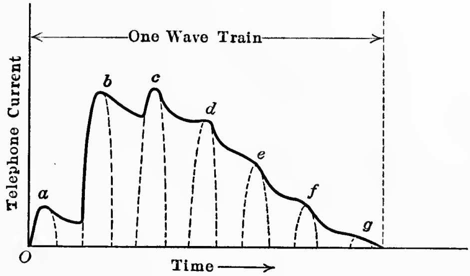


FIG. 97.

is thus, during each wave train, a more continuous attraction on the telephone receiver diaphragm, with resulting improved audibility.

The importance of using a coupled tuned circuit in receiving signals, rather than placing the detector and telephone directly in the antenna circuit will be more fully appreciated after reading the paragraphs on interference given at the end of this chapter.

FACTORS AFFECTING THE RECEPTION OF SIGNALS

From the above discussion of damped wave radio communication, it is seen that, disregarding the fact that open oscillatory circuits are used, the phenomena are the same as in the case of

FIG. 98.

two tuned coupled circuits studied in Chapter II. For simplicity in explaining the theory, the following discussion will therefore take up the factors affecting the reception of signals in a circuit *B*, Fig. 98, when excited by a circuit *A*, in which oscillatory discharges are set up periodically. The natural frequency of circuit A, which represents the transmitting circuit will be assumed to be fixed. By varying the condenser C so as to alter the natural frequency of circuit B, the familiar resonance curve is obtained, as shown in Fig. 99. The current in circuit B is given as a function of the natural frequency of that circuit.

Assuming the circuits to be loosely coupled, the curve will be a maximum when the two circuits have the same natural frequency. On the other hand, it is necessary that a certain minimum current I_t flow through the telephone receivers of circuit B in order to produce a sound loud enough to be heard. This current is determined by the sensitivity of the telephone receivers. If the natural frequency of circuit B is varied, beginning at zero, a sound will be heard in the telephone receivers between two values F_1 and F_2 of the frequency, determined as in Fig. 99, by the intersection of the resonance curve with the line of abscissa I_t, the sound being a maximum at the resonance frequency F. The range F_1F_2 of frequencies between which the receivers will produce a sound, depends evidently on the sensitivity of the receivers. Thus, with a telephone receiver less sensitive, the minimum current I_t will be greater, and will therefore correspond on the resonance curve to two points nearer the maximum, and therefore closer together.

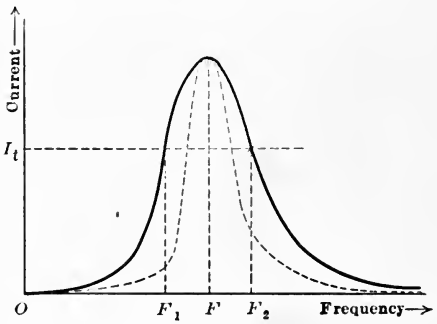


FIG. 99.

The range F_1F_2 also depends on the shape of the resonance curve. Thus, for a given telephone receiver, that is, for a given value of I_t, the range F_1F_2 will be smaller the sharper the tuning, as shown by the dotted curve of Fig. 99. From what was said in Chapter II, it is then seen that for a given value of I_t, the range F_1F_2 will be smaller the smaller the decrement of the oscillations generated by circuit A, the smaller the decrement of circuit B, and the looser the coupling of the two circuits. This last factor is equivalent, in the case of two radio stations, to increased distance between the stations. These conditions will be used in the following discussion of interference.

Elimination of Interference.—In the actual case of radio communication, a receiving station R, Fig. 100, communicating with a transmitting station T_1, is generally surrounded by other transmitting stations which may send at the same or at different wave lengths, may be near to or distant from R, and may be heard by the station R even if the latter is exactly tuned to the station T_1, by reason of the fact mentioned above that signals are heard within a certain range of frequencies F_1F_2. It is therefore of interest to discuss the conditions of such interference, and to examine some of the methods used to prevent or eliminate it.

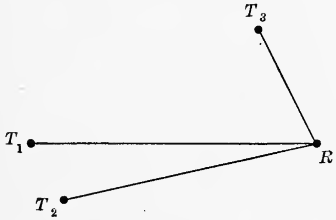


Fig. 100.

As in the previous case, we will consider for simplicity closed oscillatory circuits. Thus let M and P, Fig. 101, be two oscillatory circuits of fixed constants, having different natural frequencies, and in which oscillations are excited periodically. Let R be a receiving circuit comprising a detector, telephone receivers, and means for varying its natural frequency. First, operating each of the circuits M and P separately, and varying the natural frequency of circuit R, two resonance curves M and P of the current in R are obtained as shown in Fig. 102, having their maxima at frequencies corresponding to the natural frequencies of these two circuits respectively. If as before, I_t is the minimum telephone current producing an audible sound, it is seen from Fig. 102 that circuit M will be heard in the receivers when circuit R has a natural frequency between F_1 and F_2 while circuit P will be heard in the interval F_3 to F_4. Now if both circuits M and P are excited at the same time, it will be possible by proper tuning of circuit R to hear either one without hearing the other.

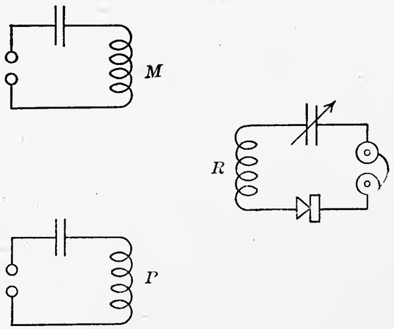


Fig. 101.

If the fixed natural frequencies of circuits M and P are closer together, as in the case of Fig. 103, it is seen that the two intervals of audibility F_1F_2 and F_3F_4 overlap. Then, if circuit R has a frequency comprised between F_3 and F_2, both circuits M and P will be heard simultaneously, and there will be interference in the reading of the signals sent by either one. In the regions F_1F_3 and F_2F_4, however, circuits M and P will be respectively heard alone, without interference. It is thus seen that the closer to each other the natural frequencies of the two transmitting circuits, the greater the overlapping range F_2F_3 and the smaller therefore the limits of frequency wherein the received signals are free from interference.

Consider again the case of Fig. 102. If the decrement of circuit P is increased without changing the frequency of its emitted oscilla-

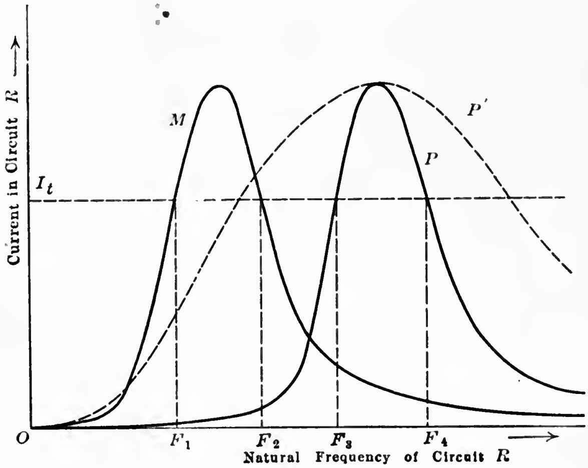


FIG. 102.

tions, its resonance curve will be broader, as shown by the dotted curve P'. And as this new resonance curve cuts the curve M above the line I_t it is seen that circuit P, because of its higher decrement, will cause interference over a part of the range of frequencies F_1F_2. This fact is important to remember: the higher the decrement, the broader the tuning, and the more difficult it is to eliminate interference due to a highly damped wave. For this reason, government regulations forbid the use of decrements greater than 0.2.

This property of highly damped waves to produce interference and to be heard in receiving sets even when they are tuned to widely different frequencies finds an application in the sending out of distress signals from a ship. The transmitting operator sending out such signals tightens the coupling of his transmitting antenna to produce high damping, as a result of which the distress calls are heard by radio receiving sets, even when tuned and working at quite different wave lengths.

Considering the case of interference as represented by the solid line curves M and P of Fig. 103, the question arises as to how this disturbing condition may be eliminated. If the coupling between circuit R, Fig. 101, and the transmitting circuits M and P is reduced, as for instance by moving circuit R away to a greater

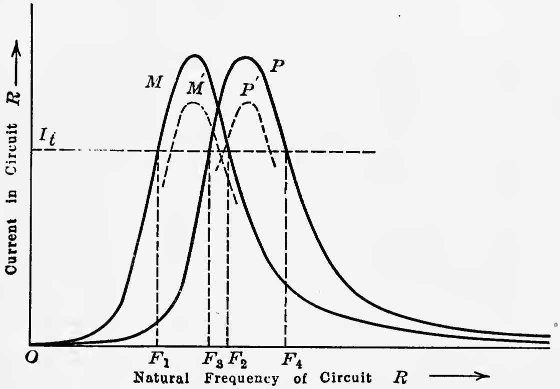


FIG. 103.

distance, the current induced in this circuit by the two transmitting circuits will be less, and the new resonance curves will be as shown by the dotted line curves M' and P' of Fig. 103. This is seen to entirely eliminate interference.

Now in practice, if a receiving station is hearing two transmitting stations simultaneously, it of course cannot be moved away to a greater distance. And this is where the use of *coupled receiving circuits* finds its great importance. Thus, consider the receiving set of Fig. 95. If, as was assumed, two transmitting stations produce interference, the natural frequency of the antenna circuit L_1C_1 is within the overlapping range of frequencies, and current is being induced in the antenna by both transmitting circuits. But the coupled circuit L_2C_2 derives its energy solely from the antenna circuit L_1C_1, and not from the distant transmitting sets, so that the antenna circuit may be considered as equivalent to two local transmitting circuits, each inducing currents in the L_2C_2 circuit. This case is then identical with that of Fig. 101, and to reduce or eliminate interference, it is simply necessary to reduce the coupling between the receiving antenna circuit and the secondary tuned oscillatory circuit. It is thus seen that the use of an intermediate oscillatory circuit between the telephones and the antenna will greatly assist in eliminating interference, if suitably coupled to the antenna.

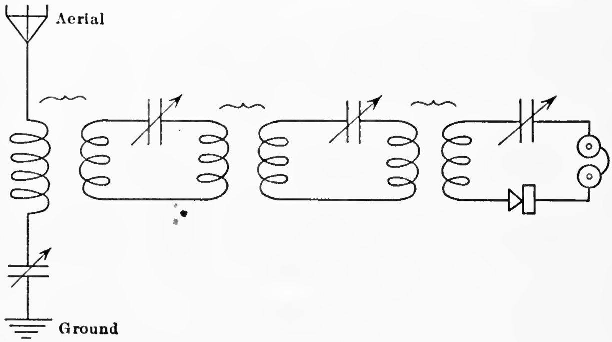


FIG. 104.

In some cases, there may be interference from several stations at the same time, and while the scheme of Fig. 95 may eliminate the interference from one or two stations, it may not eliminate the others. In such cases, the use of a number of intermediate circuits, as shown in Fig. 104, all tuned to the station to be received, and all suitably coupled will be found to successfully cope with the situation.

In some cases, such as that of Fig. 105, elimination of interference may be difficult, and unnecessary, for although it will be impossible to tune out the signals from the set producing the resonance curve M, without the use of intermediate coupled circuits, the two signals may be distinguished if the *audio frequencies* of the two circuits are different. Thus, if circuit M is excited by a 500-cycle alternator and circuit P by a 900-cycle alternator, the two messages, although heard simultaneously, will have different pitches, and the receiving operator can read the signals without error by concentrating his attention on the one pitch.

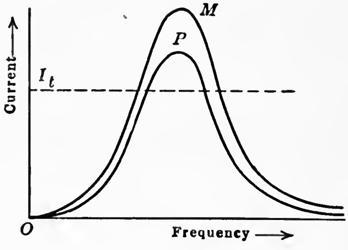


FIG. 105.

CHAPTER V

UNDAMPED WAVE RADIO TELEGRAPHY

Recalling some facts brought out in Chapters II and IV, a receiving circuit will respond with greatest amplitude to a transmitting circuit when it is tuned to the latter. The sharpness of tuning, as evidenced by the abruptness or steepness of the peak of the resonance curve, is a direct function of the decrement of the wave, and it was shown that the smaller the decrement, the sharper would be the tuning. A logical conclusion is then that exceedingly sharp tuning will be obtained with oscillations having zero decrement, that is, with undamped oscillations, or true high frequency alternating currents. It was explained at the conclusion of Chapter I that the accomplishment of radio communication necessitates the setting up of high frequency alternating currents in the transmitting circuit. In the earlier state of the science, the oscillatory discharge of a condenser provided a convenient way to obtain such high frequency currents, and was therefore used, as explained in the previous chapter. The drawback to that method, however, is that, due to the damping of the current in the transmitting circuit, the receiving circuit is excited only periodically, instead of continuously. Resonance and induction effects therefore cannot be expected to be as pronounced, perfect and effective as with true alternating currents, or undamped oscillations.

Another advantage of the use of undamped oscillations for transmitting signals may be seen from the following considerations. Suppose an antenna circuit is excited continuously by means of undamped oscillations, that is, by means of a true alternating current. The same antenna may be made to radiate the same amount of energy, at the same frequency, by damped oscillation excitation. In this case, however, instead of being radiated continuously, the energy is radiated in separate groups or wave trains. Since each wave train is of very short duration, the amount of energy radiated per wave train must be large in order that the total amount of energy radiated per second will equal that radiated in the case of undamped oscillation excitation. With the damped oscillations, this requires the setting up in the antenna circuit of considerably greater voltages and currents than with undamped oscillations. This in turn necessitates better insulation of the antenna, conductors of larger cross-section, and apparatus of greater power output. The advantage of undamped waves is then obvious. It consists essentially in using low power apparatus continuously rather than high power apparatus intermittently.

UNDAMPED WAVE TRANSMITTING CIRCUITS AND METHODS

The production of undamped oscillations presented great difficulties. It was first attempted by endeavoring to build an alternator which would deliver the high frequencies needed of thousands of cycles per second. Some special construction of generator or the application of new phenomena and principles was obviously necessary. The development of the science has now produced several methods of generating these high frequency alternating currents or undamped oscillatory currents. The methods studied in this book are the Alexanderson alternator, the Goldschmidt alternator, the oscillatory arc, and the three-electrode vacuum tube oscillator. The last method is merely mentioned here as it is fully studied in later chapters.

The Alexanderson Alternator.—The ordinary alternator construction embodying a wound rotor and stator is not practicable for the direct generation of the extremely high frequency alternating currents required in radio transmission. It is known that the frequency delivered by the ordinary alternator is given by the expression

$$\text{frequency} = \frac{\text{number of poles}}{2} \times \frac{\text{r.p.m.}}{60}$$

Thus, with a 2-ft. rotor diameter and a speed of 2500 r.p.m., the number of poles required on a machine to deliver 100,000-cycle current would be 4800, giving a pole pitch of 0.0157 in.[1] Such a construction is manifestly impracticable, and it is therefore necessary to resort to some special construction.

The Alexanderson alternator which has been used quite extensively is an inductor type machine. That is, the direct current

[1] A. N. Goldsmith, *Radio Telephony*.

field winding and the alternating current winding are both stationary, and the exciting magnetic field is made to pulsate and therefore induce an alternating emf. through the rotation of a toothed steel rotor. The actual construction is schematically represented in Fig. 106. The rotor is a steel disc, specially shaped to withstand the stresses resulting from the high operating speed which is normally 20,000 r.p.m., and accurately balanced to prevent undue vibration. This disc is slotted radially over its entire periphery, so that steel tooth projections are formed around the disc. The intervals between the teeth, which number about 300, are filled with phosphor bronze, a non-magnetic material, for the purpose of giving the disc a smooth rim to reduce windage and noise. The stator consists of two grooved steel

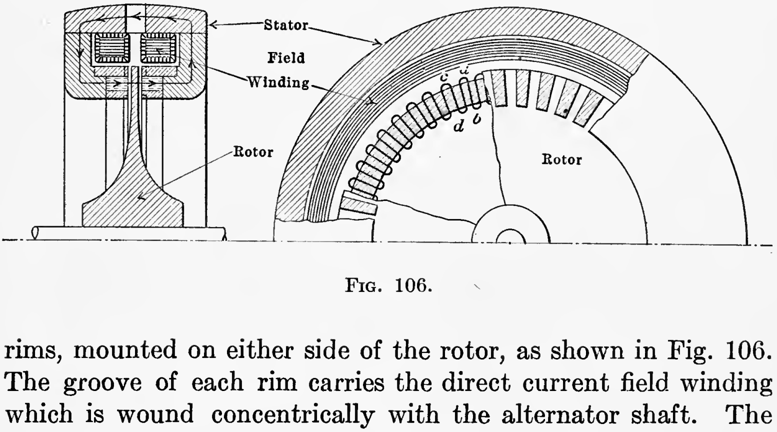


Fig. 106.

rims, mounted on either side of the rotor, as shown in Fig. 106. The groove of each rim carries the direct current field winding which is wound concentrically with the alternator shaft. The alternating current winding consists of a wire *abcd* . . . , wound in radial slots in the inner edge of the metal rims facing the rotor.

When the alternator is excited by sending a direct current through the field coils, a magnetic field is set up at each point of the circumference, as shown by the dotted lines in the left hand diagram of Fig. 106, which links with the alternating current winding. The edge of the rotor disc is in the path of this magnetic field. As the rotor is then set in motion, each section *ab*, *cd*, . . ., of the alternating current winding is passed alternately by a steel tooth of the rotor, and by a non-magnetic bronze sector. This produces a periodic variation of the permeability of the magnetic path in the gap between the pole faces and therefore a periodic variation or pulsation of the magnetic field. This in turn induces an alternating emf. in the alternating current winding on the stator discs.

The frequencies obtained with this type of machine range from 100,000 to 200,000 cycles per second, with power outputs as high as 200 kw. On account of the very high operating speed of the rotor, special precaution must be taken in the installation of the machine, bearings, oiling system, etc. It is also essential, in order to hold the frequency constant, to maintain an accurately uniform speed. This machine is used in large permanent stations only and is not economical for low power sets.

The Goldschmidt Alternator.—The Goldschmidt alternator makes use of the mutual induction phenomena between the rotor and stator windings of an ordinary alternator, to multiply the generated frequency within the alternator. The principle may be explained as follows.

An alternator with wound rotor and stator may be represented schematically as in Fig. 107, where the stator coil S is the direct

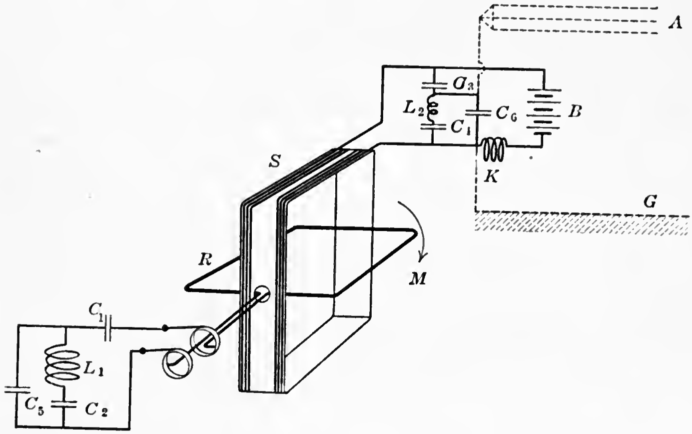


Fig. 107.

current field winding, energized by the battery B; while the rotor coil R rotates within the stator coil in the direction M. Through the rotation of the rotor, the steady magnetic field due to the coil S links the rotor coil in alternately opposite directions, this resulting in the induction of an alternating emf. in the rotor. The frequency of this emf. is equal to the frequency f of rotation of the rotor. The induced emf. reverses and passes through the value zero when the planes of the two coils R and S coincide. It is a maximum in the position shown in the figure. If the rotor is connected to an external circuit through its slip rings, an alternating current will flow through its winding.

In order to study more closely the phenomena taking place during one revolution of the rotor, refer to Fig. 108, where the stator has been omitted for simplicity, but will be assumed to be in a vertical position as considered in Fig. 107. Starting with the rotor in the position I, the emf. is zero. It steadily increases as the rotor turns to the position II. The induced emf. is say, in the direction *ABCD*. After passing position II, the emf. decreases to zero, without however changing in direction. As the rotor passes position III, the induced emf. reverses, and thus is now in the direction *DCBA*. It is a maximum for the position

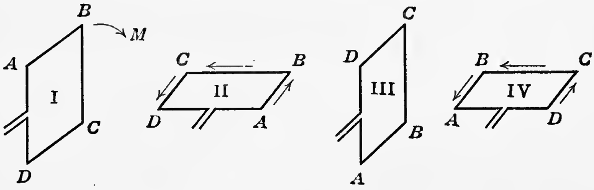


Fig. 108.

IV, and decreases again to zero when the rotor returns to the position I, when it reverses and the cycle repeats itself. The alterting current flowing in the rotor when the latter is closed follows the variations of the induced emf. and creates an alternating magnetic field around the rotor coil of the same frequency. However, as may be seen clearly from the figure, due to the simultaneous reversal of the position of the coil and of the current induced in it, the actual direction of this alternating field with respect to the stator coil will not reverse. The field will simply go in one half-cycle through the same values it had in the preceding half-cycle. The effect of the alternating current of frequency f flowing in the rotor is thus to set up around the stator coil S a pulsating magnetic field of frequency $2f$. This induces in the stator an alternating current of frequency $2f$, double the frequency of the rotor current. In order to increase these effects, the rotor current is made larger by tuning the rotor circuits to the frequency f of the rotor current. This is done by closing the rotor coil R, Fig. 107, on a circuit $C_1L_1C_2$, the natural frequency of which, taking the inductance R into account, will be equal to

f. Similarly, the stator S is tuned to the frequency $2f$ by shunting it with a suitable circuit $C_3L_2C_4$. The coil K is a choke coil preventing the alternating currents from flowing through the battery B.

Applying the same principle explained above, it is seen that the rotor now turns in the alternating field of frequency $2f$ set up by the alternating current flowing in the stator. This will induce in the rotor a current of frequency $3f$. This current is amplified by means of the condenser C_5, of suitable capacitance to make the circuit RC_1C_5 resonant for the frequency $3f$. Again, the rotor current of frequency $3f$ induces in the stator a current of frequency $4f$, which is made stronger by flowing through the suitably tuned circuit SC_3C_6.

Thus, if the original frequency f was 10,000 cycles per second, which is readily possible in an alternator having wound rotor and stator, the stator frequency will be 40,000 after four successive "reflections," or multiplications. And if the condenser C_6, Fig. 107, is made up of an aerial A and the ground G, this 40,000-cycle current will thus be impressed on the sending antenna.

The Oscillating Arc.[1]—The alternator methods of generating undamped oscillations, or high frequency alternating currents, are generally suitable only for permanent stations. They have also the great disadvantage that they require, for efficient and uniform operation, an extremely accurate speed control, since the frequency of the current generated is a direct function of the alternator speed. For these reasons, methods whereby undamped oscillations may be produced in a circuit by purely electrical rather than combined electrical and mechanical means are vastly preferable.

The oscillating arc is but a special case of a more general principle established mathematically in a later paragraph below. For the present, it is simply desired to give a rapid description of the functioning of the arc method of producing undamped oscillations.

An electric arc between carbon electrodes energized by a source of direct current has a resistance which by the definition of resistance is $R = \frac{V}{I}$, where V is the potential difference between the

[1] For more complete details of the theory of the arc and for a fairly complete bibliography on the subject, see *Proceedings of the Institute of Radio Engineers*, Vol. 4, 1916, p. 371, and Vol. 5, 1917, pp. 255 to 319.

arc electrodes, and I the current in the arc. If a curve is plotted giving values of current corresponding to various values of potentials, then the resistance is the slope of the curve, being equal to the ratio $\frac{dV}{dI}$. Depending on the shape of the curve, this ratio may be positive, as in the case of a metallic conductor, or negative, as in the case of the arc. Such a "characteristic curve" of an arc is given in Fig. 109. It is seen that the arc resistance is always negative, and increases with decreasing current.

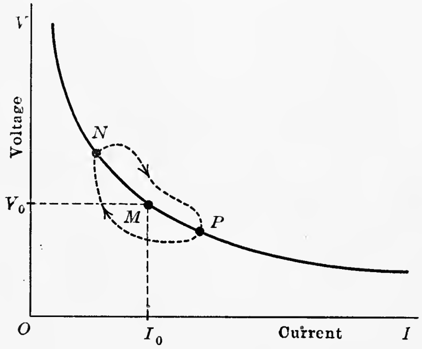


Fig. 109.

If then such an arc is supplied by a *constant current* generator, and shunted by a circuit containing inductance and capacitance, undamped oscillations will take place in the oscillatory circuit. This is illustrated in Fig. 110. A constant voltage generator G in series with a choke coil K forms the constant current supply. It is connected to the two electrodes A and B of a carbon arc. This arc is shunted by the coil L in series with the condenser C. To begin with, this shunt oscillatory circuit is disconnected from the arc, the condenser C is not charged and its plates are at the same potential. Upon being connected to the generator, the emf. applied across the arc will also be applied to the condenser through the coil L, and the condenser will begin to charge. The required energy will have to be supplied by the generator through the medium of a charging current. The total current supplied by the generator will then not flow through the arc, but a part will flow in the LC circuit. Since the generator is a constant current source of energy, any flow of current to charge C will be accompanied by a corresponding decrease in the current flowing through the arc. But, referring to the characteristic curve of the arc, a decrease in current will be accompanied by a rise in the potential difference between the electrodes and therefore across the LC

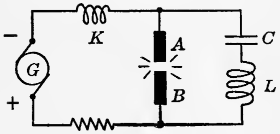


Fig. 110.

circuit. This will further increase the charging rate of the condenser. When the condenser is fully charged, the current flow in the *LC* branch circuit stops, and the current in the arc increases, since the generator current is no longer divided between the two circuits. But this increase in the arc current results in a decrease in the difference of potential at the arc electrodes, and therefore also across the *LC* circuit. The condenser thus begins discharging through the arc, increasing the current through the latter, and further reducing the emf., across the electrodes. When the condenser is fully discharged, conditions are returned to the original state, and the phenomenon repeats itself.

The relations between the instantaneous values of arc current and voltage may be represented graphically by means of the characteristic curve. The curve of Fig. 109 was explained to represent the corresponding values of arc current and voltage for *steady* conditions, that is, when the arc is not oscillating, such as is the case when the *LC* circuit is not connected to the arc electrodes. This curve is called the *static characteristic* of the arc. Now let V_o and I_o be the values of voltage and currents before the *LC* circuit is connected, *M* being the operating point. Then when the circuit is connected to the arc, as was explained above, the arc current decreases during the charging period, and the operating point of the arc moves from point *M* to point *N*, when the condenser is fully charged. Again referring to the above explanation, the condenser in discharging increases the arc current, and the operating point will oscillate between points *N* and *P* of the characteristic curve. However, certain phenomena take place within the arc when these oscillations occur, which have an effect such that the operating point *M* does not follow the static characteristic curve when oscillating between *N* and *P*.

At the instant when the condenser has fully discharged, the operating point is at *P*. The next moment will be a charging period of the condenser, during which the arc current decreases, as was explained previously. That is, the operating point *P*, will move toward the left. Now, due to the heavy current carried by the arc at the end of the discharge period when the operating point was at *P*, the carbon electrodes are quite hot and the space between them is of high conductivity. And due to the great frequency of the oscillations, the arc electrodes do not have time to cool down to their normal temperature as the current decreases, so that the resistance of the arc will be less than normal, and the voltage at the arc will be consequently smaller. In other words, the operating point in its motion from P to N will describe a curve which is *below* the normal or static characteristic curve. Similarly, as the current increases when the point moves from N to P, the electrodes do not have time to heat up to the normal value, the arc will then be of higher resistance, and the curve will lie above the static characteristic curve. As the arc oscillates, there is thus a lag of the electrode temperature behind the current, resulting in a closed characteristic curve, shown in dotted lines on Fig. 109, called the *dynamic characteristic* of the arc.

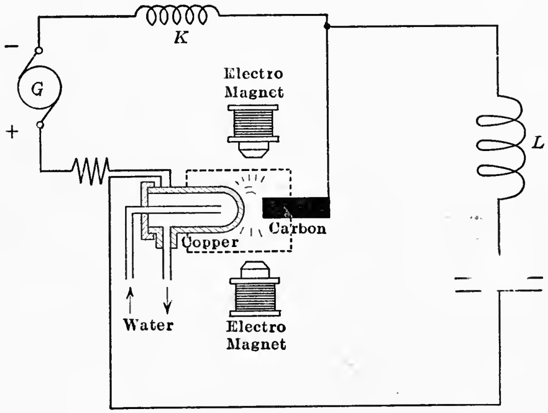


FIG. 111.

An oscillation giving a dynamic characteristic as shown in Fig. 109 was found to be possible only for small amplitudes, and frequencies not higher than a few thousand cycles per second. It is therefore not suitable for radio transmission purposes, although the shape of such an oscillation is very nearly a sine wave.

In order to obtain oscillations of great amplitude and of high frequencies, it is necessary to use an arc having a large negative resistance, that is, having a steep characteristic curve. This is obtained only for an arc which is very unstable, such as the Poulsen arc, in which the positive electrode is made of a copper jacket, Fig. 111, cooled by water circulation, and with the arc placed in a hydrocarbon atmosphere and in a strong magnetic field. With such an arc, it is possible to obtain frequencies up to two or three million cycles per second, and of great amplitudes. The charging current of the condenser may frequently be so great as to extinguish the arc every half-cycle. The oscillations generated by this method are thus not sinusoidal. The dynamic characteristic of the arc under such conditions is shown in Fig. 112.

Finally, there is a third kind of oscillation which may be obtained which corresponds to the case where the arc, after being extinguished due to the condenser charging current, is made to carry the discharge current. This current is in the opposite direction to the current normally supplied to the arc by the generator. In this case, the oscillatory current is greater than the supply current, and the arc functions more as an ordinary spark gap, producing slightly damped oscillations.

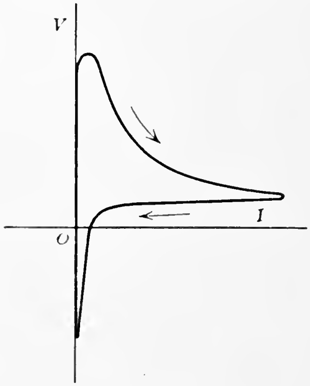


Fig. 112.

Mathematical Theory of Undamped Oscillation Generation

The following is not the mathematical theory of the oscillating arc, but a demonstration originated by A. W. Hull,[1] that undamped oscillations may be generated in an oscillatory circuit when shunted by a negative resistance and supplied by a constant direct current, provided certain critical conditions are fulfilled. This general principle will later be shown to be applicable to the three-electrode vacuum tube oscillator.

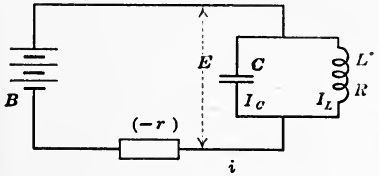


Fig. 113.

Consider an oscillatory circuit, Fig. 113, of constants L, C and R, shunted by a negative resistance $(-r)$ and a source of constant direct current B. If E is the instantaneous emf. across

[1] *Proceedings of the Institute of Radio Engineers*, Vol. 6, 1918, pp. 5 to 35.

the oscillatory circuit, then the currents in the coil, condenser, and external circuit will be, respectively,

$$I_L = \frac{E}{R} - \frac{L}{R}\frac{dI_L}{dt}$$

$$I_C = -C\frac{dE}{dt}$$

$$i = \frac{E}{(-r)} + i_0$$

where i_o is the constant current flowing in the circuit when no oscillations are taking place. Expressing the current i as the algebraic sum of the currents I_L and I_C, and eliminating i and E, we obtain the equation:

$$\frac{d^2I_L}{dt^2} + \left(\frac{R}{L} + \frac{1}{(-r)C}\right)\frac{dI_L}{dt} + \frac{1}{LC}\left(1 + \frac{R}{(-r)}\right)I_L + \frac{i_0}{LC(-r)} = 0$$

For the current I_L this equation gives a value which is an exponential, non-periodic function of time when

$$\left(\frac{R}{L} - \frac{1}{(-r)C}\right)^2 - \frac{4}{LC} > 0$$

If, however,

$$\left(\frac{R}{L} - \frac{1}{(-r)C}\right)^2 - \frac{4}{LC} < 0$$

the value obtained for I_L is an exponential and periodic function of time, and is therefore of the form of an oscillatory current, the damping factor of which is equal to

$$\frac{R}{2L} - \frac{1}{2rC}$$

where r is the numerical value of the negative resistance $(-r)$ as obtained from the characteristic curve of the negative resistance. This is seen to be less than the damping factor of a free oscillation, which was found to be equal to $\frac{R}{2L}$.

It is now seen that the damping factor of the oscillation may be positive or negative. In the first case, the oscillation is damped like a free oscillation of the type studied in Chapter II. Since the damping depends on the resistance R of the oscillatory circuit and also on the negative resistance $(-r)$, it is seen that by choosing a suitable value for $(-r)$, the damping factor, that is, the decrement of the oscillation may be made as small as desired. In the case of a negative damping factor, that is, if

$$\frac{R}{L} - \frac{1}{rC} < 0 \text{[1]}$$

or

$$Rr < \frac{L}{C}$$

the oscillations are not damped, and once started, will increase until the variations of current and voltage bring the operating point of the negative resistance to such points of the characteristic curve that the slope, that is, the value of $(-r)$, no longer satisfies the above condition. In other words, undamped oscillations are generated in the circuit.[2]

The frequency of these oscillations, as determined by the above equation, is given by the expression:

$$f = \frac{1}{2\pi}\sqrt{\frac{1}{LC} - \left(\frac{R}{2L} - \frac{1}{-rC}\right)^2} \qquad (17)$$

and is thus seen to be different from the case of free oscillations by the introduction of a term comprising the negative resistance. The demonstration given above would hold equally well for different connections of the various elements of the circuit.

Undamped Wave Radio Telegraph Transmitting Circuits

Undamped wave radio telegraph transmitting circuits do not differ in principle from damped wave circuits. They comprise an undamped wave generating circuit connected or coupled to a tuned radiating circuit. A method of connection of a Goldschmidt alternator was already pointed out in connection with Fig. 107. An Alexanderson alternator may be directly connected in the antenna circuit, as shown in Fig. 114, the antenna circuit being tuned to the alternator frequency, or the latter adjusted to the natural frequency of the antenna circuit. There is no special reason for coupling the alternator instead of connecting it directly to the antenna; and it is thus possible to increase the efficiency of the transmitting set.

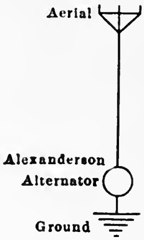


Fig 114.

[1] r denoting the absolute value of $(-r)$.

[2] A more general and complete demonstration may be found in *La Lumiere Electrique* of October 14, 1916, pp. 25 to 31, and the *Revue Générale de l'Electricité* of July 13, 1918, pp. 35 and 36.

The oscillatory arc may be connected directly in the antenna circuit, as shown in Fig. 115. It may also be made to generate undamped oscillations in an intermediate tuned circuit L_1C_1 coupled to the suitably tuned antenna, as shown in Fig. 116. In this case, everything applies which has been said in the previous chapter concerning the coupling of tuned circuits, and care must

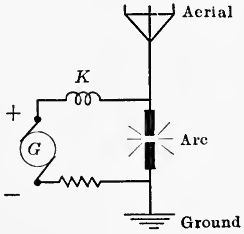


FIG. 115.

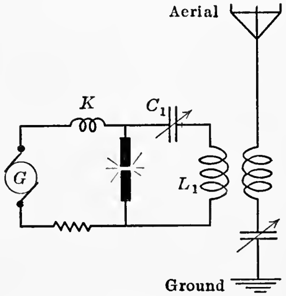


FIG. 116.

be taken, for obtaining a pure wave and good tuning, not to use so close a coupling as to produce a double humped resonance curve.

Certain remarks should be made concerning the location of the telegraph transmitting key in the above circuits. In the case of alternators of great output, delivering antenna currents of 60 to 100 amp. or more, the key cannot be used to break the antenna circuit, unless a special key relay is used. On the other hand, it is not recommended to insert it in the field circuit of the alternator, on account of the lagging effect this has on the break, due to the inductance of the circuit. A practice frequently followed, and which is possible on account of the very sharp tuning of the receiving circuits due to the use of undamped waves, is to *detune* the antenna circuit by a method such as shown in Fig. 117. The key is made to short circuit a few turns of the antenna inductance, thus slightly altering the transmitting wave length. The receiving circuit, which is tuned to the wave length transmitted with the key closed, does not respond or responds only very weakly to the detuned waves, so that dot and

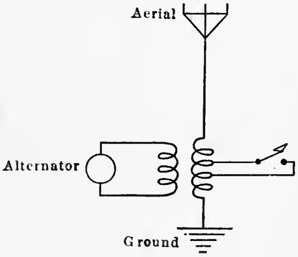


FIG. 117.

dash code sending may be accomplished by the corresponding detuning of the transmitting circuit.

In the case of an arc set, the most common practice is the detuning method just described. It is impracticable to place the key in the direct current supply circuit, since the arc once extinguished by the opening of the generator circuit would not start again upon closing the key.

UNDAMPED WAVE RECEPTION

The general principles given in the case of damped wave reception are applicable to undamped waves, if account is taken of certain modifications resulting from the different characteristics of undamped waves. The remarks on tuning and coupling of receiving circuits are of course applicable to undamped waves, and it should be noted here that the tuning is incomparably sharper than with damped waves. This means that the resonance curves show a marked and very sharply defined peak at resonance frequency, which thus greatly facilitates the elimination of interference from undamped wave sets of other wave lengths.

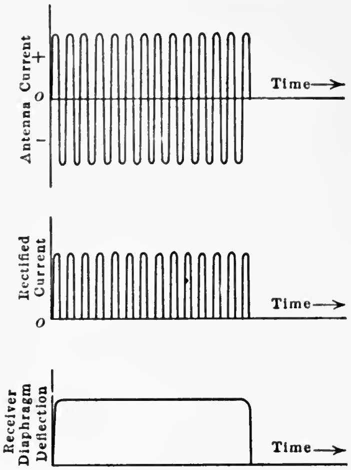


FIG. 118.

The ordinary detector methods cannot be applied in the case of undamped waves as may be understood readily from the following considerations. As the key of the undamped wave transmitting set is closed to send a dot or a dash, an uninterrupted undamped wave is sent out, inducing in the receiving circuit a true, undamped alternating current as shown in the upper curve of Fig. 118. If this wave were rectified and made to flow through the telephone receiver, the current would be as shown by the second curve of the figure. This would result in a permanent attraction of the receiver diaphragm during the entire duration of the dot or dash. The diaphragm would thus not enter into vibration, as it did in the case of damped waves which were broken up in wave trains spaced at regular intervals and occurring at some audio frequency. In the case of the undamped waves, simply a click would be heard at the start and the end of a dot or dash and this could be interpreted only with difficulty if at all. The most obvious solution of the problem is then to insert some sort of interrupter in the receiving circuit which will break the received current at regular intervals and thus give an interrupted current in the telephone circuit and set the diaphragm in audible vibration. Thus, the damped wave receiving circuit of Fig. 95 may be adapted to undamped wave reception by connecting an interrupter at M, Fig. 119.

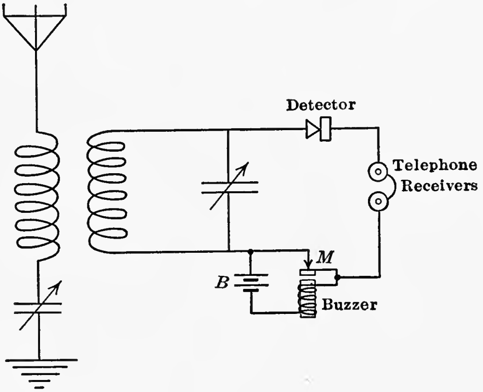


FIG. 119.

The interrupter shown here is an ordinary buzzer operated by the dry battery B. The vibrating contact interrupts the telephone circuit at point M. If the vibrator is adjusted for a frequency of say 500 to 1000 cycles per second, current impulses will flow in the telephone receivers at the same frequency whenever signals are received, and a good audible note will be heard.

Another method of receiving undamped waves is by the use of the *tikker*, which is illustrated in Fig. 120. The circuit comprises no detector but is provided with a vibrating contact T called a tikker. The telephone receivers are shunted by a condenser C. The operation is then as follows: When the tikker contact is closed, the alternating emf. set up in the tuned circuit L_1C_1 by the incoming oscillations charges the condenser C. When the tikker opens, this condenser discharges through the telephone receivers. It should be noted that no musical note is heard, since the tikker frequency bears no relation to the oscillation frequency and the condenser C is charged to different voltages for successive vibrations of the tikker. This in turn produces irregular current impulses through the receivers. This method is thus not very satisfactory but is used on account of

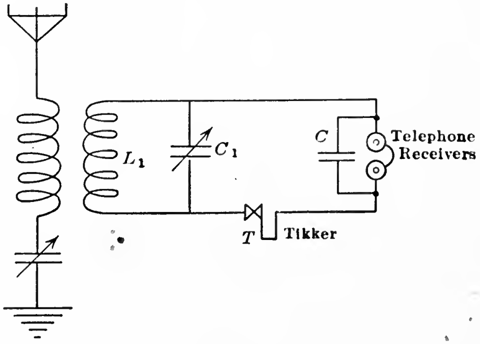


Fig. 120.

its simplicity. The tikker is generally a gold leaf vibrator contact similar to a buzzer; or else, it is made up of a steel wire held against the groove of a rotating metal pulley. The chattering of the steel wire against the pulley produces the required rapid make and break of the telephone circuit.

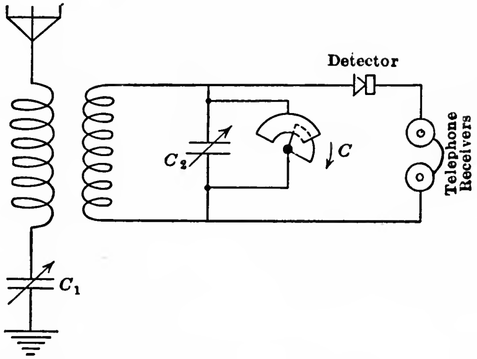


Fig. 121.

The *detuning method* of undamped wave reception is based on a different principle. It is illustrated in Fig. 121. The receiving circuit is essentially the same as a damped wave receiving circuit, except that one of the tuning condensers is shunted by a small condenser C made up of a set of movable plates which are rotated between a set of fixed plates. This rotation is effected by means of a small motor at a speed corresponding to some audio frequency. The set is first tuned by means of the condensers C_1 and C_2, with condenser C stationary. Condenser C is then rotated, periodically varying the capacitance of the secondary oscillatory circuit, thus throwing it periodically out of tune. During the detuned intervals it will not respond, or will respond only feebly, to the incoming oscillations. The oscillatory current is thus periodically varied in amplitude, or "modulated," and when rectified, produces a sound in the telephone receivers.

All of the above methods have the common disadvantage of being based on a periodical interruption of the received current. That is, the received energy is periodically not made use of in the telephone receivers. This waste of energy is objectionable, since one of the principal advantages of the undampd over the damped wave radiation is the continuousness of the energy supply to the receiving circuit, and also, since the total amount of energy received is at best very small. Therefore, a method which will utilize all the energy received is much to be preferred. For these reasons and others which will be pointed out later, the *heterodyne* method of reception is of great importance and is probably the best reception method for undamped waves.

Heterodyne Reception.—The heterodyne method consists essentially of superimposing upon the received undamped high frequency alternating current a locally generated undamped alternating current of slightly different frequency. These two currents combine, periodically adding and subtracting, and thus produce *beats* similar to those resulting from the combination of two sounds of almost the same pitch.

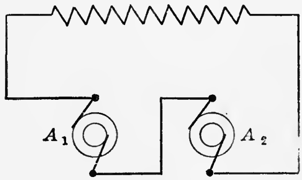


FIG. 122.

The frequency of these beats is equal to the difference of the frequencies of the component currents. These various points will be brought out in greater detail in the following paragraphs.

Consider the circuit of Fig. 122, in which an alternator A_1, generates an alternating current of frequency f_1, and an alternator A_2 generates an alternating current of frequency f_2. The resultant current in the circuit is the sum of the two component currents. Referring to Fig. 123, where the two component currents are represented respectively by the two upper curves, the total current in the circuit is at each instant equal to the algebraic sum of the component currents. The lower curve of Fig. 123 shows graphically this resultant current. It is seen to represent an alternating current of periodically increasing and decreasing amplitude, the maximum amplitude being equal to the sum of the maximum amplitudes of the component currents, and the minimum amplitude to their difference. The frequency of the beats is equal to the difference of the component frequencies.

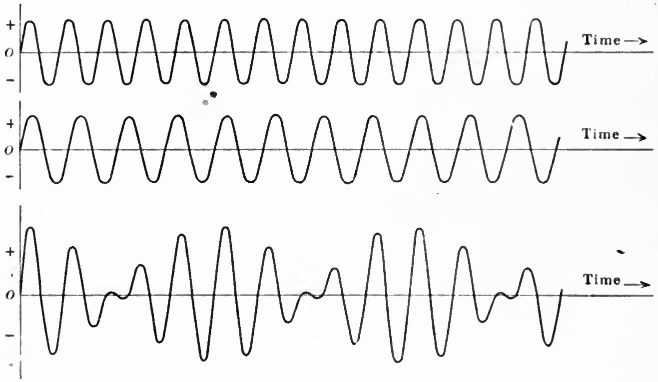


FIG. 123.

These statements may be demonstrated mathematically as follows. Let the two component currents be represented by the expressions:

$$i_1 = I_1 \sin 2\pi f_1 t$$
$$i_2 = I_2 \sin 2\pi f_2 t.$$

The resultant current i is then at each instant:

$$i = i_1 + i_2$$

which may easily be transformed into the expression:

$$i = \sqrt{I_1^2 + I_2^2 + 2I_1I_2 \cos 2\pi(f_1 - f_2)t} \cos [2\pi f_1 t + \varphi(t)]$$

which represents the beat current as described above.

Now consider the application of these principles to the reception of undamped waves. The receiving antenna circuit AG, Fig. 124, is tuned to the incoming waves. The received waves induce in the closed oscillatory circuit LC a current of the frequency f_1 of the received signals. This current may be represented by the upper curve of Fig. 123. As explained previously, this current when rectified by the detector would not produce a readable sound in the receivers. A high frequency alternator K is therefore connected or coupled to the LC circuit and is made to produce in the latter an alternating current of frequency f_2 slightly different from f_1 of the incoming signals. This may

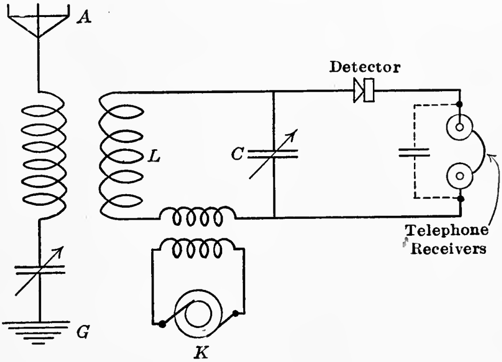


FIG. 124.

be represented by the second curve of Fig. 123. As just shown, these two currents produce beats in the LC circuit as represented in the last curve. This current when rectified by the detector and led through the telephone receivers is then of the form shown in Fig. 125, and it produces vibrations of the receiver diaphragm at a frequency equal to $f_1 - f_2$.

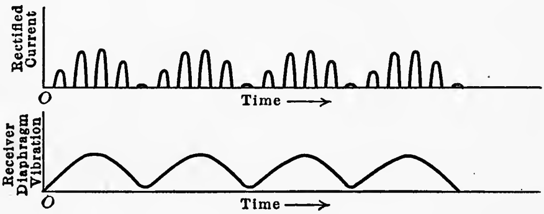


FIG. 125.

The circuit illustrated in Fig. 124 is one frequently used. There are, however, numerous modifications. Thus, the local generator circuit may be coupled to the closed oscillator circuit LC through the antenna circuit instead of directly. Another method, illustrated in Fig. 126, consists in using a special double winding telephone receiver, one winding carrying the rectified received current while the other carries the locally generated high frequency current.

Advantages of Heterodyne Reception.—The advantages of the heterodyne method of undamped wave reception are numerous, the more important ones being summarized here.

The heterodyne method does not simply permit the reception of undamped wave signals, but provides an amplification of the received signal current. This may be explained roughly by the fact that the locally generated energy is added to that received from the transmitting station to produce the beats. A brief demonstration of the fact, however, may be not without interest. Thus, if i_1 and i_2 are respectively the received and locally generated high frequency currents, the resultant current in the receiving circuit is

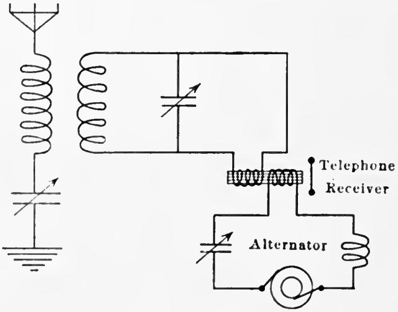


Fig. 126.

$$i = i_1 + i_2$$

The effect on the telephone receivers is however not proportional to the current, but to the energy, and is thus proportional to the square of the resultant current i and therefore to

$$i^2 = (i_1 + i_2)^2 = i_1^2 + i_2^2 + 2i_1 i_2$$

The first term, i_1^2, is the square of the received high frequency current. As this is of a frequency beyond the limit of audibility, it has no useful effect on the strength of the sound heard in the receivers. This is true also of the second term, i_2^2, which is the square of the locally generated current of nearly the same frequency. The signal intensity, as heard in the receivers, is thus represented by the term $2i_1i_2$. It is thus possible by suitably proportioning the value of the locally generated current to the received current, to obtain a greatly amplified signal.

Another advantage of the heterodyne method is that the note heard in the receivers, being dependent on the frequency of the current generated at the receiving station, may be adjusted by the receiving operator to suit his ear. This is of course impossible with damped waves, since the audio frequency is then determined wholly by the spark frequency at the transmitting station. An example of the advantage to be derived from this ability to vary the note at the receiving station is the possibility of obtaining a beat frequency equal to the natural frequency of vibration of the telephone receiver diaphragm. This makes possible a much stronger and purer note. But even without this adjustment, the note heard with the heterodyne is always of the greatest purity. It is of the nature of an even, musical whistle. This is due to the fact that the beat current curve is symmetrical, and the telephone receiver vibration thus has fewer harmonics than in the case of damped waves. This can be seen immediately by comparing the lower curves of Figs. 91 and 125.

The great advantage of the possibility of adjusting the pitch of the signals at the receiving station is increased by the fact that it assists in overcoming interference from other signals to a remarkable degree. Thus, suppose the signals to be received have a wave length of 3000 meters, corresponding to a frequency of 100,000 cycles per second, and that the station also receives interfering signals of a frequency of 100,500 cycles per second, that is, of a wave length of 2985 meters—only 15 meters less than the signals it is desired to receive. The sharp tuning afforded by undamped waves will already have minimized the interference. Now if the locally generated frequency is 101,000 cycles, the notes heard in the receivers will be respectively 1000 cycles and 500 cycles per second for the desired and the interference signals. These are two notes widely different and easily distinguished by the ear. But even better, if the locally generated frequency is adjusted to 100,500 cycles per second, the signals desired will have a pitch of 500 cycles, while no beats whatsoever and therefore no sound will be produced by the interfering signals which are thus entirely eliminated.

It would of course be possible to modulate the undamped waves at the transmitting station by combining two frequencies, instead of performing this combination at the receiving station. But then the advantages just pointed out would be lost and the advantage of the use of the simple damped wave receiving circuit thereby made possible would not compensate for the loss of these other advantages. Besides, the frequency of the modulated oscillations resulting from the combination of two undamped oscillations has a slightly variable frequency, and thus all the advantages of undamped waves would be done away with. The use of vacuum tubes has greatly simplified the operation of heterodyne circuits, as will be fully discussed in a later chapter.

CHAPTER VI

THE THREE-ELECTRODE VACUUM TUBE. GENERAL PROPERTIES

The purpose of the present chapter is to outline the properties and define certain constants of the three-electrode vacuum tube. This device has of late become of greatest importance in its applications to the art of radio communication. Its uses as detector, amplifier, oscillator and modulator are taken up in detail in the chapters following.

The three-electrode vacuum tube is essentially a device in which a stream of electrons emitted by a heated filament is controlled by the potential of an auxiliary grid electrode. The theory of operation of this device is better understood by first studying certain fundamental phenomena which control its action.

Electron Emission by Hot Bodies.—In the first chapter of this book it was shown that the gradual raising of the temperature of a material results in an increasing agitation of its constituent atoms or molecules. In the case of a metal or other substance containing free electrons, the latter participate in this thermal agitation. And, from what has been said before, since the trajectory of a free electron about its mean position is of a greater radius or length than that of an atom vibrating about its position of equilibrium, there will be an emission of free electrons from the heated body out into the surrounding space. This will occur at temperatures well below the temperature at which volatilization of the body becomes of importance. In fact, it has been found experimentally that electron emission from hot bodies increases rapidly with temperature and becomes readily appreciable for temperatures greater than that of dull red heat, that is, roughly greater than 700 deg. Kelvin.

Now if a body is heated to and maintained at a certain temperature T, it will emit electrons into space, and through the loss of these negative charges, it will become postively charged and thus will gradually tend to attract back the electrons emitted out into the surrounding space. This will then result in a condition of equilibrium, characteristic of the temperature and the substance. If the temperature is raised, then the atomic agitation of the substance will be such that the free electrons, having greater velocity and therefore greater kinetic energy, can overcome the attraction resulting from the positive charge of the body in greater number, and a new state of equilibrium will be established.

The number of electrons N emitted at a temperature T is given by the formula[1]

$$N = AT^c e^{-b/T} \tag{18}$$

where e is the base of the natural system of logarithms and A, b and c are constants depending on the chemical nature of the substance, its shape and certain other characteristics.

This emission of electrons by a heated body evidently has the effect of making the surrounding space a good conductor of electricity. This follows from the very definition of a conductor, which is a material containing free electrons. If then a positively charged body is placed near the hot metal emitting the electrons, these electrons will travel from the hot metal to the positively charged body and establish an electric current through the intervening space.

The Two-Electrode Vacuum Tube.—In order to study this phenomenon, consider a metal filament F, sealed in an evacuated glass bulb, Fig. 127, containing also a metal plate P. When the filament is heated by a battery A, it will emit electrons into the space surrounding it, in quantity expressed by the above equation. If the plate is now made positive with respect to the filament, as may be accomplished by connecting a battery B across the plate and the filament, there will be established a flow of electrons from the filament to the plate, and therefore an electric current in the circuit $FBPF$.

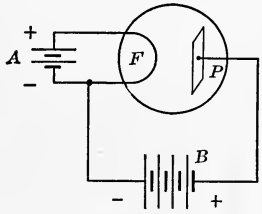


Fig. 127.

Keeping the filament temperature constant, the number of electrons N emitted by it per unit of time will similarly be constant. The number of electrons n attracted by the plate will be depend-

[1] O. W. Richardson, *Phil. Trans.* (A), Vol. CCI, 1903, p. 543.

ent on the plate potential with respect to the filament. Thus, starting with a potential difference of zero volts and gradually increasing it, the number of electrons reaching the plate and therefore the current in the plate circuit will gradually increase. However, a potential will be reached at which all the electrons emitted by the filament per unit of time will be attracted by the plate, so that further increase of plate potential will not produce any increase of plate current. This maximum current, beyond which there is no increase for increased plate potential, is said to be the *saturation current* of the tube for the corresponding filament temperature or current. This is represented by the curve of Fig.

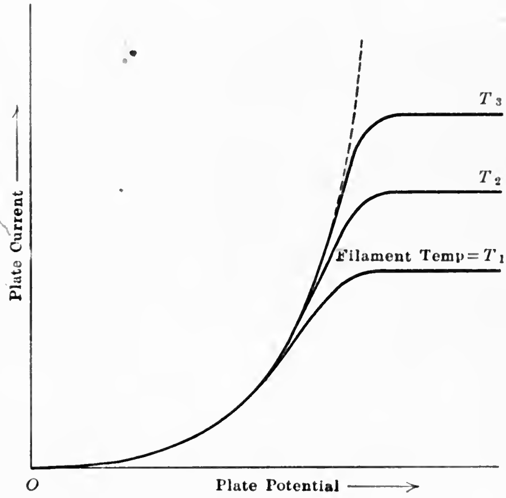


Fig. 128.

If it is then desired to have a greater current in the plate circuit, it is necessary to raise the filament heating current. For each filament temperature there is a definite value of saturation current which obtains when the plate attracts the emitted electrons at the same rate they are emitted.

The saturation current is seen to be a measure of the number of electrons emitted by the filament at the corresponding filament temperature. This current may therefore be expressed by a relation of a form similar to that given above for the electron emission by a hot body. It has been found that the constant c of the above equation is equal to ½. Thus, the saturation plate cur-

10

rent I for a given filament temperature T in a tube having a perfect vacuum, is given by the equation

$$I = A\sqrt{T}e^{-b/T} \tag{19}$$

where A, b and e have the same meaning as above. The variation of saturation current with filament temperature is represented by the curve of Fig. 129.

Now consider the tube with the filament cold (battery A disconnected) and the plate maintained at a certain constant potential V_1, by means of the battery B. Since the filament is then emitting no electrons, there will be no current in the plate circuit.[1] If the filament is then gradually heated, by gradually increasing the current from the battery A, it will emit more and more electrons as its temperature is raised and the plate current will increase as expressed by the equation above. Due to the steady stream of electrons passing from the filament to the plate, there are at any instant a number of electrons in the space between the filament and the plate. These electrons all move toward the plate and are absorbed by it, but, at the same rate, new electrons are emitted by the filament so that the number of electrons present at any moment in the space between the electrodes depends on the rate of absorption or attraction by the plate and on the rate of emission by the filament. The steady increase of the filament temperature will increase the electron emission from the filament, and therefore also the number of electrons present in the space of the tube. This group or cloud of electrons between the plate and the filament produces a negative *space charge*,

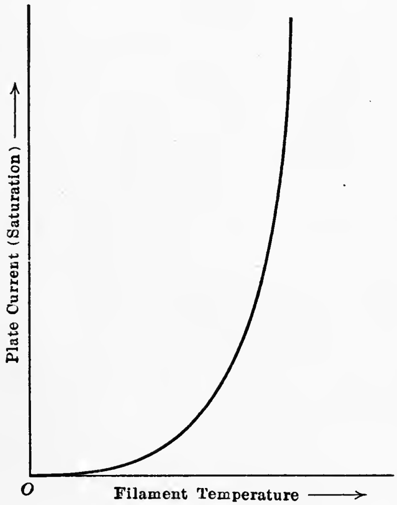


FIG. 129.

[1] This assumes the space within the bulb to be a perfect dielectric. In case there is gas present, ionization may take place. This case is considered later.

the effect of which upon the electrons leaving the filament is opposite to that of the plate potential.

As the filament current is increased far enough, an equilibrium is reached between the opposite effects of the plate potential and the space charge, and any further increase of filament temperature produces no increase of plate current because any additional electrons in the space of the tube makes the negative space charge overbalance the positive charge on the plate and repel the emitted electrons back toward the filament. If a greater plate current is then desired, it is necessary to increase the plate potential. Thus, for every value of plate potential, there is a corresponding filament temperature beyond which no increase in plate current is observed. This is represented by the curves of Fig. 130.

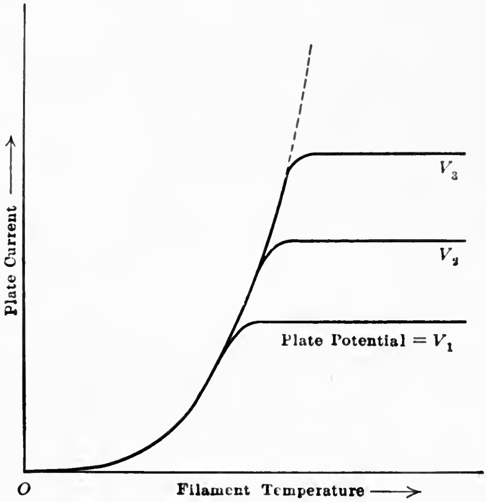


FIG. 130.

From the above explanation, it is seen that the maximum plate current which can be obtained for a certain value V of plate voltage, depends essentially on the shape, dimensions and spacing of the two electrodes. For a tube having a perfect vacuum, this current i, in the case of a straight filament 1 cm. long, placed along the axis of a cylindrical metal plate of radius r is given by the relation[1]

$$i = \frac{2\sqrt{2}}{9}\sqrt{\frac{e}{m}}\frac{V^{3/2}}{r}$$

[1] I. Langmuir, *General Electric Review*, 1915, p. 330.

where e and m are the charge and mass of an electron. Substituting numerical values, the current is

$$i = 14.65 \times 10^{-6} \frac{V^{3/2}}{r} = aV^{3/2},$$

i being in amperes, V in volts, r in centimeters and a a constant.

Effect of Gas on the Plate Current.—The above paragraphs are strictly applicable only in the case of a perfectly exhausted tube, where the electrons move unimpeded in the space separating the filament and the plate. If the tube contains gas, even a very small amount, the phenomena are modified when the plate potential is such that the tube operates at voltages greater than the ionization potential. Thus, consider the solid

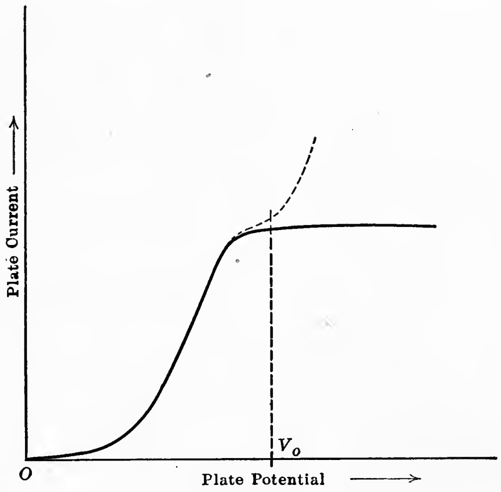


FIG. 131.

line curve of Fig. 131, showing the relation between plate current and plate potential for a constant filament temperature, in a tube having no trace of gas. The higher the plate potential, the greater the velocity of the electrons in their travel from the filament to the plate, especially after the current has reached its saturation value. Now if the tube contains a trace of gas, a certain potential V_0 will be reached for which the velocity of the electrons is great enough to disrupt the atoms of the gas by collision. This splits the atoms into free electrons and positively charged ions which travel, respectively, toward the plate and the filament, producing an increase in the current in the plate circuit. This is shown by the dotted line curve of Fig. 131.

Studying the effect of the presence of gas more minutely, let us consider one of the electrons emitted by the hot filament. It is under the constant action of the repelling force from the negatively charged filament and the attracting force from the positively charged plate. The resultant velocity of the electron thus constantly increases as the electron travels toward the plate. The electron then has maximum speed upon reaching the plate, and ionization by collision will therefore generally begin in the immediate neighborhood of the plate. This may be observed as a blue glow surrounding the plate.

This blue glow may be observed even in a highly evacuated tube if operated at a plate potential much greater than that for which the tube was designed. In this case, the glow is due not so much to ionization of the gas which may have been left in the tube, as to the ionization of gases resulting from the volatilization of the plate under the effect of the impact of the electrons against it at such great velocity. This may be avoided by careful manufacture of the tube, as will be explained at the end of this chapter. Ionization of the gas left in the tube is generally prevented at very low gas pressures, since the mean free path of the electrons is then greater than or at least of the order of the distance between the filament and the plate. Under such conditions, the gas molecules are so far apart that the number of collisions is so reduced as to be negligible.

The two-electrode tube described above finds application as a rectifier, since no current will flow in the tube if the plate is made negative with respect to the filament, for the electrons are then prevented from leaving the filament by the repulsion from the plate, but current will flow when the plate is made positive to the filament. It should be noted that this rectifying property is destroyed if the plate is heated to incandescence, as occurs readily when too high a positive potential is applied between the plate and the filament, thus increasing the velocity of the electrons and consequently the heat developed at the plate by the bombardment and sudden stoppage of the electrons. The quantity of this heat may be considerable, when it is considered that the velocity of the electrons may reach many miles per second, thus giving to each electron a fairly large amount of kinetic energy which is partly transformed into heat as the plate stops the motion.

THE THREE-ELECTRODE VACUUM TUBE

The three-electrode vacuum tube differs from the two-electrode tube studied above in the addition of a third electrode *G*, Fig. 132. between the filament and the plate in the path of the electrons. This third electrode may be a perforated plate or mesh or grid of fine wires, through the openings of which the electrons must pass in their travel from the filament to the plate. By applying a positive or a negative potential to the grid electrode with respect to the filament, it is possible to accelerate or decelerate the electrons moving from the filament to the plate, and to counteract or increase at will the effect of the space charge. The third electrode or grid thus offers a means of controlling the current in the plate circuit, without changing the plate potential or filament temperature. One of the advantages of this method of control is that while the plate current may be quite large, and the energy set in motion in the plate circuit quite considerable, yet the energy required to charge the grid to the desired potential is extremely small, due to the small capacitance of the grid with respect to the filament.

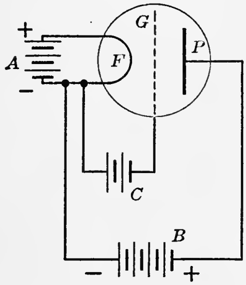


Fig. 132.

Characteristic Curves.—The operation of the three-electrode vacuum tube may be explained in a general manner as follows. Suppose that no difference of potential is established between the grid and the filament, the grid not being connected to any circuit. The tube will then operate like a two-electrode tube and a steady stream of electrons will flow from the filament *F* heated by the battery *A*, to the plate *P*, under the influence of the emf. of the battery *B*, Fig. 132. This plate current is limited, as was explained for the two-electrode tube, by the space charge due to the electrons present in the space between the plate and the filament.

Now if the grid is given a *negative* potential with respect to the filament, for instance by connecting a battery *C* with its plus terminal to the filament and its minus terminal to the grid, the space charge effect will be increased. That is, the negative charge of the grid will repel the electrons emitted by the filament back toward the latter, the result being a decrease in the plate current. This decrease will be greater, the more negatively the grid is charged, and finally, the grid may be made sufficiently negative to entirely stop the flow of electrons from the filament to the plate.

If the grid is now charged *positively*, the negative space charge due to the electrons in the tube will be partly neutralized, and

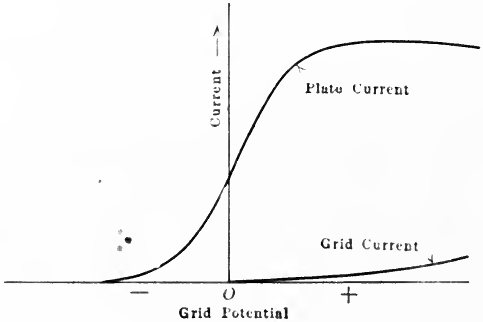


FIG. 133.

the plate current consequently increased, assuming the filament temperature and plate potential to be constant. In other words, the positive grid will assist the plate in attracting the electrons

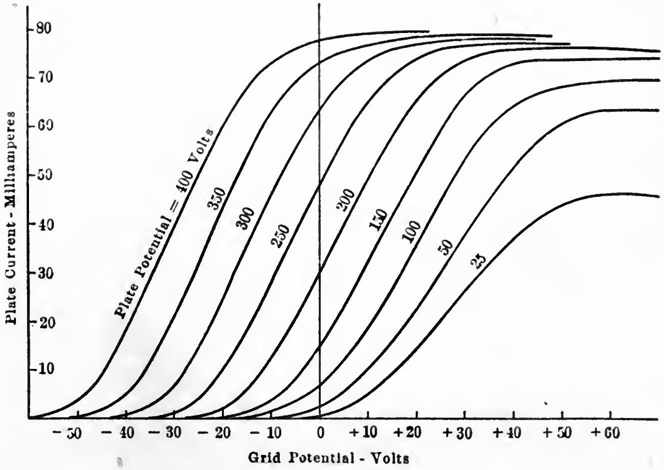


FIG. 134.

emitted from the filament. Thus, for increasing positive grid potentials, the plate current increases until the saturation current corresponding to the existing filament temperature is reached.

These variations of plate current with the grid potential are shown graphically in the curves of Fig. 133, which are called the *static characteristic curves* of the tube for the plate voltage and filament current (or temperature) specified. There is such a pair of characteristic curves for each value of plate potential and filament temperature. Assuming the temperature to be constant, a family of curves may be obtained for various plate voltages, as shown in Fig. 134. The higher the plate potential, the more the curve is shifted to the left. These relations may also be shown as in Fig. 135.

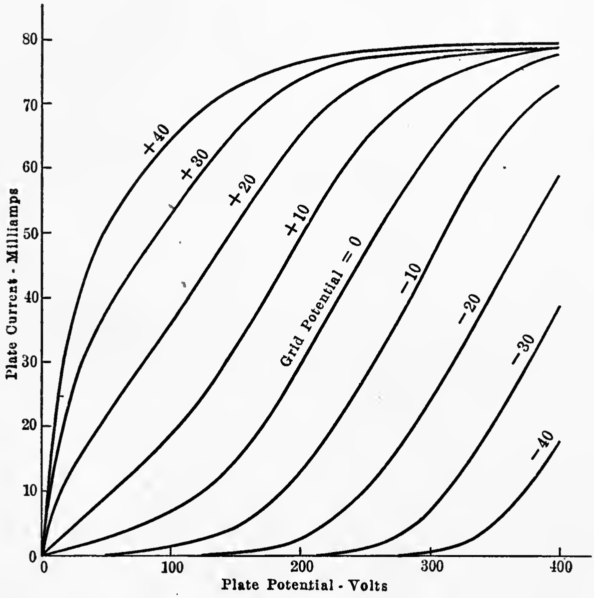


Fig. 135.

It should be noted that when the grid is made positive, it collects a few electrons itself, giving rise to a current in the grid circuit *FCGF*, Fig. 132. In the useful operating range of the tube, this current is a very small one, ranging from zero to about $\frac{1}{100}$ to $\frac{1}{10}$ of the plate current. Also, the higher the plate voltage, the smaller the grid current for a given grid potential.

The plate current increases, as was explained above, with increasing positive grid potential, until the saturation current is reached. All the electrons emitted by the filament are then attracted to the grid and plate, and further increase in the grid potential cannot result in further increase in the plate current. In fact, as the grid is made more and more positive, it attracts a greater number of electrons to itself, thereby increasing the grid current in the circuit *FCGF*. And since the total number of electrons emitted per unit of time remains constant for a fixed filament temperature, this increase in grid current brings about a corresponding decrease in the plate current. The three-electrode vacuum tube is not considered under such conditions in the circuits described in this book.

Mathematical Theory of Vacuum Tube Properties

It is helpful in gaining a thorough understanding of the operation of the three-electrode vacuum tube and in the design of vacuum tube circuits to express the properties explained above in mathematical terms. This is indicated here in a general manner.[1]

First, suppose that the grid is connected directly to the negative terminal of the filament, which will be taken as the reference point for the various potentials. The grid potential is then zero. The electrons emitted by the filament, however, are attracted to the plate by that part of the electrostatic field of the plate which is not screened by the grid, and which can act through the openings of the grid. This so-called "stray field" is evidently directly proportional to the plate potential E_p and dependent on the relative positions and shape of the electrodes in the tube. Thus, neglecting a generally small constant factor, the stray field E_s may be expressed by the relation

$$E_s = \frac{E_p}{k},$$

where k is a constant greater than unity.

If a potential E_g is now applied to the grid, equal to E_s but of opposite polarity, the effect of the plate potential upon the electrons emitted by the filament will be exactly neutralized, and

[1] In writing the following paragraphs, use has been made of the following papers:

H. J. Van der Bijl, *Physical Review*, Vol. xii, No. 3, 1918, p. 171.

J. M. Miller, *Proceedings Institute Radio Engineers*, Vol. vi, 1918, p. 141.

S. Ballantine, *Proceedings Institute Radio Engineers*, Vol. vii, 1919, p. 129.

the conditions will be the same as if the plate potential E_p had been reduced to zero, instead of the potential

$$E_g = -E_s = -\frac{E_p}{k}$$

actually applied to the grid. In other words, the grid potential E_g produces around the filament a field of the same strength as the plate potential E_p equal to kE_g. Similarly, a change of the grid potential equal to dE_g has an effect on the plate current equivalent to a change of the plate potential dE_p, equal to kdE_g. Since a change of grid potential has the same effect on the plate current as a change of plate potential equal to k times the grid variation, it follows that the plate current of the tube for a given value E_p of plate potential and E_g of grid potential, will be the same as if the tube had no grid, and was operated at a plate potential V equal to $E_p + kE_g$. Referring then to the expression for the plate current in a two-electrode tube, it was shown that the plate current I_p was given by the relation, $I_p = aV^{3/2}$. The plate current in a three-electrode vacuum tube is then expressed by the relation[1]

$$I_p = a(E_p + kE_p)^{3/2}$$

where a is a design constant for the tube. This relation holds only for that portion of the plate-current grid-voltage characteristic curve where the effect of the saturation current is negligible.

From this expression, it is seen that there will be no plate current if a potential is applied to the grid equal to

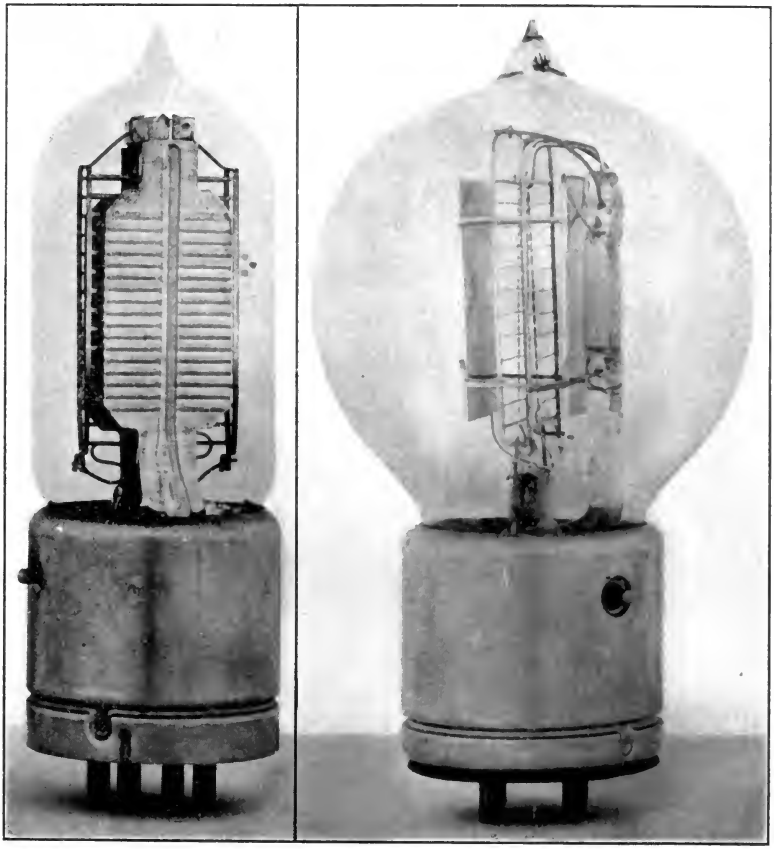
$$E_g = -\frac{E_p}{k}.$$

Also, it was explained in the case of the two-electrode tube that no current will flow to the plate if the potential V is negative.

[1] This equation does not exactly represent the conditions in certain designs of vacuum tubes. In certain cases, a more nearly correct equation is

$$I_p = a(E_p + kE_g)^2.$$

The value of the exponent depends on the construction of the tube, and also on the portion of the characteristic curve considered. If only the lower part is used, the exponent 2 is evidently better and more representat ve than $3/2$. If only a small portion of the central straight part of the curve is used, then a straight line equation with the exponent equal to unity may be suitable.



(*A*) (*B*)

Plate 3.—(*A*) Three-electrode vacuum tube used mainly as detector and amplifier. Signal Corps tube type VT–1. (*B*) Three-electrode vacuum tube used mainly as oscillator and modulator. Signal Corps tube type VT–2.

(*Facing page* 154.)

Plate 4.—Three-electrode vacuum tube used as high power oscillator and modulator. The plate has been forced downward to expose the grid and filament axially mounted. Note the cooling flanges on the plate. Signal Corps tube type VT–18.

This holds true for the three-electrode tube. Thus, there will be no plate current if

$$E_p + kE_g = \text{or } < 0$$

or, if

$$E_g = \text{or } < -\frac{E_p}{k}.$$

Amplification Factor.—The voltage amplification factor of a three-electrode vacuum tube is by definition the ratio of the plate and grid potential variations which will produce an identical variation of the plate current. From the preceding discussion, it is obvious that this factor is equal to the term k of the previous equation.

The amplification factor varies inversely as the spacing between the grid wires, since the stray field due to the plate will be smaller the closer the mesh of the grid. It also varies directly as the ratio of the plate-filament and grid-filament distances, as may readily be understood from the fact that the closer the grid is placed to the filament, the smaller the grid potential required to set up a field around the filament equivalent to the stray field of the plate. Thus, for obtaining very large amplification factors, it is necessary to use a fine grid, mounted at a small distance from the filament as compared to the distance of the plate from the filament. Voltage amplification factors of several hundred have been obtained in some tubes. Such a factor results in a steep static characteristic curve which is of great advantage in vacuum tube applications, as will be seen in later chapters.

The amplification factor k is actually not constant for a given tube but varies somewhat with the plate and grid potentials. This is due to the fact that when these potentials are changed, the average distribution of the electrons in the space of the tube is altered so that the section of the path from filament to plate which has the maximum density of the electron "cloud," is shifted with respect to the grid and plate. The effect of this shift is correspondingly to alter the relative effects of the plate and grid potentials upon the electron flow in the tube. This variation is small however and is neglected here.

Internal Plate Resistance and Impedance.—In order to obtain a current flow across a vacuum tube, it is essential to establish and maintain a difference of potential between the plate and the heated filament, making the plate positive with respect to the filament. This necessitates the use of a generator or battery *B*, Fig. 132, connected between the filament and the plate, the function of which is to furnish the energy required to accelerate the electrons emitted by the filament and attract them to the plate. Without this battery, the positive charge of the plate would very soon become neutralized by the electrons collected by it and the electron flow would cease.

In order to understand fully the function of this generator or battery *B*, we may follow one of the electrons in its travel around the plate circuit *FPBF*. All the electrons making up the plate current are under the same conditions, so that tracing the action of one of them will give an idea of the part played by the battery *B* in setting up and maintaining the current flow in the plate circuit.

As the electron is ejected by the heated filament, it escapes from the metal making up the filament and is then in the evacuated space of the tube—a space empty of all matter—and there, is subjected to the electrostatic field set up between the filament and the plate by the battery *B*. This is therefore similar to the case, studied in Chapter I, of an electric charge without material support and free to move in empty space. Applying the results of this previous study, the following phenomena are seen to take place in the vacuum tube.

As soon as the free electron is in the empty space of the tube, it comes under the full influence of the electrostatic field of the plate and is set into motion in the direction toward the plate. As was explained in Chapter I, the setting into motion of the electron results in a distortion of its electrostatic field and the transformation of a part of its electrostatic energy into electromagnetic energy. This transformation was shown to be at the expense of the electrostatic field producing the motion of the electron, that is, in the present case, of the field of the plate. Thus, finally, some of the electrostatic energy of the field of the plate is transformed into electromagnetic energy and stored in the electromagnetic field of the moving electron.

The electron moves with increasing velocity as it comes nearer to the plate, and its energy is thus stored in its magnetic field in increasing amounts, with a corresponding decrease of its electrostatic energy. Now it was also shown in Chapter I that if the electrostatic field producing the motion of a charge was reversed, the charge would come to rest and the electromagnetic energy of the charge would revert back to the original form of electrostatic energy, and there would be a restoration to the electromotive field of the energy derived from it and temporarily stored as magnetic energy around the moving charge. In the case of the vacuum tube, however, the electron is not brought to rest in this manner, but is made to collide with the plate, so that the energy stored in the magnetic field of the moving electron by the battery *B*, instead of being restored to the battery, is surrendered to the plate as heat.

An approximate conception of this process may be obtained by assuming that the electron is brought practically to rest by collision with one or more of the atoms making up the metal of the plate. Its store of electromagnetic energy, which may be considerable, due to the great velocity gained during its travel from filament to plate, is then transformed into electrostatic energy and is spent in the atomic collision which results in a momentary distortion of the atoms of the plate, and in their consequent vibration as they oscillate about their position of equilibrium. This increased agitation of the atoms of the plate results in an increase of temperature of the plate. Thus the energy supplied by the battery *B* serves the useful purpose of attracting the electrons to the plate so that they may deliver their negative charges to it and thereby produce a current flow in the external plate circuit, but the energy from the battery which is represented by the motion of the electrons is permanently lost in heat at the plate as the result of the collision. There is thus a permanent expenditure of energy on the part of the plate battery of a nature similar to that due to the ohmic resistance of a metallic conductor, as interpreted in Chapter I, that is, of the nature of a frictional energy loss due to collisions of the moving electrons with the atoms of the metal.

From the above explanation, it is seen that the energy loss is directly proportional to the electromagnetic energy of the electron moving in the empty space of the tube. And since for the total electron current, this is proportional to the square of the current, as shown in Chapter I, the total energy *W* spent by the battery *B* per unit of time for maintaining the plate current *I*, may be expressed by the relation

$$W = RI^2$$

where *R* is a constant. Neglecting the ohmic resistance o the external plate circuit *PBF*, this constant is then the *internal plate resistance* of the vacuum tube.

As an illustration of the heating of the plate under the bombardment of electrons, it may be said that when the plate of a vacuum tube is held at a fairly high potential with respect to the filament, the electrons reach considerable velocity and the plate becomes red or white hot within a few seconds.

The internal plate resistance of a three-electrode vacuum tube may be defined as the ratio,

$$R = \frac{dE_p}{dI_p}$$

And since

$$dE_p = k dE_g,$$

we may write

$$R = k \frac{dE_g}{dI_p} \tag{20}$$

The internal plate resistance is thus the product of the amplification factor of the tube and the reciprocal of the slope of the plate-current grid-voltage static characteristic curve. By reference to Fig. 133, it is seen that the internal resistance of a vacuum tube operating at constant plate potential and filament temperature may be varied by changing the grid potential.

Another expression of the internal plate resistance of the tube may be obtained as a function of the plate current and plate and grid potentials as follows. From the above expression of the plate current, the internal plate resistance will be

$$R = \frac{dE_p}{dI_p} = \frac{1}{2a(E_p + kE_g)},$$

and multiplying the numerator and denominator of this fraction by $(E_p + kE_g)$, the resistance is

$$R = \frac{E_p + kE_g}{2a(E_p + kE_g)^2} = \frac{E_p + kE_g}{2I_p}. \tag{21}$$

From the above equations it is seen that *the internal d.c. plate resistance of a tube is a minimum for the grid potential corresponding to the maximum slope of the curve,* which is the point of inflexion located midway along its rising branch. The resistance is infinite at the point of zero plate current, a minimum at the point of inflexion, and rises again rapidly after the saturation current has been reached. For a given value of grid potential, the internal plate resistance decreases gradually from infinity for zero plate potential as the plate potential is raised, until saturation becomes noticeable, when it rises again.

It should be noted that when an alternating potential is impressed upon the grid, as is done in radio work, a pulsating current flows through the plate circuit, and the capacitance of the plate with respect to the filament may not always be neglected, especially at the higher frequencies. The tube is then equivalent for the alternating current component to a condenser having the capacitance of the plate to the filament, shunted by a resistance equal to the effective internal plate resistance of the tube. The combination has an impedance equal to the *internal plate impedance* of the tube. In order to simplify computations in this book, the internal plate impedance will generally be considered as a pure resistance, and only in cases of extremely high frequencies will the capacitive reactance of the tube be considered.

Mutual Conductance.—The conductance of a conductor is defined as the reciprocal of its resistance. The reciprocal of the internal resistance of a three-electrode vacuum tube is equal to

$$\frac{1}{R} = \frac{1}{k}\frac{dI_p}{dE_g}$$

The quantity $\frac{dI_p}{dE_g}$, which is the slope of the grid-voltage plate-current static characteristic curve is called the *mutual conductance* of the tube. It follows that the mutual conductance G_m is given by the relation

$$G_m = \frac{dI_p}{dE_g} = \frac{k}{R} \tag{22}$$

It is seen to be a maximum at the point of inflexion of the curve, and to be equal to zero at the points of zero plate current and maximum saturation current.

Measurement of Vacuum Tube Constants.—There are many methods of measuring the various constants defined above. They are only indicated here in a general way, reference being made to the papers mentioned in the foot note, page 153. In general, it may be said that there are the static and the dynamic methods of testing the tubes. The former makes use of the static characteristic curves referred to above, or of conditions existing when the currents and voltages have steady values. The latter method attempts to reproduce conditions similar or equivalent to the operating conditions of the tube. Both methods yield interesting results from which it is possible to obtain data required for the design of vacuum tube circuits.

Amplification Factor.—The static method[1] of measuring the amplification factor of a three-electrode vacuum tube is to apply a certain positive potential to the plate, and then impress sufficient potential to the grid to bring the plate current down to zero, as indicated by an ammeter in the plate circuit. The amplification factor is the ratio of the plate and grid voltages prevailing at the time this condition of zero plate current obtains. The measurement is repeated for various values of plate potential, and the average value of the results is usually taken as the correct figure.

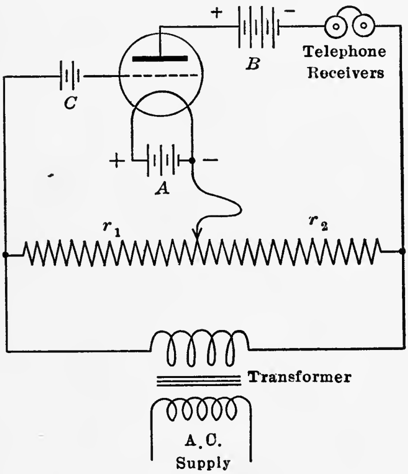


FIG. 136.

The dynamic method[2] is based on the theorem demonstrated in Chapter VIII, that the alternating current flowing in the plate circuit of a three-electrode vacuum tube and superimposed upon the normal direct current as a result of an alternating emf. e_g impressed between the grid and filament is the same as would flow in the external (metallic) plate circuit if it were connected to an alternator of emf. ke_g and internal resistance R, the factors k and R being respectively equal to the voltage amplification factor and the internal resistance of the tube. This is seen to be a consequence of the preceding discussion.

The circuit used is illustrated in Fig. 136. The filament of the tube is heated by the battery A, while the plate and grid potentials are obtained from the batteries B and C respectively. These are connected through the resistances r_1 and r_2, which are adjustable by means of a sliding contact connected to the filament. The

[1] Van der Bijl, *loc. cit.*
[2] J. M. Miller, *loc. cit.*

plate circuit also comprises a telephone receiver head set. An alternating emf. is then applied across the slide wire resistance. This results in a simultaneous application of alternating grid and plate emf's., which when measured with respect to the filament terminal are 180 deg. out of phase as may be readily seen by reference to the figure. If the sliding contact is then so adjusted that the drop across r_1 is equal to that across r_2, these two emf's. will have equal and opposite effects on the electron flow in the tube, and there will be no alternating current component in the plate circuit. The existence of this condition is denoted by the absence of sound in the receivers. The amplification factor is then obviously given by the relation

$$k = \frac{r_2}{r_1} \tag{23}$$

By this method, it has been found that the amplification factor varies somewhat with the plate and grid potentials applied by the batteries B and C, as previously explained.

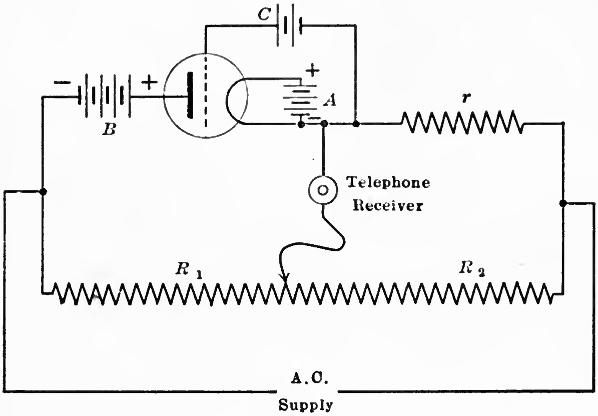


Fig. 137.

Internal Plate Resistance.—The internal plate resistance of a tube may be determined for any value of grid and plate potential by computing the slope $\frac{dE_p}{dI_p}$ of the plate-voltage plate-current static characteristic curve corresponding to the specified grid potential; or else by determining the reciprocal of the slope $\frac{dI_p}{dE_g}$ of the plate-current grid-voltage characteristic. The resistance is then $k\frac{dE_g}{dI_p}$.

11

The dynamic method[1] is illustrated by the circuit of Fig. 137. The plate-to-filament path of the tube is connected as one branch of a Wheatstone bridge, one other branch of which is a known resistance r, and the two other branches are adjustable by means of sliding contacts connected to the bridge circuit which contains a telephone receiver. The latter may be connected through an amplifier for greater sensitivity. The internal resistance is determined for fixed plate and grid potentials furnished by the batteries B and C. The bridge is connected to a supply of alternating current and balanced until the receivers are silent, when the internal plate resistance of the tube is given by the well known expression:

$$R = \frac{R_1}{R_2} r \tag{24}$$

Mutual Conductance.—The mutual conductance of a three-electrode vacuum tube may be determined by measuring the slope of the grid-voltage plate-current static characteristic curve. Dynamic methods may be used as described in Mr. S. Ballantine's paper,[1] but these are not given here. The mutual conductance may also be calculated if the amplification factor and internal resistance of the tube are known, from the relation

$$G_m = \frac{k}{R}$$

Construction of Vacuum Tubes

The types of construction of three-electrode vacuum tubes which have been used are numerous. The best ones have followed along the ideas given here. The filament is preferably centrally located in the tube and the grid and plate electrodes disposed symmetrically about it in order to equalize the mechanical stresses on the filament resulting from electrostatic attractions or repulsions.

A type of construction quite widely used is illustrated in Fig. 138 and Plate 3 opposite page 154. This tube, made by the Western Electric Company, and known in the U. S. Signal Corps as its type VT-1 which was used principally as a detector and amplifier, has a V-shaped filament made of a strip of platinum covered with barium oxide, which gives a large emission of electrons at a dull red heat. The free emission at this low temperature

[1] S. Ballantine, *loc. cit.*

lengthens the life of the filament. The plate comprises two punched sheet steel elements mounted on either side of the filament and parallel to its plane. They fit around the glass stem at the bottom and are clamped at the top about a small cube of insulating material. The grid is also made of two punched sheet steel elements, supported between the plates and the filament by the lead-in wires at the bottom and by wires embedded in the small insulating cube at the top. The cube also supports a small spring wire hook which holds the V-shaped filament taut. The entire construction is quite rugged and will withstand considerable vibration. The terminals are brought out to four studs in the base of the tube, providing contacts for the grid, plate and two filament terminals. A special socket with four spring contacts is used in conjunction with this tube.[1] This tube operates with a filament voltage of 4 volts, a filament current of 1.1 amp., and a plate potential of 20 to 40 volts. The voltage amplification factor is about 7 and the internal plate resistance for a grid potential of zero volts is about 17,000 ohms.

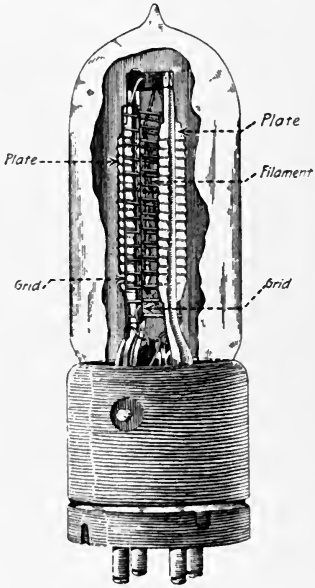


Fig. 138.

Another tube used very successfully by the Signal Corps, U. S. A., during the war was known as its type VT-2 which was designed for use primarily as an oscillation generator. In appearance it differs but little from the type VT-1, though the shape of the bulb is different and the spacing of the electrodes is altered as seen in Plate 3. It operates at a plate potential of 250 to 400 volts. The filament is heated by a 7-volt supply and takes a current of 1.36 amp. The tube has a minimum internal plate resistance of 5000 ohms and its voltage amplification factor is about 6 to 7.

Another common tube construction is illustrated in Plate 4 opposite page 155. It is made by the General Electric Company

[1] See *Electrical World*, vol. lxxiii, No. 8, 1919, p. 358.

and has a filament of uncoated tungsten, helically wound and mounted along the axis of a spiral grid and a cylindrical plate. The filament operates at 10 volts and 6.5 amp., and the normal plate potential is 700 to 800 volts. The minimum internal plate resistance is about 50,000 ohms, and the voltage amplification factor is about 10. The plate is mounted on four wire supports and is equipped with cooling flanges which radiate away the heat developed at the plate. The filament is operated at a yellow incandescent temperature. This tube is known in the U. S. Signal Corps as its type VT-18 and its use was that of fairly high power oscillation generation.

A tube[1] used extensively by the French during the last year of the World War, Plate 5 opposite page 172, has a straight tungsten filament about 0.83 in. long mounted along the axis of a spirally wound grid and a cylindrical plate. The grid is made of nickel wire about 0.01 in. in diameter, wound in the form of a spiral of 11 or 12 turns, 0.16 in. in diameter and 0.7 in. long. The plate is of sheet nickel, bent in the form of a cylinder 0.4 in. in diameter and 0.6 in. long. The operating filament voltage is 4 to 5 volts and the plate potential, 200 to 400 volts. The voltage amplification factor is about 10 and the minimum internal plate resistance about 24,000 ohms. The French used this tube for all three functions—detector, amplifier and oscillator—preferring to sacrifice some refinements of the tubes for special functions, for the sake of simplifying the supply and construction problems.

In order to have a tube with stable characteristics, unchanging during its life, it is essential that all air or other gases be exhausted from the tube, glass walls and the metal of the electrodes. This requires special precautions in the manufacture of the tubes, which are briefly outlined here. After the tube electrodes (filament, grid and plate) have been assembled and mounted on the glass stem of the bulb, the entire bulb is sealed to a fine glass tube connected to a vacuum pump, in the same manner that is followed in making ordinary incandescent lamps used for illumination purposes. Before exhausting the bulb, it is first heated in a special oven to a temperaure of about 500° C., and then is pumped out while maintained at that temperature. This is done to drive the gases out of the glass walls of

[1] *Revue Générale de l'Electricité*, April 26, 1919.

the tube. In order to heat the electrodes, an electric current is sent through the filament which is thus made incandescent, and a positive potential is applied to the plate and grid which are then heated by the bombardment of the electrons emitted by the filament. During all this time, the vacuum is gradually increased until it reaches a value of a few microns of mercury, when the bulb is finally sealed by melting off the fine glass tube connecting it to the vacuum pump.

CHAPTER VII

THE THREE-ELECTRODE VACUUM TUBE AS A DETECTOR

RECEPTION OF DAMPED WAVE SIGNALS

Two-Electrode Vacuum Tube.—An understanding of the use of the three-electrode vacuum tube as a detector of radio frequency damped oscillations is facilitated by first considering the application of the two-electrode tube or Fleming valve for the same purpose. This is illustrated in Fig. 139. The circuit, as may be seen, is exactly the same as the ordinary crystal detector receiving circuit for use with damped waves, except that the two-

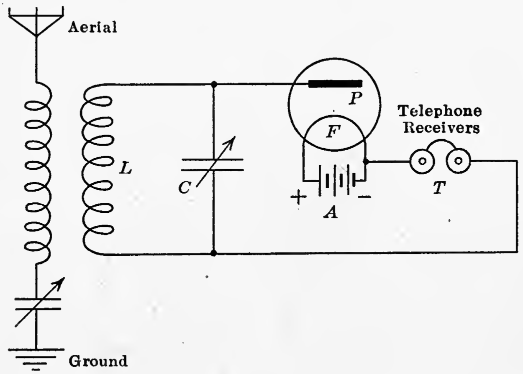


FIG. 139.

electrode tube with its filament heating battery is substituted for the crystal detector. There is no battery connected between the plate and filament, so that normally there is no electron flow from filament to plate, and therefore no current in the plate circuit *PLTFP*.[1]

[1] This is true only when the plate is connected to the negative extremity of the filament, as shown in Fig. 139. If connected to the positive end of the filament, the plate is positive with respect to the negative end of the filament, and will attract the electrons emitted by the more negative portion of the filament.

Now when a high frequency oscillation is made to energize the antenna circuit, a similar oscillation is set up in the tuned oscillatory circuit *LC*, as a result of which, an alternatng emf. is established across the condenser *C* and therefore also between the plate and filament of the tube. The plate is thus made alternately positive and negative, and a current will flow in the plate circuit and therefore in the telephone receiver, during those half cycles (and those only) when the plate is positive. The oscillatory current is thereby rectified, and damped or modulated wave oscillations will produce a sound in the telephone receiver in the same manner as when using a crystal detector. (See Chapter IV.)

Three-Electrode Vacuum Tube

Circuits Using No Grid Condenser.—The three-electrode vacuum tube may be used as a detector by taking advantage of the fact that the plate-current grid-voltage static characteristic curve is not a straight line, especially at its upper and lower ends.

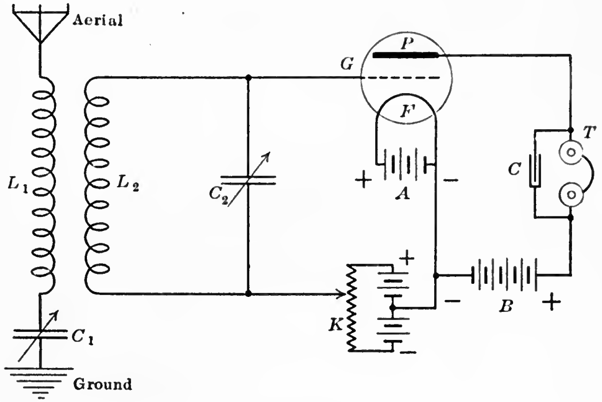


Fig. 140.

Thus, consider a three-electrode tube connected, as shown in Fig. 140, to a radio receiving circuit. The filament *F* is heated by the battery *A*. The plate circuit *FBTPF* comprises the telephone receiver *T* and is energized by the battery *B* which establishes a difference of potential E_p between the plate and filament of the tube, and maintains a current I_p in the plate circuit. The grid *G* is connected to the filament through the inductance

L_2, and a battery and potentiometer. By means of the latter, it is possible to adjust the grid potential to such a value E_g that, under the steady conditions pervailing when no oscillations are being received, the operating point of the tube is some point M or N of either bend of the static characteristc curve, Fig. 141. If the plate potential E_p furnished by the battery B is suitably chosen, it is possible to do away with the grid battery and potentiometer. Thus, referring to Fig. 134 it is seen that for a plate potential of 100 volts, the lower bend of the curve corresponds for the particular tube considered to a grid potential of about zero volts, which makes the use of a grid battery unnecessary.

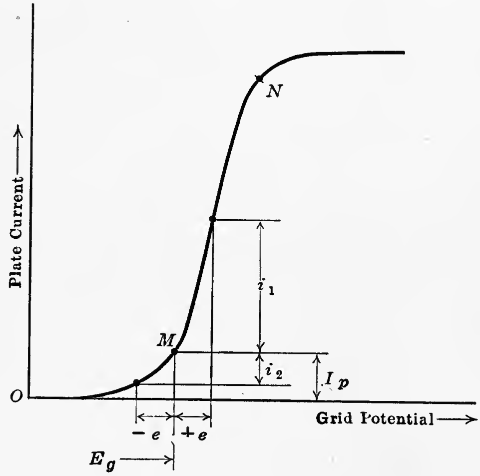


FIG. 141.

Referring then to Fig. 141, assume that conditions are such that when no oscillations are received, the plate potential E_p and grid potential E_g result in a steady current I_p in the plate circuit, the operating point M being on the lower bend of the curve. Now if an oscillation is received by the antenna, an oscillatory current will be set up in the circuit L_2C_2, Fig. 140, and an alternating emf. will appear at the plates of the condenser C_2. This alternating emf. is impressed between the filament and the grid of the tube, producing at every cycle approximately equal and opposite variations $\pm e$, of the steady grid potential E_g. Plotting these in Fig. 141, it is seen that the resulting variations i_1 and i_2 of the plate current I_p are unequal due to the bend of the characteristic curve. Thus, symmetrical variations of the grid voltage produce asymmetrical variations of the plate current, as a result of which rectification occurs.

This may perhaps be better understood by plotting the grid potential and plate currents against time. Referring to Fig. 142, the upper curve shows the symmetrical variations of the grid potential about its normal steady value E_g under the effect of the incoming oscillations. The resulting plate current is given for the various instantaneous values of grid potential by referring to the static characteristic curve of Fig. 141, and the

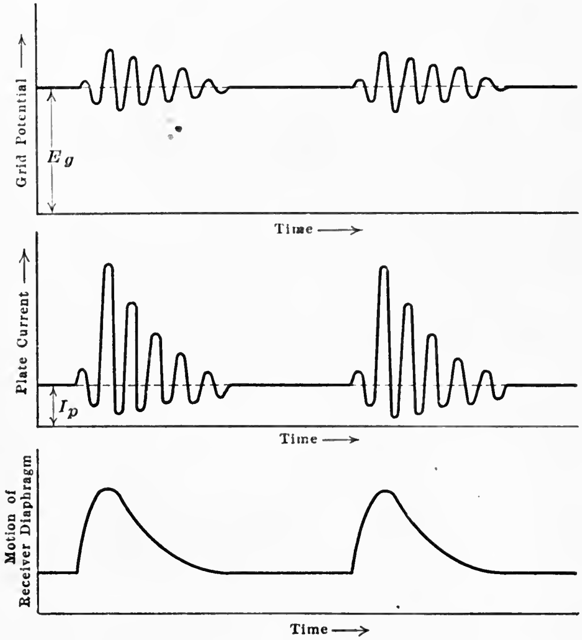


FIG. 142.

plate current curve of Fig. 142 thus obtained. It is seen that the symmetrical variations of grid potential produce asymmetrical variations of the plate current about its normal steady value I_p. The resulting deflection of the telephone receiver diaphragm is shown in the lower curve. As in the case of reception with the crystal detector, the operation of the telephones will be improved if they are shunted by a condenser C. (See Chapter IV.)

It is thus seen that by this method, detection of damped oscillations is possible The advantage of this method over the other type of detectors previously described is that whereas in other detectors the oscillatory electromotive force is acting directly, it is here acting on the grid, and its effect in the plate circuit is therefore multiplied by the amplification factor k of the tube. In other words, the signals are not merely detected, but are at the same time amplified very much. This makes the three-electrode vacuum tube the most sensitive detector available.

It is obvious that the operation of the tube about a point N of the upper bend of the curve is similar to that about the lower bend, except that the incoming oscillations produce at every cycle a large decrease and small increase of the plate current instead of a large increase and small decrease as in the case studied above. One reason which makes the operation about the lower bend a better practice is that the steady current I_p flowing in the plate circuit when no oscillations are being received is smaller than in the case of the upper bend, and the battery B furnishing this current is therefore not so rapidly exhausted. It should be noted that detection will not occur if the operating point of the tube is in the central part of the curve, which is practically a straight line and would produce equal variations above and below the normal value of I_p.

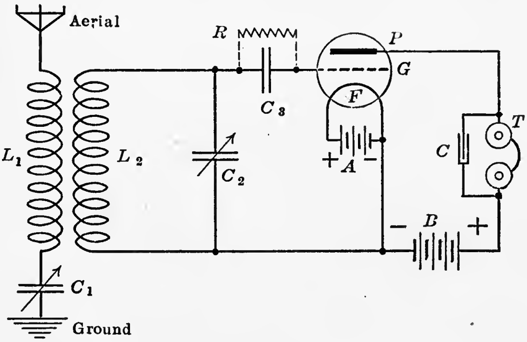


Fig. 143.

Circuits Using a Grid Condenser.—There is another method of using the three-electrode tube as a detector, which is illustrated in Fig. 143. The circuit is the same as that of Fig. 140 except that there is no battery in the grid circuit, and that a condenser C_3 is connected in series with the grid. With this circuit, when no oscillations are being received, the plate potential is E_p and a steady current I_p flows in the plate circuit under the action of the battery B. There is no current flowing in the grid circuit, since the grid is insulated from the filament by the condenser C_3. The grid then assumes the potential corresponding to zero current in the grid circuit. In the case of a tube free from all gas, as is assumed here,[1] this grid potential is zero. When the antenna circuit is then energized by incoming oscillations, an oscillatory current is set up in the circuit L_2C_2, and an alternately positive and negative emf. appears at the terminals of the condenser C_2. This alternately charges the grid positively and negatively through the condenser C_3. In the positive half-cycle, the grid

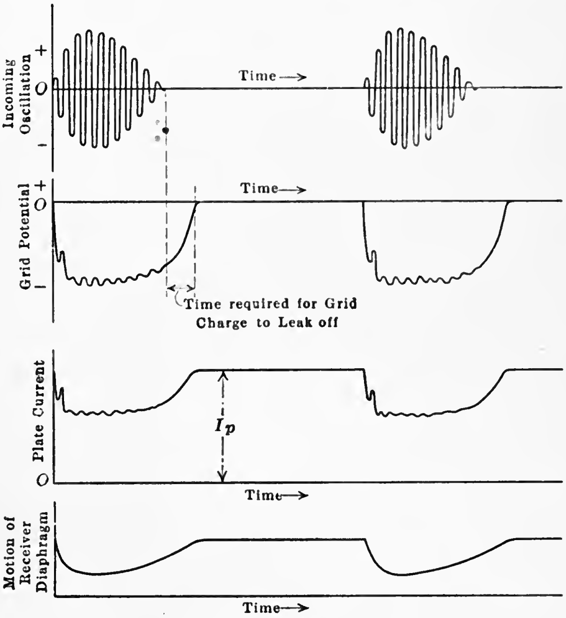


FIG. 144.

attracts some of the electrons present in the tube. In the negative half-cycle, however, it does not lose these electrons again, and a negative charge thus builds up on the grid at every cycle, the cumulative effect of which is to produce a decrease of plate current during the entire wave train. After the incoming damped oscillation has died out it is necessary to remove the negative charge from the grid in order to restore the original conditions for the arrival of the following wave train. In the case

[1] The case of a tube containing an appreciable amount of gas is fully studied in an article by Ralph Bown, *Physical Review*, vol. x, No. 3, September 1917, p. 253.

of a tube containing gas, this charge automatically leaks off through the gas of the tube. In case of a highly evacuated tube, it is necessary to shunt the condenser C_3 with a high resistance R of 1 to 5 megohms in order to provide a leakage path for the charge. The operation of the tube is illustrated in the curves of Fig. 144, which are self-explanatory.

This method of using a grid condenser has the advantage over the method previously explained that the tube will operate satisfactorily as a detector at any point of its characteristic curve. This therefore does away with the potentiometer adjustment of the grid and simplifies the operation. In fact, the operation of the tube as a detector with this method will be best when the operating point is at the center of the curve, where the greatest slope prevails. This may be understood by considering that an incoming oscillation will produce the accumulation of a certain negative charge on the grid, and therefore a certain decrease of the grid potential. Bearing the curve of Fig. 141 in mind, this decrease will evidently produce the greatest change of plate current at the point of maximum slope of the curve.

Amplification of Received Signals.—As was already pointed out, the three-electrode tube detector amplifies the signals due to the fact that the variations of grid potential produced by the incoming oscillations act in fact as variations of plate potential equal to k times the grid potential variations, k being the voltage amplification factor of the tube. This amplification is therefore an inherent property of the tube.

In the above discussion of the detector function of the tube, no account was taken of the effect of resistance, inductance and capacitance in the plate circuit. These have, however, a great importance and may be made to produce a pronounced amplification of the signals under suitable conditions. This additional amplification is thus a function of the circuit rather than a property of the tube itself. It will be studied in detail in the next chapter, which takes up the use of the tube for amplification purposes.

RECEPTION OF UNDAMPED WAVE SIGNALS

The reception of undamped wave signals was shown (Chapter V) to involve two operations: the generation of a local high frequency alternating current, and the rectification or detection

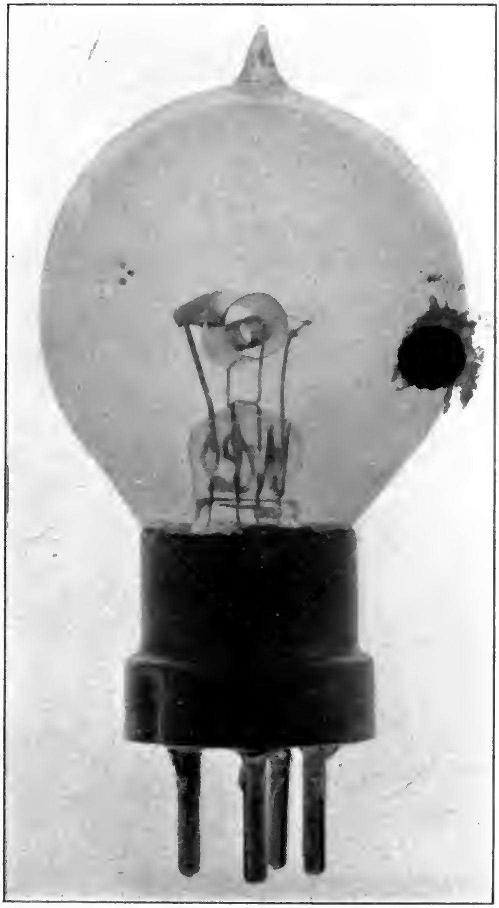
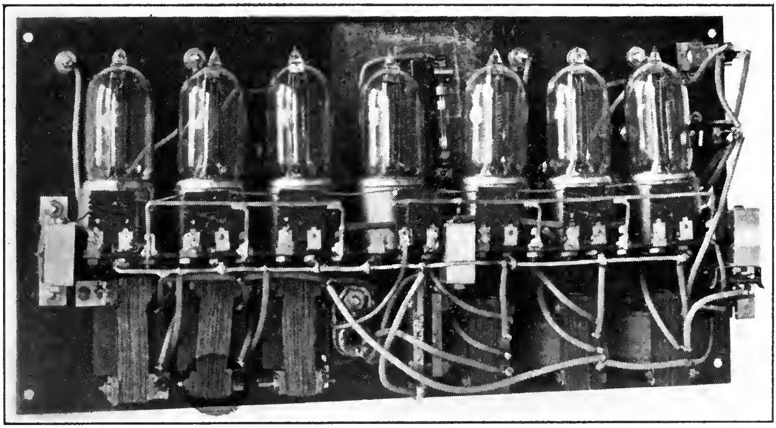
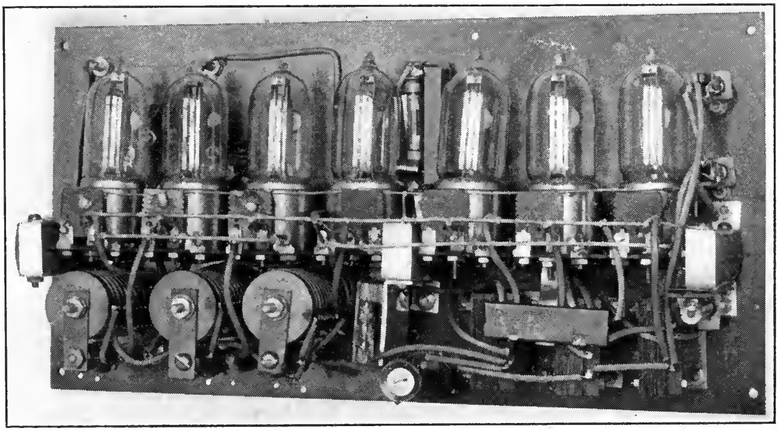


Plate 5.—Three-electrode vacuum tube used by the French Army. Note the cylindrical plate and helical grid with the filament axially mounted.

(*Facing page* 172.)



(*A*)



(*B*)

Plate 6.—Two three-electrode vacuum tube cascade amplifiers. From left to right the tubes are three radio frequency amplifiers, a detector and three audio frequency amplifiers. Above (*A*) note that the radio frequency amplifier tubes are coupled by special iron core transformers while below (*B*) air core transformers are used.

of the "beats." This latter function can be accomplished by the three-electrode vacuum tube, and is the same as the detection of damped wave signals. It will be shown later (Chapter IX) that the generation of the local undamped oscillations may also be done by means of a vacuum tube, and that *both functions may be accomplished by the same tube.* For this reason, the explanation of undamped wave reception by means of the tube is postponed to a later chapter. (See Chapter IX.)

CHAPTER VIII

THE THREE-ELECTRODE VACUUM TUBE AS AN AMPLIFIER

VOLTAGE AMPLIFICATION

As pointed out at the conclusion of the last chapter, the three-electrode vacuum tube may be used as an amplifier when connected to a circuit having suitable electrical constants. The amplifier function of the tube consists in reproducing on a larger scale in the plate or output circuit the variations impressed in the grid or input circuit. The fundamental property on which the amplifier action of the tube is based was previously mentioned in Chapter VI, but is more fully treated here.

Mathematical Theory.—The amplification property of the three-electrode vacuum tube may be stated as follows, for the particular case of alternating currents, which is the one generally encountered in radio work. If an alternating emf.

$$e_g = E \sin 2\pi ft$$

is impressed between the grid and the filament of a tube, a pulsating current will flow in the plate circuit, resulting from the superimposition of an alternating current upon the normal steady plate current. *This alternating current flowing in the plate circuit is the same as would flow if the external plate circuit of the tube were connected to an alternator* (instead of to the tube) *having an internal resistance equal to the internal plate resistance of the tube, and generating an alternating emf. equal to* ke_g, *k being the amplification factor of the tube.*

This important theorem may be demonstrated as follows.[1] Consider the circuit of Fig. 145. The filament F of a three-electrode tube is heated by a battery A. The plate P is connected to the filament through a battery B and a circuit having an impedance Z. This impedance Z will be assumed to include

[1] A more general demonstration may be found in Van der Bijl's paper mentioned in a previous note.

the internal resistance of the battery B, but not the internal plate resistance R of the tube. The grid G is connected to the filament through a battery C which determines the operating point of the tube on the static characteristic curve, and through an alternator D which impresses between the grid and the filament the emf. to be amplified.

In order to facilitate reference, the following letter symbols used are tabulated.

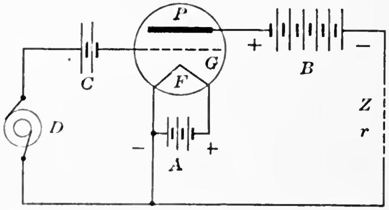


FIG. 145.

R = internal plate resistance of the tube.

E_g = instantaneous difference of potential between grid and filament.

E_p = instantaneous difference of potential between plate and filament.

e_g = instantaneous alternating grid emf.

E_B = plate battery emf.

E_C = grid battery emf.

I_p = instantaneous total current in the plate circuit.

I_B = direct-current component of the plate current.

i_p = alternating current component of the plate current.

Z = impedance of the external plate circuit.

r = resistance of the external plate circuit.

k = amplification factor of the tube.

It was seen in Chapter VI that the current in the plate circuit for various values of plate and grid emf's., as represented by the static characteristic of the tube, may be expressed by the relation

$$I_p = a(E_p + kE_g)^m$$

where m was seen to be equal to 2 or $3/2$. If however only the straight central portion of the characteristic curve is considered, as is generally the case for amplification, or if the operation extends only over a small part of the curve, then m may be made equal to unity, and the equation becomes

$$I_p = a(E_p + kE_g) \qquad (25)$$

and the internal plate resistance of the tube is

$$R = \frac{E_p + kE_g}{I_p}, \qquad (26)$$

so that the equation becomes

$$I_p = \frac{1}{R} E_p + \frac{k}{R} E_g \qquad (27)$$

On the other hand, from the above symbols, we have

$$\begin{cases} I_p = I_B + i_p \\ E_g = E_C + e_g \\ E_p = E_B - I_B r - i_p Z, \end{cases}$$

so that equation (27) above will give, for the total plate current and its direct current component, respectively, the two expressions,

$$\begin{cases} I_B + i_p = \frac{1}{R}(E_B - I_B r - i_p Z) + \frac{k}{R}(E_C + e_g) \\ I_B = \frac{1}{R}(E_B - I_B r) + \frac{k}{R} E_C \end{cases}$$

Subtracting the latter expression from the former, the alternating current component i_p of the plate current is obtained,

$$i_p = \frac{k}{R} e_g - \frac{1}{R} i_p Z \qquad (28)$$

from which

$$k e_g = R i_p + Z i_p$$

and finally

$$\frac{Z i_p}{e_g} = k \frac{Z}{R + Z} \qquad (29)$$

The first member of this equation is the ratio of the alternating emf. operating in the plate circuit to that impressed on the grid circuit. This ratio, as expressed by the the second member is seen to be less than the amplification factor k of the tube, but tends to reach this value k as the impedance Z is increased toward an infinitely great value. Thus, for infinite impedance in the plate circuit, the alternating component of the plate current is zero, but the voltage amplification is a maximum.

Physical Interpretation.—The above conclusions may be arrived at in a somewhat different manner, which is perhaps easier to understand. Assume first that the external plate circuit of Fig. 145 has no impedance or resistance, that is, that Z and r are equal to zero. Then, the plate potential E_p is always equal to the emf. E_B of the battery B, and the alternating grid potential

e_g produces synchronous variations of the plate current along the static characteristic curve of the tube corresponding to the plate voltage E_B. Thus, if the tube is operating at a plate potential of 200 volts, its characteristic curve will be the curve of Fig. 146 for a plate potential of 200 volts, and if the grid emf. is varied between − 20 and + 20 volts, the plate current will vary along the curve "200," from 3 to 67 milliamperes.

Suppose now, Fig. 145, that the plate circuit contains a resistance r. That is, to simplify the problem, suppose that the impedance Z of the external plate circuit is a resistance of value

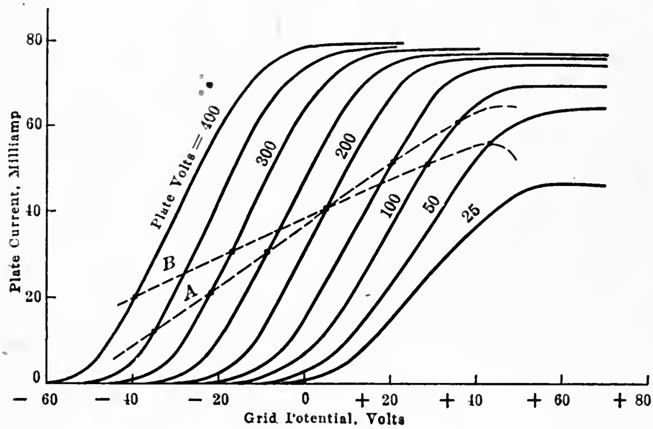


FIG. 146.

r. The emf. E_B of the battery B is then equal to the potential drop rI_p along the resistance r plus that, $I_pR = E_p$, between the filament F and the plate P of the tube, I_p being the current flowing in the plate circuit. Thus,

$$E_B = E_p + rI_p.$$

from which

$$E_p = E_B - rI_p \tag{30}$$

which expresses the difference of potential between the plate and filament of the tube.

If now the grid potential E_g is varied so as to increase the current I_p in the plate circuit, it is seen that the resistance drop rI_p in the external plate circuit will correspondingly increase. It then follows from equation (30) that, with the battery voltage E_B remaining constant, *the plate potential E_p will de-*

crease. Conversely, if the current I_p is decreased by decreasing the grid potential, the plate potential E_p will increase. Thus, on account of the external resistance of the plate circuit, the plate potential is no longer constant, but is a function of the plate current, and varies in opposite direction to the latter. In other words, the tube, operated under such conditions, has a *negative plate resistance reaction* as defined in connection with the oscillating arc, Chapter V. This will be taken up in greater detail in a later paragraph.

It follows from the above discussion, that when the external plate circuit has resistance and the grid potential is varied, the operating point of the tube no longer follows the static characteristic curve of the tube, which corresponds to a constant value of plate potential, but follows a different curve, called the *dynamic characteristic.* The shape of this characteristic curve depends on that of the static characteristic of the tube (and therefore on the tube construction and constants) and also on the resistance or impedance of the external plate circuit. This is shown in Fig. 146. The solid line curves are the static characteristic curves of the tube, while the dotted lines A and B represent the dynamic characteristics of the tube having an external plate circuit resistance of 5000 and 10,000 ohms, respectively. Thus, with an alternating grid potential and no resistance in the plate circuit, the tube follows the static characteristic curve corresponding to the potential E_B of the battery B. With an external plate circuit resistance of 5000 ohms, the operating point follows curve A, and with a resistance of 10,000 ohms, it follows curve B. Each of the curves A and B corresponds to a certain value of plate battery voltage. If this is changed, then the dynamic characteristic curve for a given external plate impedance will be shifted accordingly.

As expressed by the last equation, the greater the resistance r of the external plate circuit, or, more generally, the greater its impedance Z, the greater the variation of plate potential E_p resulting from a given change of plate current I_p brought about by a given variation of grid potential. This is shown in the curves of Fig. 146, where it may be seen that the slope of the dynamic characteristic curve decreases as the resistance (or impedance) of the external plate circuit is increased. As an extreme case, for infinite impedance, the curve would be parallel to the grid voltage axis of Fig. 146, showing that variations of grid potential would produce no variation of plate current, but maximum variations of plate potential. These would be *amplified variations of the grid (or input) potential variations.*

To illustrate this last statement, consider the circuit of Fig. 147, where the plate circuit comprises the battery B and a coil L of infinite, or say of very large impedance. When an alternating emf. is impressed between the grid and filament, by means of an alternator as shown here but which may be the emf. induced in a receiving antenna by incoming signals, an alternating current will flow in the plate circuit. Or rather, it will tend to flow, since any change in the current flowing through the coil L will induce in the latter a counter-emf. of such direction and magnitude as to effectively oppose the change of current. And, as was assumed, if this inductance L is large, the current in the plate circuit will remain practically unchanged, while the alternating counter-emf. induced across the coil L will be equal and opposite to the alternating emf. tending to produce the plate current variations, that is, equal and opposite to ke_g, k being the amplification factor of the tube and e_g the alternating emf. impressed on the grid. The emf. across the coil L is thus an amplified reproduction of the grid emf.

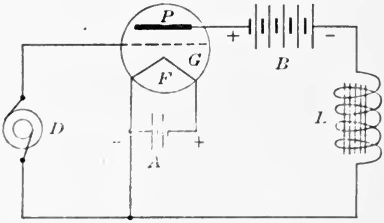


FIG. 147.

From what has been said previously, another possible arrangement would be to connect a large resistance in place of the coil L of Fig. 147, the operation of such a circuit being about the same as studied in the preceding discussion.

Adjustment for Maximum Voltage Amplification.—It was shown in the previous section that maximum voltage amplification occurs when the external plate circuit has infinite impedance. It is impossible of course to obtain this condition. The two methods of securing a high impedance illustrated above were by the use of a high resistance or of a high inductance in series with the plate circuit.

In the case using a high resistance, the d.c. potential drop across the resistance is so great that in order to impress a sufficiently high potential between the plate and filament of the tube, and obtain an appreciable current, it is necessary to use a very high battery voltage, which is generally undesirable in practice and frequently impossible to obtain.

In the case using an inductance, as shown in Fig. 147, since the amplification is due to the reactive or induced emf. of the coil, the resistance of the circuit may be reduced and made as low as desired without affecting the operation. It is then possible to obtain a large voltage amplification with a low resistance circuit and with a consequently small plate battery voltage.

It was shown at the beginning of this chapter that the external plate circuit would be under identical conditions if it were connected to an alternator having an internal resistance equal

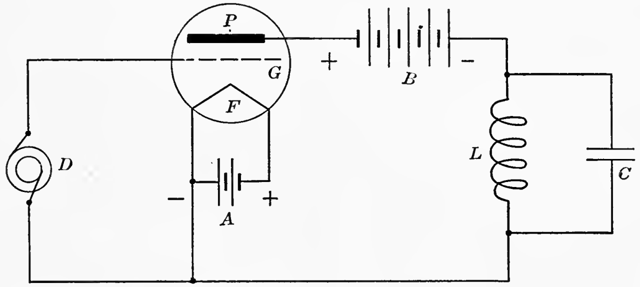


FIG. 148.

to that of the tube, and generating an alternating emf. ke_g. It was also shown in Chapter II that if the coil L is shunted by a condenser of capacitance C, such that the natural frequency of the oscillatory circuit is equal to the alternator frequency, this circuit will oppose infinite reactance to the current generated by the alternator.

In order to obtain maximum voltage amplification, it is then simply necessary, Fig. 148, to shunt the coil L by a condenser C of such capacitance that the natural frequency of the circuit LC will be equal to the frequency of the alternating emf. impressed between the grid and filament of the tube. The impedance of the external plate circuit will then be a maximum, and will be greater the smaller the resistance of this circuit.

It should be noted that if the capacitance of the condenser C is steadily increased, starting from a zero value, the voltage amplification will increase until the natural frequency of the circuit LC is equal to the input frequency, after which it will decrease again. At the same time, as may be understood by reference to Chapter II, the plate current through the tube will become a minimum when the conditions for maximum voltage amplification are obtained.

Power Amplification

It was shown that when an alternating emf. is impressed between the grid and filament of a tube, an alternating emf. and current are set up in the plate circuit and superimposed on the steady plate emf. and current due to the plate battery. The external plate circuit offers a certain impedance Z to this alternating current. Using the letter symbols previously established, the power expended in the plate circuit, is then equal to

$$W = \frac{(ke_g)^2 Z}{(R + Z)^2} \qquad (31)$$

If then the external impedance of the plate circuit is varied, the power output of the tube will be a maximum when $R = Z$, that is, when the internal plate resistance (or impedance) of the tube is equal to the external plate or output circuit resistance (or impedance). This is illustrated by the curve of Fig. 149 which has been plotted for the tube having the characteristic curves of Fig. 146. The maximum power output is obtained when the external plate circuit has an impedance of 5000 ohms, equal to the internal plate impedance of the tube at the direct current operating point of the characteristic.

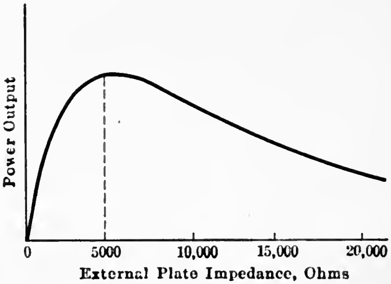


Fig. 149.

In order to adjust the circuit of Fig. 147 for maximum power amplification, the inductance of coil L may be varied, or the resistance of the circuit. Or else, the external plate circuit impedance may be left unchanged, and the internal plate resistance of the tube varied by adjusting the filament temperature or the plate potential, or by inserting in the grid circuit a battery which will supply to the grid a constant potential and bring the operating point of the characteristic curve to a point where the internal resistance of the tube is equal to that of the external plate circuit. It is not advisable, when using the tube for amplifying radio signals, to adjust the external impedance by means of a condenser shunting the coil L, since it would then be necessary to give to this oscillatory circuit a natural frequency different from that of the incoming signals.

AMPLIFIERS USED IN RADIO COMMUNICATION

In radio communication, the radio signals sent out by a transmitting station are made to set up alternating currents and emf's. in a tuned receiving antenna circuit, which is then coupled or connected to the detector and telephone receiver circuit. In many cases, the energy received in the antenna circuit is so extremely small that after being transferred to the telephone receivers it is too small to operate them properly and no sound, or only a very weak sound, is heard. It then becomes necessary to amplify the received signals. Although there are several types of mechanical amplifiers in use, they are not adaptable to radio frequency oscillations. The vacuum tube amplifiers, on the other hand, have been very successfully applied to this purpose.

It was pointed out in the preceding chapter that the three-electrode vacuum tube detector was inherently an amplifier of signals. From what has just been said in connection with power amplification, it may easily be understood that the plate circuit of the tube of Fig. 140, which comprises resistance and inductance, may be adjusted for maximum power amplification by a suitable choice of its constants. This amplification due to the detector tube is, however, frequently insufficient, and it has been found necessary to use several tubes in succession for effective amplifying purposes. These constitute so-called cascade amplifiers, which are studied below.

Cascade Amplification

Cascade amplifiers make use of the amplifying properties of the three-electrode tube, as studied above. It was shown that a tube may be used in two manners, namely, for voltage amplification and for power amplification. These two methods are utilized in the various types of cascade amplifiers.

Resistance Coupled Amplifiers.—The alternating emf. to be amplified will, for simplicity, be assumed here to be generated by a small alternator A, Fig. 150, although the explanation of the operation of the circuit would be exactly the same if the emf. to be amplified were supplied to the amplifier circuit by the receiving antenna and detector circuits. Thus, consider a three-electrode vacuum tube 1, Fig. 150, with the alternator A connected between its grid G_1 and filament F_1. The latter is heated by a battery, as explained before, but this is not shown here in order to simplify the circuit diagrams.

The plate circuit $P_1B_1R_1F_1P_1$ of this tube comprises the battery B_1 and the high resistance R_1 of say 100,000 ohms. This is similar to the circuit of Fig. 145. As was explained in

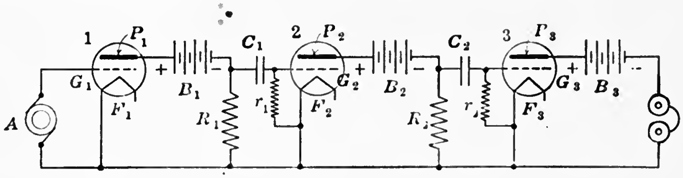


FIG. 150.

the previous paragraph, the small alternating emf. applied between the grid and filament by the alternator A is reproduced, amplified, across the resistance R_1, the ratio of this amplified emf. to the original input emf. being approximately equal to the amplification factor k of the tube 1.

This amplified emf. may then in turn be applied between the grid and filament of a second tube 2, connected up the same as tube 1, and can thus be itself amplified across the resistance R_2. In order to impress it upon the tube 2, the filament F_2 of this second tube may be connected to one end of the resistance R_1, and the grid G_2 to the other end. A condenser C_1 is interposed between the grid G_2 and the resistance R_1 in order to prevent the potential of battery B_1 from being impressed upon the grid G_2. The alternating emf. across R_1, which is an amplified reproduction of the alternator emf., is thus impressed between the grid and filament of the second tube 2. The presence of the condenser C_1 in the grid circuit of the latter necessitates that the grid G_2 be connected to the filament F_2 through a high resistance r_1, of the order of 1 to 5 megohms, in order to provide a leakage path for the negative charges which would otherwise accumulate on the grid, as explained in Chapter VII.

As a result of this arrangement, an alternating emf. is established across R_2, equal to k times the emf. across R_1, that is, equal to k^2e_g where e_g is the emf. input to the first tube, assuming that the successive tubes are identical and have equal voltage amplification factors.

In a similar manner, this emf. k^2e_g which may be thousands of times greater than e_g, is applied to a third, fourth, fifth, sixth and seventh tube if required. The last tube is connected, as shown here for the third tube, to a telephone receiver, and it is thus possible to amplify faint or inaudible signals to many times their received intensity. The total amplification ratio is equal to the product of the amplification ratio of the various individual amplifier steps.

The drawback to this type of amplifier is, as already mentioned, the necessity for high plate battery potentials. It is nevertheless useful, especially for amplifying very high frequency currents, for which purpose the inductance coupled amplifier described in the next section is often impracticable due to the distributed capacitance of the coils.

Choke Coil Coupled Amplifiers.—From the previous study of voltage amplification by means of an inductor or choke coil in the plate circuit of a three-electrode vacuum tube, the choke coil coupled amplifier of Fig. 151 may readily be understood.

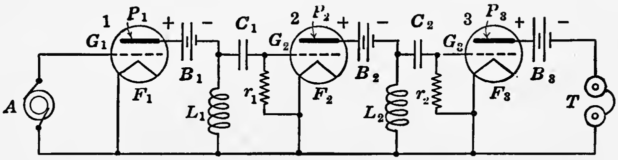


Fig. 151.

The alternating emf. to be amplified and assumed here to be generated by the alternator A is impressed between the grid and filament of the first vacuum tube. This causes pulsations of the plate current, with the consequent generation across the coil L_1 of an induced counter-emf., which is an amplified reproduction of the input grid emf. This has been fully explained in connection with Fig. 147. This emf. induced in the coil L_1 is impressed upon the grid of the second tube in a manner similar to that just explained in connection with the resistance coupled amplifier. The various steps of the amplifier are thus connected in cascade and the plate circuit of the last tube comprises the telephone receivers T for the reception of the amplified signals.

This choke coil coupled amplifier has some marked advantages. One already mentioned is that the external plate circuit of each tube may be made of low ohmic resistance, and thus there is only a minor amount of d.c. energy lost in the external plate circuit, and thus also the use of lower voltage plate batteries is permitted. Another and very great advantage, when using the amplifier for the reception of radio signals, is the use of condensers shunting the inductance coils L_1, L_2, L_3, . . ., in order to make the reactance of the plate circuits infinite and thus obtain maximum amplification, as previously explained. For, at the same time, this produces *tuned amplification* which greatly helps in eliminating interfering signals.[1] This is more fully treated in the following paragraph. This method is used mostly for radio frequency amplification rather than audio frequency amplification, in other words, it is used to amplify the received signal current before detection rather than after. The reason for this is that a rectifying detector such as a crystal or three-electrode vacuum tube operates best when the signals are of comparatively large

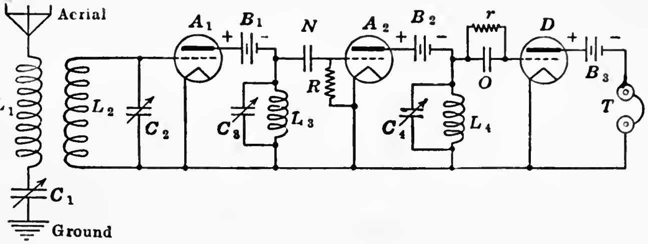


Fig. 152

amplitude, while they will rectify very poorly, almost negligibly, when the received signal oscillations are very weak. It is thus of advantage to amplify the received high frequency oscillations before rectifying them.

In order to thoroughly understand this question, consider the radio receiving circuit of Fig. 152. The incoming signals set up oscillations in the antenna circuit which is tuned to their frequency by means of the inductance L_1 and condenser C_1. Coupled to this antenna circuit is a tuned oscillatory circuit L_2C_2

[1] See Chapter IV on the use of successive coupled tuned circuits for reception of signals.

connected between the grid and filament of a three-electrode vacuum tube A_1, which is used to amplify the high frequency alternating emf. set up across the circuit L_2C_2 by the received oscillations. For this purpose, its plate circuit is adjusted for maximum voltage amplification, as explained in a previous paragraph. This is done by connecting in series with the plate battery B_1, a large inductance L_3. The reactance of the external plate circuit of the tube A_1 is then made infinite (condition for maximum voltage amplification) by shunting the coil L_3 by a condenser C_3 of such capacitance that the natural frequency of the circuit L_3C_3 is equal to the frequency of the alternating emf. to be amplified, that is, the frequency of the oscillations being received. Then when the oscillatory emf. is impressed on the grid of the tube A_1, an amplified oscillatory emf. is set up across the coil L_3 as previously explained. This emf. is then transferred to a second amplifier tube A_2, through the condenser N, as explained previously. The second amplifier tube is adjusted for maximum amplification by means of the circuit L_4C_4, similarly tuned to the frequency of the incoming oscillations. The oscillating emf. across L_4, which is a considerably amplified reproduction of the original emf., is then impressed upon the grid of the detector tube D, through the grid condenser O and grid leak r which acts as explained in Chapter VII. The function of the intermediate tuned circuits L_3C_3 and L_4C_4 is thus seen to be a double one; first, to obtain maximum voltage amplification; and second, to eliminate interfering signals of slightly different frequency, by the application of the principles of interference elimination established in Chapter IV.

It should be noted that the condensers C_3 and C_4 are not essential to the operation of the amplifier, provided the inductances L_3 and L_4 are large. (These should not be provided with iron cores, since excessive hysteresis losses would occur at radio frequencies.) Furthermore, for certain high frequencies, the resonant condition of the plate circuit of the amplifier tubes will be fulfilled automatically when the distributed capacitance of the coils L_3 and L_4, or the capacitance of the plate with respect to the filament of the tubes, is such that the natural frequency of the plate circuits is equal to that of the incoming signals.

Transformer Coupled Amplifiers.—A third kind of cascade amplifier is the transformer coupled amplifier illustrated in Fig. 153. The alternating emf. to be amplified is impressed between the grid and filament of a three-electrode vacuum tube A_1, either directly or through the medium of a step-up transformer M_1 of suitable ratio and impedance. This alternating grid emf. produces in the plate circuit of the tube A_1 a pulsating current which, flowing through the primary of the coupling transformer M_2, produces similar potential variations on the grid of the tube A_2. These potential variations are amplified reproductions of the grid potential variations of the tube A_1, provided the transformer M_2 is suitably designed, as explained below. The second tube A_2 can then be made to further amplify the original alternating current or emf. through the medium of a third transformer M_3 and tube A_3, and so on. In the plate circuit of the

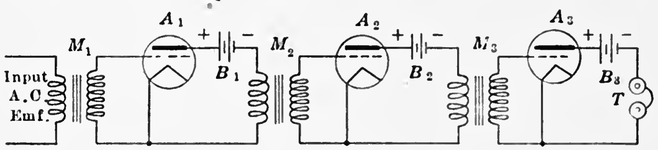


FIG. 153.

last tube is connected a telephone receiver T for the reception of the signals. This method of amplification is quite widely used, and is applicable to high (radio) or comparatively low (audio) frequency amplification. In the case of radio frequency amplification, the transformers M used for coupling successive tubes are built without iron cores, while for the amplification of currents of audio frequency they are constructed with finely laminated iron cores or even with iron wire cores. Certain amplifiers used by the Signal Corps, U. S. Army, have iron core transformers for radio frequency amplification, which seem to improve the stability and general operation of the amplifier. These, however, require the use of special enamelled soft steel laminations of a thickness of the order of one thousandth of an inch (.001 in.).

In order to build an efficient amplifier, it is seen from the above that the design of the transformers must be governed by the characteristics of the tubes to be used.[1] Thus, the plate circuit of each tube, which comprises the primary winding of a transformer, must be designed for maximum power amplification, while at the same time, the ratio of the transformer must be such as to deliver maximum potential to the grid circuit of the

[1] For description of several types of transformers used in vacuum tube amplifiers, see *Electrical World*, vol. lxxiii, No. 12, 1919, p. 568.

next tube. These conditions are best fulfilled when the transformer is given a step up ratio M such that Z_p and Z_g, respectively the impedances of the plate circuit of one tube and the grid-to-filament circuit of the next tube, will satisfy the equation

$$M^2 = \frac{Z_g}{Z_p}. \tag{32}$$

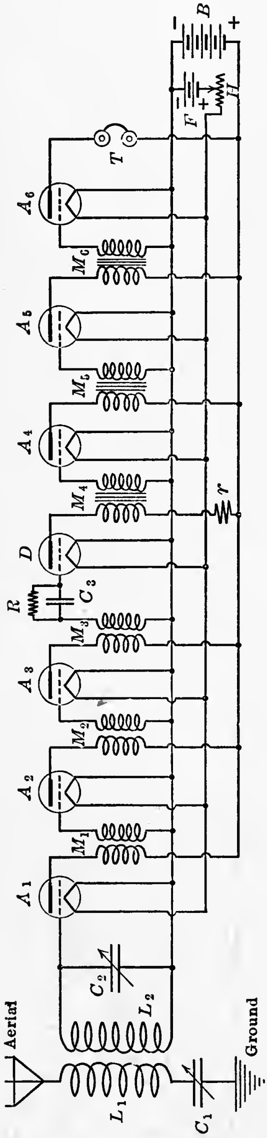


Fig. 154.

Properly designed amplifiers, that is, amplifiers having transformers suitable for the type of vacuum tube used, may be constructed which will considerably amplify signals which otherwise would be entirely inaudible. A Signal Corps amplifier, known as the type SCR-72, gives amplifications of from 40,000 to 60,000 times. It is only a two-stage amplifier, using two type VT–1 vacuum tubes, an input transformer and an intermediate coupling transformer.

As an illustration of a radio receiving circuit using vacuum tubes for high and low frequency amplification and for rectification or detection of signals, the amplifier of Fig. 154 is briefly described. This also represents a Signal Corps set[1] developed during the war and giving an output signal strength about 10^{14} times as great as would be obtained with a simple vacuum tube detector circuit. The circuit comprises a tuned antenna circuit L_1C_1 coupled to a tuned secondary circuit L_2C_2. The latter is connected between the grid and filament of a three-electrode vacuum

[1] The circuit described here is slightly different from the one actually used in the U. S. Army. Plate 6 opposite page 173 shows the interior of the Signal Corps amplifier referred to.

tube A_1, normally operating at the central portion of its static characteristic curve. That is, symmetrical variations of grid potential give rise to symmetrical variations of plate current. The incoming signals are thus successively amplified by the three tubes A_1, A_2 and A_3 coupled to each other by air core transformers or special radio frequency iron core transformers M_1 and M_2. The amplified radio signals are then impressed through the medium of a similar transformer M_3, between the grid and filament of a tube D which is made to operate as a rectifying detector by the use of a grid condenser C_3 shunted by a leak resistance R. The most favorable operating point of the characteristic curve is obtained by inserting a suitable resistance r in the plate circuit of this tube D. The rectified signals, that is, the audio frequency pulsations of the plate current of tube D are then amplified in succession by the tubes A_4, A_5 and A_6, and are finally sent through the telephone receivers T. The plate circuits of all the tubes are energized by the battery B, while all filaments are heated by a battery F. The amount of amplification may be varied by means of the rheostat H inserted in the filament circuit, with which it is possible to vary the filament current and therefore the filament temperature. This in turn varies the electron emission and the plate current of each tube.

This amplification is of such considerable magnitude that when suitably connected to a small loop antenna (see Chapter XI), very weak signals from a distant station may be reproduced with great strength.

It should be noted that the last described amplifier has probably the largest number of steps in cascade that can be used without very special precautions. The reason is that due to the successive couplings of plate and grid circuits, certain reactions take place between the circuits which lead to the generation of oscillations, as explained in the next chapter, and bring about disturbing noises known as howling, and resulting mostly from blocking of the tubes, that is, from a periodical interruption of the current in the plate circuits of the tubes.

Regenerative Amplification

A different application of the amplifier principle is used in the regenerative method of amplification, first conceived by E. H. Armstrong. This is illustrated in Fig. 155, which shows a radio receiving circuit using one three-electrode vacuum tube for detection and regenerative amplification. The tuned antenna circuit comprises the aerial, inductance L_1, variable condenser C_1, and the ground or counterpoise. Coupled to this circuit is a vacuum tube detector circuit, essentially the same as that of Fig. 143 except for the fact that coils L_3 and L_4 are inserted in the grid and plate circuits, respectively, and coupled to each other. An incoming damped oscillation sets up an oscillatory current in the antenna circuit L_1C_1 of the same frequency and general damping characteristics. Through the coupling M_1 of the antenna circuit and the secondary tuned circuit $C_2L_2L_3$, energy is transferred to the latter, setting up in it a damped oscillatory current. As explained in the previous section, the

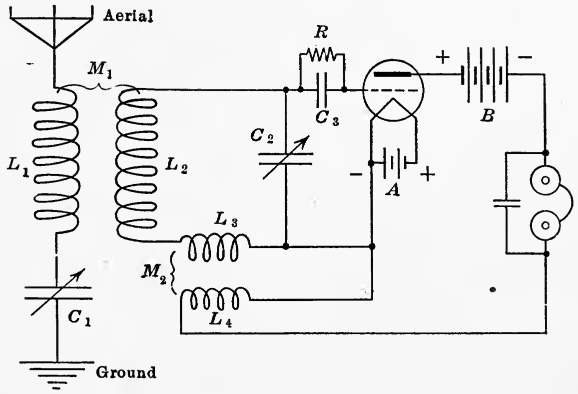


Fig. 155.

resulting alternating difference of potential between the filament and grid of the tube produces pulsations of the plate current at the same frequency as the oscillations of the circuit $C_2L_2L_3$. In other words, an alternating current is superimposed on the normally unvarying current in the plate circuit. This alternating current, flowing through the coil L_4, induces in the coil L_3 an emf. which, under conditions to be set forth later, is in phase with the oscillatory emf. operating in the circuit $C_2L_2L_3$. In other words, energy is synchronously supplied by the plate circuit to the oscillatory grid circuit, which partly compensates for the resistance losses in that circuit, and thereby increases the amplitude and decreases the damping of the oscillation. The duration and amplitude of every incoming wave train, as impressed upon the grid of the tube, are thus increased and the signals heard in the telephone receivers are correspondingly louder.

This elementary explanation does not fully represent the actual operation, but this will be better understood when reading the next chapter. It may already be understood that the regenerative action is primarily a function of the degree of coupling M_2 of the grid and plate coils L_3 and L_4. As this coupling is gradually made closer, more and more energy is transferred from the plate to the grid oscillatory circuit, in which the oscillations are correspondingly less and less damped. A value of coupling will then be reached when this damping becomes zero, and then the plate circuit furnishes to the circuit $C_2L_2L_3$ an amount of energy exactly equal to that dissipated or lost as heat in that circuit. The oscillations in the system are then no longer damped, and become undamped oscillations. Thus, when this critical value of coupling is reached, an oscillation started in the grid oscillatory circuit will persist indefinitely, the necessary energy being supplied by the battery B.

If the coupling is further increased, the plate circuit supplies more energy to the grid circuit than is lost in the latter, and the decrement becomes negative, that is, the oscillation increases in amplitude until the correspondingly increased losses exactly compensate the amount of energy supplied by the plate circuit. It is thus seen that the vacuum tube under such conditions may be made to sustain undamped oscillations in an oscillatory circuit.

The theory of the regenerative amplifier is thus closely related to that of the vacuum tube oscillator. For this reason it is taken up somewhat differently and more completely in the following chapter. The many methods and circuits for obtaining regenerative amplification will then be shown to be the same as those used for oscillation generation by means of the tube, with the exception that the degree of coupling of the grid and plate circuits must be reduced below the critical value at which the oscillations are sustained by the mutual reactions of the circuits.

It may be of interest to give a somewhat different interpretation of the regenerative method of amplification which links it more closely with the cascade method studied above, and more particularly with the circuits making use of double winding transformers.

In the cascade amplifier, the alternating emf. to be amplified is impressed between the filament and the grid of a three-electrode tube, in the plate circuit of which it produces an alternating current and an amplified reproduction of itself. This amplified emf. is then transferred from the plate circuit of the first tube to the grid circuit of the next tube through the medium of a coupling transformer. In the regenerative amplifier, the procedure differs in that the amplified reproduction of the original grid emf. is not handed to the grid circuit of a second tube, and so on, but is impressed upon the grid circuit of the first tube itself, through the same medium of a coupling transformer. The regenerative amplifier may thus be considered as a special case of the cascade amplifier, the cascade or chain being closed upon itself, so to speak. The design of the "feed back" transformer of a regenerative amplifier circuit, that is, the transformer coupling the plate and grid circuits of the tube, is then governed by the same fundamental principles indicated above for the cascade amplifier coupling coils.

A feature of the regenerative amplifier not possessed by the cascade amplifier is that the amplified or output oscillation has a smaller decrement than the original or input oscillation. This is due to the very fact that the cascade or amplifier chain is closed upon itself and hence that the output oscillation, which would otherwise be of the same character as the input oscillation, is made to react upon, reinforce, and thereby modify the latter.

An application of this principle whereby a vacuum tube amplifier may be used for regenerative action, will be found in the next chapter, certain oscillator circuits being derived in this manner.

Resistance Compensated Receiving Circuits

The various methods of selective reception of radio signals examined in the previous chapters of this book, whether damped or undamped, are based on the reactance tuning of the receiving circuit. That is, through the use of suitably chosen values of inductance and capacitance, the radio receiving circuit is made to have a natural frequency equal to the frequency of the signals to be received. The circuit then has zero reactance at that frequency and therefore minimum impedance, and the currents induced in it are a maximum.

As was shown previously (see Chapters II and IV) the current in the receiving circuit, tuned in this manner, is limited only by the resistance of the circuit, and is greater the smaller that resistance. The selectivity of the circuit, that is, its ability to respond to oscillations of a given frequency to a greater extent than to oscillations of any other frequency, is thus seen to be due solely to the variation of its reactance with the frequency, while the resistance is, practically speaking, approximately constant, and has merely a limiting action on the resonance effect.

It is possible by means of the three-electrode vacuum tube to also make the *resistance* of the receiving circuit as low as desired at one particular frequency and yet have this resistance assume its normally high value for all other frequencies.[1] This effect superimposed upon the selective resonance effect for zero reactance tuning, affords a circuit of almost perfect selectivity and almost entirely non-sensitive to oscillations of any frequency but the one for which it is adjusted.

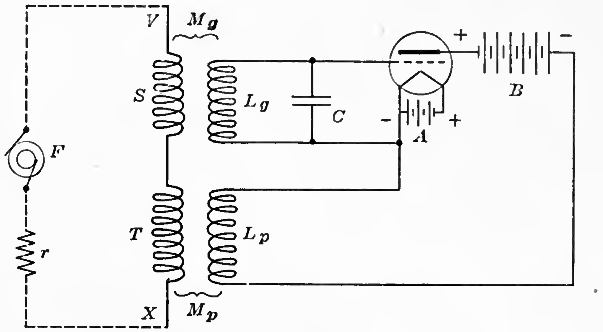


Fig. 156.

In order to fully understand this minimum resistance tuning method of reception it is first necessary to study the negative resistance reaction of the tube in somewhat greater detail than has been done before. Consider the circuit of Fig. 156 which represents a three-electrode vacuum tube having a grid circuit tuned to a frequency f by means of the circuit L_gC, and having its grid and plate circuits independently coupled to a conductor $VSTX$. If an alternating current of frequency F is then sent through the conductor VX, say by means of an alternator connected to its extremities, an alternating emf. of the same fre-

[1] *Revue Générale de l'Electricité*, Feb. 15, 1919, p. 270 abstract of French patents Nos. 485533, Sept. 12, 1916, and 20499, May 22, 1917, issued to M. I. Pupin and E. H. Armstrong.

quency will be induced across L_g and the grid of the tube, which in turn will set up a similar alternating current in the plate circuit and coil L_p. This induces an alternating emf. in the coil T. The phase of this last emf. with respect to the alternating current originally sent through the conductor VX is evidently dependent on the relative values of the frequencies F and f. On the other hand, its amplitude is determined by the amplification factor k of the tube, and the values of M_g and M_p.

Now the alternating current sent through the conductor VX by the alternator induces in coils S and T, counter-electromotive forces opposed to that of the alternator. The coils S and T thus have the effect of a resistance upon the current and hence may be said to have a positive resistance reaction on the current. Now if the emf. induced by the plate current of the tube in the conductor VX through the coupling M_p is in phase with the current in VX (and therefore 180° out of phase with the counter-electromotive forces in VX), and of an amplitude greater than that of the induced counter-electromotive forces of coils S and T, the resultant resistance reaction or effective resistance of the conductor $VSTX$ to the alternator current will be negative.

This may be checked up by connecting the conductor $VSTX$ coupled to the tube, as one arm of a Wheatstone bridge energized by a source of alternating current. The conductor $VSTX$ will then oppose the alternating current with an effective resistance which will vary with the frequency of the current energizing the bridge in a manner approximately as shown by the solid line curve of Fig. 157. In this curve, the ordinates represent the effective resistance of the conductor $VSTX$ as measured on the bridge, and the abscissæ represent the ratio of the frequency F of the current sent through the conductor VX and the natural frequency f of the tuned grid circuit of the tube.

Now if the conductor $VSTX$ is in series with an ohmic (positive) resistance r, Fig. 156, the total or resultant resistance reaction of the circuit will be equal to the algebraic sum of the resistance r and the resistance reaction of the conductor VX. It is thus seen that by a suitable adjustment of the natural frequency f of the tuned grid circuit of the tube it will be possible to make the ratio F/f of such a value that the resistance reaction of the conductor $VSTX$ will be negative at the operating frequency F of the alternator circuit, and the resistance of the circuit be thereby considerably reduced. It may even become zero if r and the resistance reaction of the conductor VX are equal and opposite.

An application of these properties to the reception of radio signals is illustrated in the diagram of Fig. 158. This represents an ordinary tuned antenna circuit having an aerial, inductance L_A, capacitance C_A, resistance R_A, and connected to the ground or counterpoise. In series with the circuit, however, a conductor $VSTX$ is connected, which may be given a negative resistance reaction for one certain frequency as explained above. The antenna circuit is coupled to a detector and telephone receiver

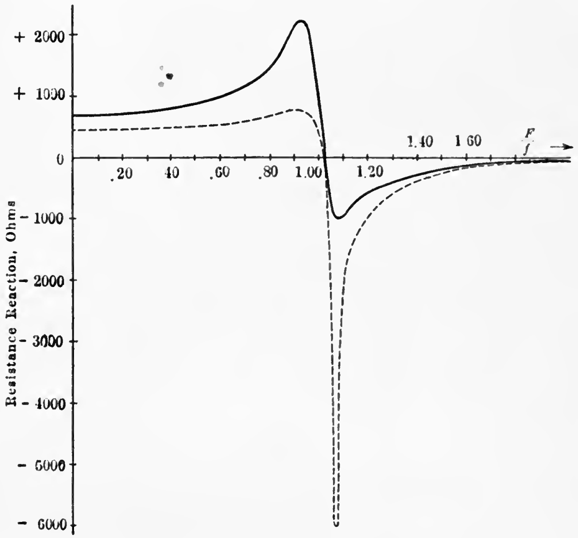


FIG. 157.

circuit in the usual manner. Now, if F is the frequency of the oscillations to be received, the antenna circuit is tuned to that frequency by means of its adjustable inductance and capacitance. The antenna circuit is thus made resonant for that frequency F, at which it will have zero reactance. On account of its resistance R_A, this circuit would, however, respond to a greater or less extent to oscillations of frequencies other than F. Also, at that frequency F, the current induced in the antenna would be limited. This was fully studied in previous chapters. By a suitable adjustment of the condenser C, however, it is possible, as may be understood from the above, to adjust the natural frequency f of the circuit L_gC to such a value that the conductor $VSTX$

will oppose a negative resistance to currents of the frequency F of the oscillations to be received. Under those conditions, and with suitable values of M_p, M_g and k, it may now be understood that the antenna circuit may be made to oppose only a very low resistance, or even zero resistance, to currents of the signal radio frequency F, due to the counteracting of its positive resistance reaction by the negative resistance reaction of the conductor VX, while it will have appreciable or even considerable resistance for all other frequencies, in addition to an appreciable reactance. The selectivity of the system may thus be increased many times over the ordinary reactance tuned circuit.

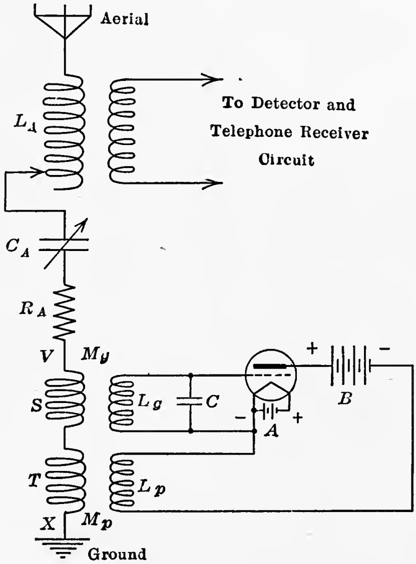


FIG. 158.

If the positive resistance of the antenna circuit has a rather high value, the negative resistance reaction of the conductor VX must be correspondingly large. This necessitates the use of a vacuum tube having a large amplification factor, or else, the use of some form of vacuum tube amplifier, connected either between the grid and the plate input and output circuits of the tube, or in cascade between the coils L_g, L_p and the coils S and T.

There is another method of amplifying the negative resistance reaction of the conductor VX, which is both simple and very effective, since the amplification may be infinite. This is illustrated in Fig. 159, which differs from the preceding circuit in that the conductor $VSTX$ is shunted by a resistance R_2 and is in series with a condenser D. This condenser is merely to compensate by its capacitive reactance the inductive reactance of the coils S and T, so that the reaction of the system ST will be that of a non-inductive resistance. Under these conditions,

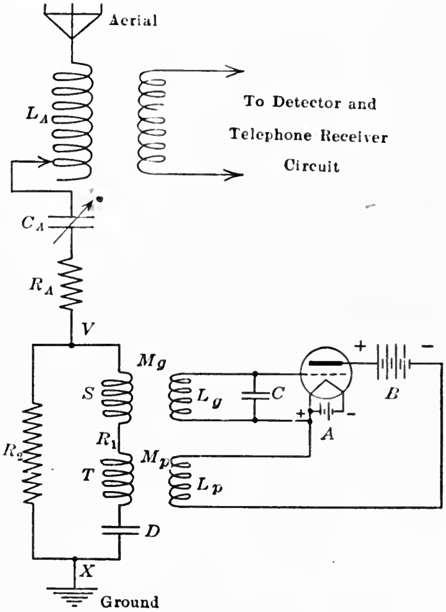


FIG. 159.

the resistance reaction R_1 of the system ST, and the resistance R_2 shunting it will give between points V and X a total resistance reaction equal to

$$R = \frac{R_1 R_2}{R_1 + R_2}.$$

If then the resistance R_1 of the conductor ST is adjusted as above explained to be negative for the frequency F of the signal waves to be received, the positive shunt resistance R_2 may be so adjusted as to have a value barely greater than the absolute value of the negative resistance R_1. The joint resistance R of the conductor VX will then be a negative resistance of extremely great absolute value. Thus, the dotted curve of Fig. 157 represents the resistance of the conductor VX for a shunt resistance R_2 equal to 1200 ohms. It is seen that a negative resistance of -6000 ohms is obtained at one value of frequency. If then the antenna circuit has a positive resistance R_A of $+6100$ ohms, it will have very considerable effective resistance for all frequencies except that one frequency F, when it will have a resistance of but 100 ohms. It is thus seen that the circuit may be made practically aperiodic and unresponsive to all frequencies outside a very narrow range about any desired value of frequency.

The enormous advantages of this method are evident, and it permits almost perfect elimination of interference from wave lengths other then the signal. It should be noted that the negative resistance of the conductor VX should never outweigh the positive resistance of the antenna, as the tube would then generate undamped oscillations in the antenna circuit, which of course is not the purpose sought here. The method just described may thus be considered as a special case of regenerative amplification.

CHAPTER IX

THE THREE-ELECTRODE VACUUM TUBE AS AN OSCILLATOR

It was pointed out at the conclusion of the preceding chapter that, under suitable conditions, the three-electrode vacuum tube may be made to sustain undamped oscillations in an oscillatory circuit by coupling the grid and plate circuits to each other and to some tuned oscillatory circuit. This property of transforming the direct current energy supplied to the plate circuit into an undamped alternating current of any desired frequency, as determined primarily by the constants of the coupled oscillatory circuit, has found wide and varied applications of great importance. It has permitted the construction and simple operation of undamped wave radio telegraph transmitting sets of light weight, small dimensions, low power, long range and sharp tuning. It has been adapted remarkably well to the heterodyne method of reception of undamped waves, thus increasing the importance and usefulness of the undamped wave transmitting sets just mentioned. Without the vacuum tube oscillator, radio telephony would hardly be possible under the conditions encountered on an airplane, and its use would be decidedly more limited than it is.

Finally, the vacuum tube oscillator has found a considerable number of applications in the laboratory, in multiplex wire telephony, and many other fields, where its qualities as a flexible and portable means of obtaining alternating currents of absolutely unvarying frequency, and of any desired frequency, are a fundamental requirement. This will be well understood when it is considered that undamped alternating currents of perfectly constant frequency and amplitude have thus been obtained with the greatest ease between the limits of about one-half cycle per second and 150 million cycles per second. Circuits for obtaining these extreme frequencies are described in later paragraphs of this chapter.

Although the preceding chapter gave some idea of the possibility and process of generating undamped oscillations by means of the vacuum tube, the subject will be taken up more thoroughly and from a different viewpoint in the following pages.

Theory of Oscillation Generation

General Principles.—The use of the three-electrode vacuum tube for setting up undamped oscillations in an oscillatory circuit is the application of the more general principle that any amplifying device having a non-linear characteristic may be made to sustain an undamped oscillation if disturbed from its condition of equilibrium under suitable coupling of its input and output circuits or members. If the device has a linear characteristic it is evident that a disturbance will not be transformed into a periodic oscillation, but will be indefinitely amplified.

Before considering the special case of the three-electrode tube, attention is called to the circuit of Fig. 160, which was studied in the preceding chapter. In this circuit, the alternator *A* impresses an alternating emf. *e* between the grid and filament of the tube, the effect of which in the plate circuit is equivalent to that of an alternating plate emf. *ke*, *k* being the amplification factor of the tube. If the circuit *LC* is assumed to have no resistance and is tuned to the frequency of the alternator *A*, then, as explained in Chapter II, the alternating emf. *ke* will set up an undamped alternating current in this circuit, without any expenditure of energy on the part of the battery *B*. In this case, the current which flows in the plate circuit *PBLFP* remains constant, as was explained in the previous chapter for pure voltage amplification. In practice, the resistance of the circuit is not zero, but is very small. An alternating current component then flows in the plate circuit, in synchronism with the grid emf., and 180 deg. out of phase with the alternating plate emf., due to the negative resistance reaction of the tube developed under such conditions. This property, although fundamental, is not explained here, having already been treated in detail in Chapter VIII. The pulsating plate current was also shown to follow a dynamic characteristic curve different from the static characteristic curve of the tube.

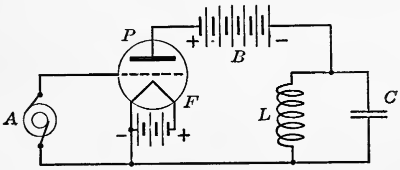


Fig. 160.

When the tube itself is used as the oscillation generator, the grid potential is no longer supplied by an external alternator, but is obtained by the coupling of the grid and plate circuits. Thus, the circuit of Fig. 161 differs from that of Fig. 160 simply in that the grid is no longer connected to the filament through an alternator A, or other external source of alternating emf., but is connected to the filament through an inductance coil L_g inductively coupled to the coil L of the plate oscillatory circuit. The functioning of this circuit was analyzed at the end of the previous chapter. It is briefly recalled here.

When an oscillation is in some manner started in the oscillatory circuit LC, the alternating current flowing in this coil L induces an alternating emf. e in the coil L_g, which is then impressed between the grid and filament of the tube. As just explained, the alternating grid emf. e has the same effect in the plate circuit as an imaginary alternating emf. ke impressed between the plate and filament, and therefore across the circuit LC. It is then seen that there are three possible cases, according to whether this emf. is equal to, greater than, or smaller than the emf. operating in the circuit LC.

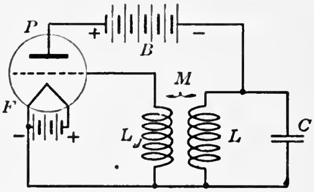


FIG. 161.

In the first case, if the emf. ke is equal to the oscillating emf. operating in the circuit LC, and in phase with it, the oscillation in that circuit will evidently be sustained or undamped.

If the emf. ke is greater than the emf. operating in the circuit LC, then the oscillation will grow in amplitude in that circuit. So will also the emf. e induced in the grid coil L_g, and the effect will be cumulative, the oscillation in the plate oscillatory circuit increasing until the grid voltage variations are such that the operating point of the tube describes the entire portion of the characteristic curve comprised between the upper and lower bends, when it stops increasing. This may be explained by the fact that a further increase of grid potential would no longer produce a corresponding or proportional increase of plate current, and this therefore sets a limit to the induction effect between the grid and plate coils.

If, finally, the emf. e induced in the grid circuit is such that its equivalent plate emf. ke is smaller than the emf. operating in the circuit LC, then the oscillation will not be sustained in that circuit, but will decrease in amplitude, and the induced grid emf. will decrease with it. The oscillation is thus damped, but the damping is less than that of a free oscillation of the circuit LC, since some energy is supplied by the plate battery, although insufficient to overcome the losses in the circuit.

It is thus seen that under suitable conditions, the three-electrode tube will set up undamped oscillations in the circuit LC. These conditions are that the plate emf. ke resulting from the various reactions just described and which is, in effect, the sustaining factor of the oscillations, must be of a sufficiently great amplitude to at least equal the operating emf. of the oscillating circuit, and that it must be in phase with the latter. The limitation is the grid-potential plate-current characteristic, and other structural constants of the tube.

The study of these conditions can best be accomplished by first establishing the characteristic curve of the entire circuit. As explained, the circuit of Fig. 161 differs from that of Fig. 160 in that the alternator A in the grid circuit is replaced by the grid coupling coil L_g. The characteristic curve of the system, that is, the curve giving the plate current as a function of the grid potential may then be derived from that of the circuit of Fig. 146 studied in the previous chapter, by introducing suitable factors to express the change made in the circuit to obtain the oscillator circuit of Fig. 161.

Thus, a property which was found to exist in the alternator excited circuit of Fig. 160 and which evidently holds true for the self-excited circuit of Fig. 161 is that the tube has an apparent negative resistance. That is, an alternating grid emf. sets up in the plate circuit an alternating current, which was shown to produce between the plate and filament a potential drop varying oppositely to the current. Thus, an increase of grid potential produces an increase of plate current and a decrease of plate potential. In other words, the grid potential and plate current vary synchronously, while the plate potential is 180 deg. out of phase. This having been fully treated in the previous chapter is merely mentioned here.

A feature which is peculiarly a property of the self-excited circuit of Fig. 161 is that the emf. impressed on the grid is the emf. induced in the coil L_g by the alternating current flowing in the coil L. The arrangement of the two coils L and L_g thus forms a sort of transformer, having a definite ratio depending on the mutual inductance of the coils. The variations of the grid and plate potentials are thus proportional to each other, and this may be expressed by the relation

$$\Delta E_p = n\,\Delta E_g$$

Furthermore, in order to make these variations of opposite directions or polarities, and obtain the required negative resistance effect of the tube, it is seen that the coil L_g must be so connected that the mutual inductance of the two coils is negative. This will be established mathematically in a later paragraph.

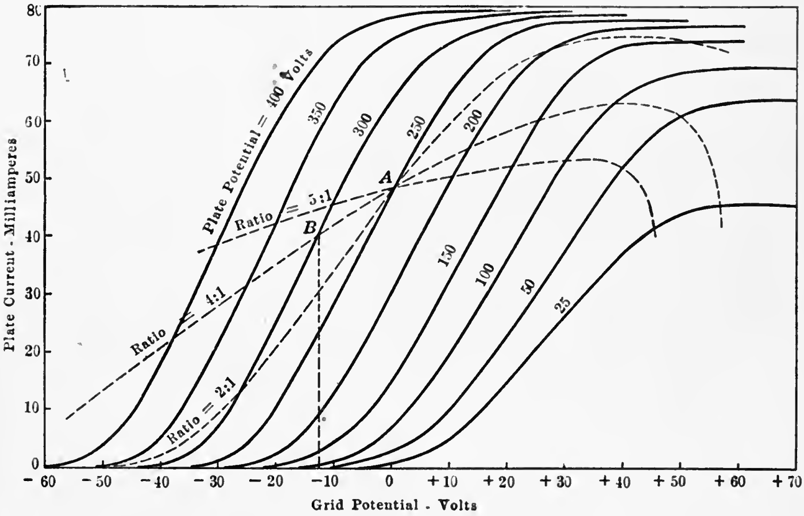


FIG. 162.

These remarks may be represented graphically in the form of a dynamic characteristic curve giving the plate current as a function of the grid potential. It should be remembered that this curve gives the *in-phase* variations of plate current and grid potential for a given tube used in a given circuit. Thus, consider the static characteristic curves of the tube itself, as given in solid lines, Fig. 162, and let A be the operating point of the tube at a certain instant of the oscillation cycle. During every cycle, the grid potential oscillates back and forth between two extreme limits. The operating point A does not, however, describe the static characteristic curve of the tube, since the plate potential varies proportionally with the grid potential, and in opposite directions.

To fix ideas, suppose that the ratio of plate to grid potential variation is 4 to 1. Then, as the grid potential decreases, the plate current decreases, while the plate potential rises, and the operating point A moves over to B along the dotted curve AB. The reverse variation of grid voltage brings about a reverse motion of the operating point. The dynamic characteristics for three values of the ratio have been shown in dotted lines in Fig. 162. Each of these curves corresponds to a certain value of coupling between the coils L and L_g. The effect of variable coupling is thus seen to be to vary the steepness of the curve. This is studied in the following paragraph.

Suppose the circuit of Fig. 161 is oscillating. If the coupling coefficient of the two coils L and L_g is gradually reduced, a similar grid potential variation induced in the coil L_g will require a gradually increasing alternating current in the coil L. In other words,

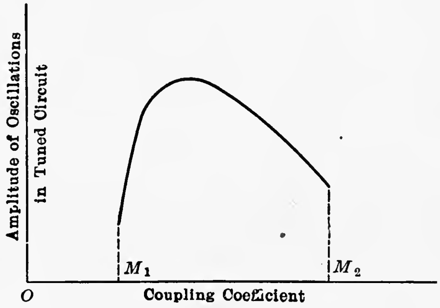


FIG. 163.

the looser the grid plate coupling, the stronger the oscillation necessary for a given grid potential variation. The oscillation thus increases with decreasing coupling coefficient until the plate current varies over the entire portion of the curve comprised between the upper and lower bends. A further decrease in coupling cannot then increase the plate current, and the emf. induced in the grid coil L_g therefore decreases with the coupling until a value of coupling is obtained where the oscillations stop entirely.

Conversely, if the coupling is gradually increased, starting at zero, a value is reached when the oscillations suddenly begin. They then increase with increasing coupling, reach a maximum, decrease, and finally stop when the coupling is made too tight. This is represented in Fig. 163.

It should be noted that the oscillations start and stop abruptly for two values of coupling. There is no transition value of coupling corresponding to very weak oscillations, the reason for this being best explained in the mathematical theory given below.

There are many different arrangements of circuits which will permit the use of the three-electrode tube as an oscillation generator, but it is considered advisable first to study mathematically the conditions which must be met by the circuit constants to permit oscillation generation. This is done in the following section.

Mathematical Theory of Oscillation Generation.—While the above discussion may give a qualitative idea of the phenomena involved in oscillation generation by means of the vacuum tube, it is essential to express these phenomena in mathematical terms when designing oscillator circuits. The purpose of the following paragraph is to establish such equations for the more important phenomena involved. Several methods have been followed by various authors.[1] The method given here is believed to be particularly helpful in gaining an understanding of the processes involved. It makes use of the idea of negative resistance, which is already familiar to the reader.

It was explained in the previous paragraphs of this chapter that when a three-electrode vacuum tube is made to sustain undamped oscillations in an oscillatory circuit, by establishing a suitable value of the grid-plate coupling, its alternating current filament-to-plate internal resistance is negative. That is to say, its plate current and plate potential vary in opposite directions, the one increasing while the other decreases, contrary to what happens in an ordinary wire resistance, where the current variations are directly proportional to and in phase with the voltage variations. Under these conditions, the plate circuit of Fig. 161, omitting the grid circuit including the coil L_g, bears a great similarity to that of Fig. 113, to which it is, in fact, equivalent. Referring to the mathematical theory given on pages 131, 132 and 133, in connection with this circuit, it was shown that undamped oscillations will take place in the oscillatory circuit LC

[1] Vallauri, *Electrotechnica*, 1917, Nos. 3 and 4.
Béthenod, *La Lumière Electrique*, Oct. 14, 1916.
Appleton, *The Electrician*, Dec. 27, 1918.
Hazeltine, *Proc. Inst. Radio Engineers*, 1918.
Gutton, *Revue Génerale de l'Electricité*, July 5, 1919.

when this is connected to a negative resistance, the absolute value of which is less than $\frac{L}{RC}$, where L, R and C are the constants of the oscillatory circuit.

Conditions for Oscillation Generation.—In order to mathematically express the conditions under which the circuit of Fig. 161 will oscillate continuously, it is then necessary to state that the filament-to-plate alternating current internal resistance r of the tube is negative under the existing grid coupling conditions, and that the absolute value of this resistance is less than $\frac{L}{RC}$.

The data available for this computation are the constants of the oscillatory plate circuit, the constants of the vacuum tube under the conditions of use, and the value of grid-to-plate coupling. Thus, let R, L and C be the resistance, inductance and capacitance of the oscillatory circuit, M the value of grid coupling, R_1 the internal plate resistance of the tube and k its amplification factor, the last two constants being obtained from the static characteristic curve of the tube corresponding to the values of filament and plate direct current potentials used.

Now, when there is an oscillatory current in the circuit, an alternating current of constant maximum amplitude is set up in the oscillatory circuit LC, and an alternating emf. E is established across the condenser C and coil L, which may be expressed by the relation

$$E = E_0 \cos \frac{t}{\sqrt{LC}}$$

where E_0 is the maximum value of this emf., and t is the time. This equation also shows that the frequency of this emf. is equal to the natural frequency of the oscillatory circuit.[1] Neglecting the resistance of the battery B and the connecting wires, it is seen that this alternating emf. is also impressed between the filament and plate of the tube, producing in the latter an alternating current superimposed upon the steady plate current of the battery B and given by the relation

$$i_1 = \frac{E}{R_1} = \frac{1}{R_1} E_0 \cos \frac{t}{\sqrt{LC}}$$

[1] This is only a first approximation which is however permissible here, the main purpose being to gain an understanding of the process involved.

Now, the alternating current i_L which flows in the coil L induces in the grid coil L_g an alternating emf. e_g given by the equation

$$e_g = -M\frac{di_L}{dt} = \frac{M}{L}E = \frac{M}{L}E_0 \cos\frac{t}{\sqrt{LC}}$$

the current in the grid circuit being considered as negligible.[1] This alternating emf. e_g is impressed between the grid and filament of the tube and it was shown in Chapters VI and VIII to have the same effect on the plate current as an alternating plate emf. equal to ke_g. This emf. will then produce in the plate circuit an alternating current i_2 superimposed upon the other currents in that circuit and equal to

$$i_2 = \frac{ke_g}{R_1} = \frac{k}{R_1}\cdot\frac{M}{L}E_0 \cos\frac{t}{\sqrt{LC}}.$$

The total current I flowing in the plate circuit is then equal to the algebraic sum of the steady plate current I_p due to the battery B, and the two alternating currents i_1 and i_2, thus,

$$I = I_p + i_1 + i_2$$

The last two terms represent the total or resultant alternating plate current i, which is the only quantity of interest in the present discussion. This is equal to

$$\begin{aligned} i &= i_1 + i_2 \\ &= \frac{1}{R_1}E + \frac{kM}{R_1L}E \\ &= \frac{1}{R_1}\left(1 + k\frac{M}{L}\right)E \end{aligned} \qquad (33)$$

From this, the alternating current internal plate resistance of the tube is then, in absolute value,

$$r = \frac{E}{i} = \frac{R_1}{1 + k\dfrac{M}{L}} \qquad (34)$$

And since the condition for oscillation generation is that this re-

[1] A study of the effect of grid current on the functioning of the tube may be found in a note by E. V. Appleton, *Phil. Mag.*, January, 1919, p. 129.

sistance shall be negative and of an absolute value smaller than $\frac{L}{RC}$, the condition to be satisfied is seen to be

$$r < -\frac{L}{RC}$$

or

$$\frac{R_1}{1 + k\frac{M}{L}} < -\frac{L}{RC} \qquad (35)$$

from which

$$M < -\frac{L + CRR_1}{k}$$

All the terms of the right hand member being positive, it is seen that the condition requires that the grid-to-plate coupling be made negative. This condition was also found from the previous qualitative discussion.

Another property already pointed out, namely the variation of the amplitude of the oscillatory current with the coupling M, may also be found from a consideration of equation (33) above, which expresses the value of the alternating plate current i. Thus, when M is decreased in absolute value, as occurs when the grid coupling is made looser, the quantity in parenthesis increases, since M is negative, and the current i therefore increases. This was fully discussed previously and is merely called to attention here.

Some interesting properties may be found by solving equation (35) with respect to the internal direct current plate resistance R_1 of the tube. Thus, the relation is obtained,

$$R_1 < -\frac{L + kM}{RC}$$

and since M must be negative, it is seen that the tube will generate oscillations if

$$R_1 < \frac{L - kM}{RC}$$

Now, it was shown in Chapter VI that the internal resistance of the tube, that is, the ratio of plate potential to plate current, depends to quite a great extent upon the filament temperature. It is thus seen that, with a given oscillatory circuit and a given value of grid coupling, the circuit will oscillate continuously only if the direct current internal resistance R_1 of the tube is below a certain value. This value may be altered by adjusting the filament temperature, through a suitable adjustment of filament current. Suppose then that the circuit of Fig. 161 is connected up with certain fixed values of inductance, capacitance and resistance in the oscillatory circuit, and with a negative grid coupling of suitable value, so that the only condition lacking for oscillation generation is a sufficiently low internal tube resistance. Starting with the filament cold (battery A disconnected), the resistance R_1 of the tube is infinite, and there is no plate current and no oscillation generation. If the filament temperature is gradually increased, by increasing the filament current, there will be a gradually increasing plate current, as explained in

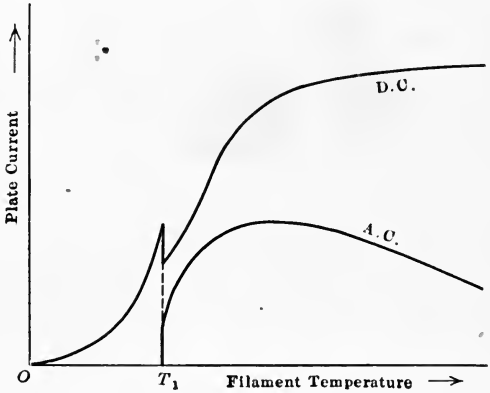


FIG. 164.

Chapter VI, with a corresponding decrease of the internal plate resistance R_1 of the tube. This is represented graphically by the curve "D.C." of Fig. 164.

A certain temperature T_1 will be reached when this internal resistance R_1 is low enough to satisfy the above condition, and oscillation generation will begin in the circuit. This is shown by the curve marked "A.C.," Fig. 164, which represents the alternating current in the circuit. It should be noted that equation (33) above gives values for the alternating current for all values of R_1, irrespective of whether or not the resistance R_1 satisfies the condition for oscillation generation. The values corresponding to a resistance too great for oscillation generation are therefore meaningless. But it follows that as soon as the resistance R_1 comes within the oscillation limits, the alternating current, as given by the equation, will have a certain definite value which may be quite large. The alternating current, which was zero up to the temperature T_1, is thus seen to *suddenly* pass from zero to a certain value, after which it increases gradually with the filament temperature. The sudden appearance of an alternating plate current is also accompanied by a decrease in the direct current component of the plate current, as shown by the break in the D.C. curve. This is due to the fact that the battery B which up to the temperature T_1 supplied the d.c. plate current only, is now made to supply also the a.c. losses of the oscillating circuit. As the filament temperature is further increased, the tube amplification factor k varies, the characteristic curves are shifted, and the alternating current does not increase indefinitely as might be assumed from equation (33). It is thus seen to reach a certain maximum value, beyond which it slowly decreases again.

Finally, a last conclusion which can be drawn from the above equations is the effect of the resistance R of the oscillatory circuit. Taking into account the fact that M is negative, equation (35) shows that the condition for oscillation generation is that

$$R < \frac{L - kM}{CR_1},$$

M being in this relation the absolute value of the mutual inductance of the grid and plate coils. Now since the function of the vacuum tube in the circuit considered is to maintain in the oscillatory circuit an undamped oscillation by synchronously replenishing the energy losses in that circuit due to its resistance,[1] and since these losses are directly proportional to the resistance, this condition may be interpreted to mean that the vacuum tube can furnish to the oscillatory circuits only a limited amount of power, which, for a given external tube circuit, is determined by the constants R_1 and k of the tube under its conditions of use. In other words, there is a maximum limit to the power output of the tube. To give a concrete example, the type VT-2 tube described in Chapter VI will have under ordinary conditions, an a.c. power output of not more than about 4 watts. The type VT-18 tube may have an output as high as 40 watts. It is thus seen that, if the oscillatory circuit has a resistance too high for the tube

[1] This resistance is the sum of the ohmic resistance, radiation resistance, etc., of the circuit.

used, the latter will not be capable of maintaining an undamped oscillation in the circuit.

Effect of Grid Potential on Generated Oscillations.—In the above discussion, no mention was made as to whether or not a battery was connected in the grid circuit. It was assumed that there was no such battery, the d.c. grid voltage being then zero. Consider the dynamic characteristic curve in Fig. 165 for the tube as connected in Fig. 161. When no grid battery is used, the d.c. grid potential is zero, and the operating point of the tube when not oscillating is at A. If an oscillation is started, as for instance by slightly disturbing the grid potential, the point A will describe the dynamic characteristic, and will continuously oscillate

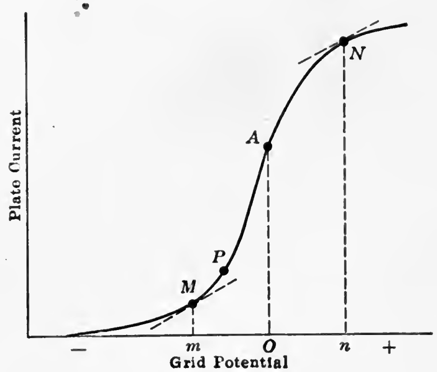


Fig. 165.

between the two points M and N at which the slope of the curve corresponds to a value of negative resistance of the tube, beyond the limiting conditions established above. That is, beyond these points M and N, the variation of plate current for a given variation of grid voltage, is, as shown by the slope of the curve, too small to maintain the oscillations through the mutual reactions in the circuit.

For a given circuit, it is then essential that the grid be given a potential between the limits m and n. If a battery is then connected in series with the grid, oscillations will be generated, provided the d.c. grid potential furnished by the battery is between the limits m and n.

If a battery is now inserted in the circuit of Fig. 161 between the grid and filament, it is possible for suitable values of grid, plate and filament battery voltages to bring the d.c. operating point of the tube to some point P, Fig. 165, of the lower part of the dynamic characteristic curve. As seen in Fig. 162, this corresponds to a higher plate battery potential and a lower continous plate current than obtains for point A. It follows that while point P is still within the oscillating range MN of the tube, the oscillations will be weaker due to the higher internal plate resistance R_1. In other words, the oscillations in a given circuit will be maximum for the d.c. grid potential which brings the operating point of the tube on the point of maximum slope of one of the static characteristic curves of the tube. This position of the operating point will be at the intersection of the oscillator dynamic characteristic curve with that static

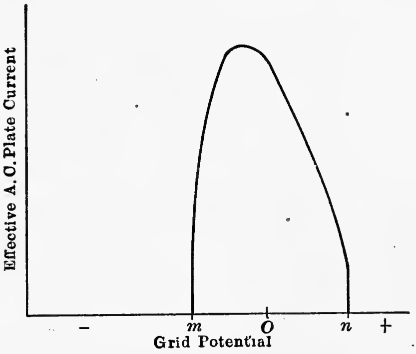


FIG. 166.

characteristic curve for which the internal d.c. plate resistance is a minimum, or the slope a maximum. For a tube and circuit having a dynamic characteristic as shown in Fig. 165, the effective value of the undamped oscillations generated under various d.c. grid potentials is shown approximately in Fig. 166. In these two figures, the grid potentials only have been plotted at the same scale.

Regenerative and Absorbing Functions.—As a result of the preceding mathematical discussion of oscillation generation, and further analyzing the relations of the oscillating vacuum tube circuits considered above, the important regenerative and absorbing properties of the vacuum tube may be set forth.

When the circuit of Fig. 161 has been set to oscillating, the alternating current flowing in the circuit LC induces in the coil L a counter-electromotive force, equal and opposite to that producing the oscillation, and therefore tending to prevent further oscillation of the circuit. This action was explained in Chapter I. Now the emf. induced by the same current in the grid coil L_g is (in that coil) of the same polarity or direction as the counter-emf. in coil L, and proportional to the latter and to the coefficient of mutual induction M. If then the coil L_g is connected to the grid and filament with its terminals reversed, that is, if the coefficient M is made negative, then the grid emf. e_g induced in coil L_g by the current in coil L, and its equivalent plate emf. ke_g will be of a direction opposite to that of the counter-emf. in coil L. The effect of the grid emf. e_g on the plate oscillatory circuit will then be opposite to that of the induced counter-emf. in coil L and will counteract the latter's tendency to stop or dampen out the oscillation. In other words, and as was already explained, when the coefficient M is negative, the oscillation is undamped, or, at least, its damping is lessened.

If, however, the terminals of the coil L_g are so connected that M is positive, then the grid emf. e_g and its equivalent plate emf. ke_g will be in the same direction as the counter-emf. induced in coil L, and will strengthen the latter's effect, so that the damping of the oscillation in the circuit LC will be increased.

In other words, depending on whether the grid-to-plate coupling is negative or positive, or, which is equivalent, depending on whether the internal a.c. resistance of the tube is negative or positive, the vacuum tube will supply energy to the oscillatory circuit, or absorb energy from it, and correspondingly sustain or dampen out the oscillation started in the oscillatory circuit. It is thus seen that if such a circuit is used in connection with a radio receiving circuit, it may be made to increase the incoming oscillations, as in the case of regenerative amplification studied in the previous chapter, or it may be used to absorb the energy of an interfering oscillation.

Vacuum Tube Oscillator Circuits

In the previous discussion, the conditions were established which must be satisfied if a vacuum tube circuit is to sustain undamped oscillations. Qualitatively speaking, it is simply necessary to couple the plate and grid circuits of the tube to a common oscillatory circuit (or to some arrangement of coupled oscillatory circuits), and in such a manner that the resulting grid-to-plate coupling will be negative. The number of circuits which answer these requirements is very large, and the actual quantitative relations are special almost in each instance. They are, however, in general, very easily obtained in a manner similar to that followed in the previous sections for the particular circuit considered. Despite their many differences, all these vacuum tube oscillator circuits may be shown to be derived from one of a comparatively small number of so to speak fundamental or elementary or generic circuits. These will be briefly described and discussed here.

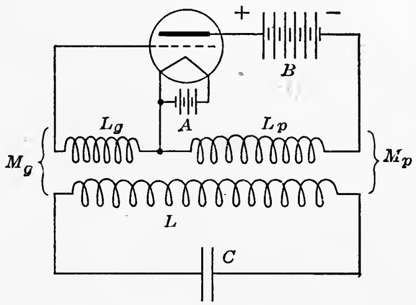


FIG. 167.

The first fundamental circuit is that of Fig. 167, where the plate and grid circuits of the tube are independently inductively coupled to the coil L of an oscillatory circuit LC, through coils L_p and L_g inserted in those circuits, M_p and M_g being the respective mutual inductances. The operation of this circuit, which from the previous paragraphs should not require any explanation, is however briefly outlined here. If an oscillation is in some manner started in the circuit LC, the resulting oscillatory current induces through the mutual inductance M_g an alternating emf. in the grid coil L_g, which in turn sets up a corresponding pulsating current in the plate circuit. This current, flowing in L_p, induces through the mutual inductance M_p, an alternating emf. in coil L, which, for suitable coupling and coils, is in synchronism with the emf. operating in the oscillatory circuit LC, and just of the required amplitude to exactly compensate for the energy losses in that circuit.

Special cases where either M_p or M_g are equal to L give the circuits of Figs. 168 and 169, which evidently function in the manner just described for the circuit of Fig. 167. It will be noted that the circuit of Fig. 168 is precisely the same as Fig. 161 which was used in previous sections of this chapter. It is well to note also that these two circuits, although based on the same principle—inductive coupling of the grid, plate and oscillatory circuit—are used for different purposes, as will be explained in a later part of this chapter.

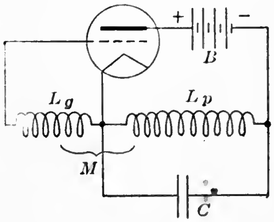


Fig. 168.

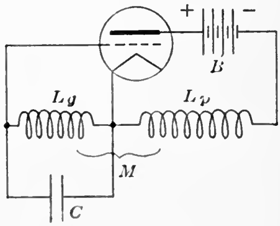


Fig. 169.

Another modification of the general circuit of Fig. 167 is shown in Fig. 170. It is evident from the fact that the oscillating current flows in the grid and plate coils connected in series, that this circuit does not require any mutual inductance between these coils. The reactive voltage drop across them, due to the common alternating current flowing through them, then determines the alternating plate and grid emf's. This circuit is of special value in its modified form, Fig. 171, where the filament is connected

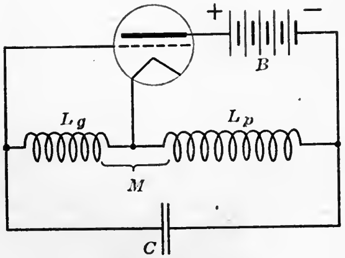


Fig. 170.

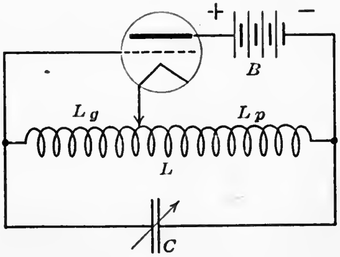


Fig. 171.

to the grid and plate through a sliding contact, the coils L_p and L_g being actually parts of a single coil wound on a straight cylindrical form. The coupling of the grid and plate circuits is then adjusted by means of the filament sliding contact, while the frequency of the generated oscillations may be varied by means of the variable condenser C. The advantage is that both adjustments are independent.

Electrostatic or capacitive coupling of the circuits may be used instead of the inductive coupling shown above. Thus, the circuit of Fig. 172 has been used extensively as the basis of a variety of capacitively coupled circuits. It is the counterpart, so to speak, of the circuit of Fig. 171. It should be noted that in this electrostatically coupled circuit, it would not be possible to connect the plate battery at N, as was done in the previous circuits, since the direct current plate-to-filament circuit would then be unable to convey the d.c. plate current, the plate being insulated from the filament by the condensers C_p and C_g. It is then necessary to connect the plate battery B as shown, directly between the filament and plate. The choke coil K is required in order to prevent the high frequency oscillations from passing through the

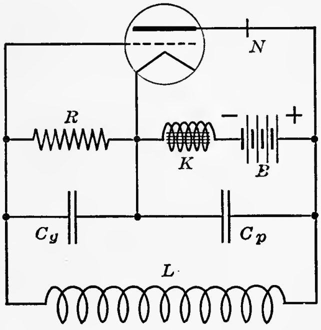


FIG. 172.

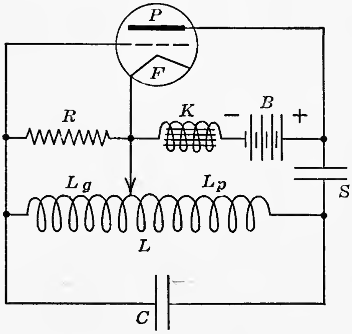


FIG. 173.

battery, and thus the coil prevents the battery from short circuiting the condenser C_p. Similarly, it is seen that the grid is insulated from the filament, so that it becomes necessary to connect it to the latter through a high resistance R, in order to provide a leakage path for the negative charge which would otherwise accumulate on it, and after a very short time would stop the flow of current in the plate circuit and block the tube. A choke coil may be connected in series with this resistance, in order to avoid short circuiting the condenser C_g. This may, however, be dispensed with in general, due to the high value of the resistance R.

The same method of supplying the direct current energy to the plate circuit may be used in the case of electromagnetic coupling. This is shown in Fig. 173. The operation of the circuit does not require additional explanation. When using this method of connecting the plate battery, it is necessary to insert a condenser S, as shown, between the battery and plate circuit inductance L_p in order to prevent the direct current plate circuit $BKFPB$ from being short circuited by this low resistance coil. The stopping condenser is made of sufficiently large capacitance to oppose only a low impedance to the alternating current component of the plate current.

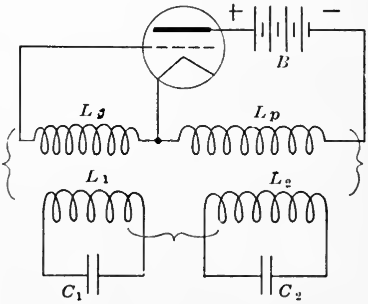


FIG. 174.

Finally, another possible combination is given in Fig. 174, showing the use of separate plate and grid oscillatory circuits. Two modifications of this circuit are shown in Figs. 175 and 176. The analysis of the properties of these circuits is not easily made qualitatively. A very complete mathematical treatment may be found in L. A. Hazeltine's paper mentioned previously.[1] Taking the simplest case, where the two oscillatory circuits have, independently, the same natural frequency, it will be understood by reference to Chapter III, that on account of their close coupling in the vacuum tube circuits considered here, this system will have

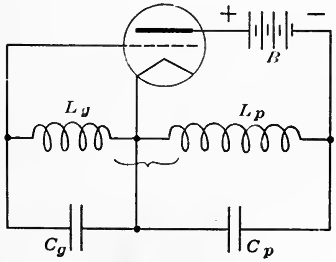


FIG. 175.

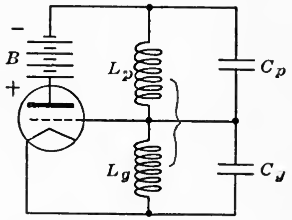


FIG. 176.

two resonance frequencies, one above and one below the natural frequency of the oscillatory circuits. By choosing suitable values of coupling and of the circuit constants, it is possible to use the circuit of Fig. 176 as a selective receiving circuit, acting as a regenerative amplifier between two given frequency limits, and as a

[1] *Proc. Institute Radio Engineers*, 1918.

power absorbing device for other frequencies. This will be taken up again in a later paragraph.

It was explained previously that oscillation generation is due to the amplification property of the vacuum tube, and is accomplished by coupling to each other the input and output circuits of the tube, as shown by the various circuits just considered. The oscillation will be greater the greater the amplification ratio of the tube, as determined by the ratio of the output and input voltages. It is then evident that if a cascade amplifier is used instead of a single tube, an oscillation will be obtained which is considerably stronger. Thus, consider the circuit of Fig. 177,

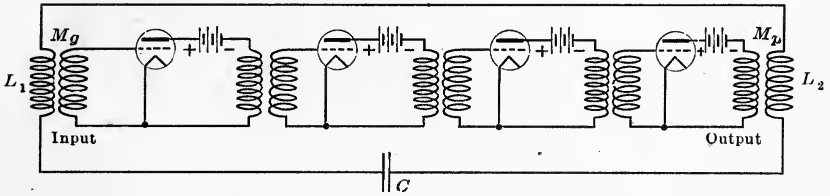
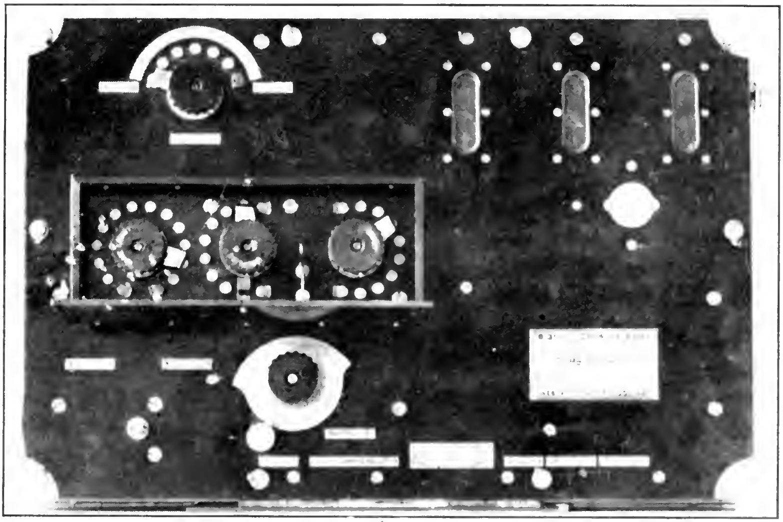


FIG. 177.

where the input and output circuits of a multi-stage cascade amplifier are coupled inductively through the oscillatory circuit L_1L_2C. This circuit is directly derived from that of Fig. 167, and it functions in the same manner. If an oscillation is started in the oscillatory circuit, an emf. is induced in the input grid circuit of the amplifier through the mutual inductance M_g, an amplified reproduction of which is impressed on the grid circuit of the last tube, resulting in a plate current variation in the last tube which, through the mutual inductance M_p, supplies the required energy for maintaining the oscillation. The arrangement is equivalent to using a single tube having an amplification factor equal to the overall amplification factor of the several tubes in cascade.

It may be seen from the above that a large number of vacuum tube oscillator circuits are possible. This permits the design of circuits especially suited to certain problems, each circuit having some individual characteristics resulting from its peculiar arrangement.

While vacuum tube circuits find application in laboratory work when it is desired to have a source of strictly constant and easily adjustable frequency, their most important applications are to

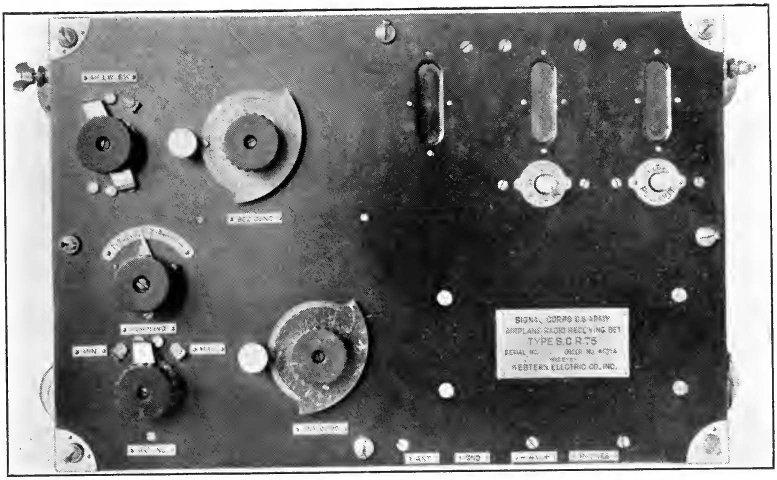


(*A*)

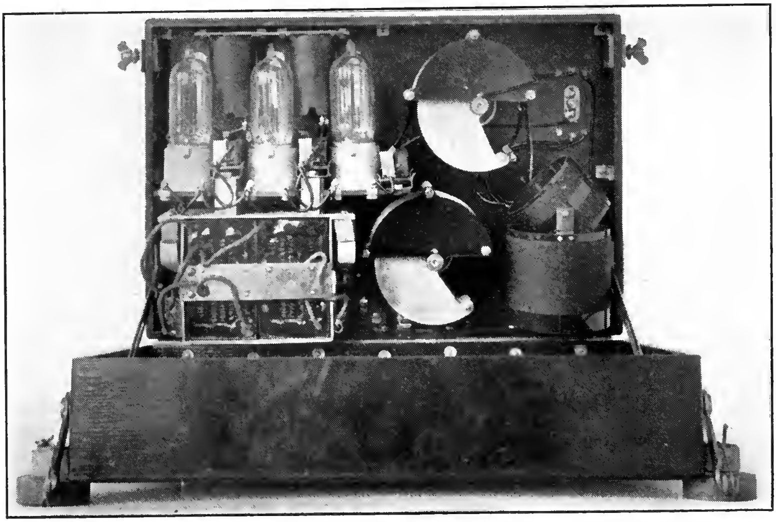
(*B*)

Plate 7.—A two-way radio telephone set designed for airplane use. The coil enclosing the oscillator and modulator tubes is so placed to save space. The tube at the extreme left is a ballast lamp serving to keep constant the filament current in the tubes. The three tubes at the right are the detector and two amplifiers. Signal Corps set type SCR–68.

(*Facing page* 218.)



(*A*)



(*B*)

Plate 8.—A damped or modulated wave receiving set having tuned primary and secondary circuits. The oscillation transformer seen at the bottom right hand corner has a pivoted secondary coil to vary the coupling. One detector and two amplifier tubes are used. Signal Corps set type SCR–75.

be found in radio telegraph and telephone transmitting and receiving work. Circuits pertaining to radio telegraph will therefore be studied first, after which some other applications of the oscillator circuits will be taken up in this Chapter and in Chapter XII. Radio telephony is studied in Chapter X.

Application of the Vacuum Tube Oscillator to Radio Telegraphy

Radio Telegraph Transmitting Circuits.—Vacuum tube oscillator circuits of the nature described above may be used for undamped wave radio telegraph transmitting. This is accomplished by exciting in the antenna oscillatory circuit undamped alternating currents generated by a vacuum tube circuit combination suitably tuned and coupled. The antenna circuit being of itself an oscillatory circuit, may be used directly in any one of the vacuum tube circuits studied above, in place of one of the oscillatory circuits coupled or connected to the tube circuits. Since the object in view is to transmit signals, care must be taken to arrange the circuits with a view to obtaining maximum oscillating energy in the antenna Some of the special conditions resulting from this application of the vacuum tube oscillator circuits are considered in the following paragraphs.

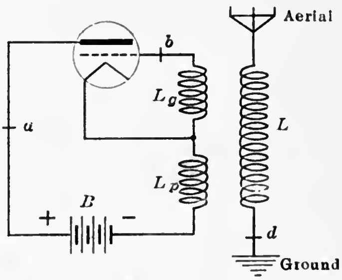


Fig. 178.

A simple circuit is shown in Fig. 178. This circuit is essentially the same as that of Fig. 167 with the exception that the oscillatory circuit LC of the latter is here an open radiating oscillatory circuit (antenna) instead of a closed non-radiating circuit. The circuit may be so designed as to require no coupling adjustment for a fairly large range of wave lengths. The only adjustment to be made is then that of tuning the antenna circuit so that it will have a natural wave length equal to the wave length to be transmitted. This is done by means of a variable inductance or capacitance inserted in the antenna circuit.

This circuit of Fig. 178 has all the advantages of the indirectly excited damped wave transmitting circuit of Fig. 84, with the added advantages resulting from the use of undamped oscillations. Thus, the antenna circuit of Fig. 178 like that of Fig. 84 is seen to have low ohmic resistance, and will therefore carry a high resonance current. In addition, in the case of the vacuum tube circuit, the antenna circuit is the only oscillatory circuit present, so that there are no delicate tuning and coupling operations required as in the case of the damped wave circuit having intermediate coupled tuned circuits. The wave is therefore very pure, and, having zero decrement, it permits of very sharp tuning at the receiving station. The only adjustment required is that of the plate and grid couplings, in order to come within the conditions of oscillation of the tube.

In this connection, it should be remembered that the greater the decrement of the oscillatory circuit, the greater the amount of power to be supplied to that circuit by the plate circuit of the tube in order to compensate for the energy losses and sustain an undamped oscillation. On the other hand, there is an upper limit to the alternating current power which may be supplied by any tube operating under given direct current conditions of filament, grid and plate voltages, which is reached when the a.c. grid potential oscillates between values corresponding to the upper and lower bends of the plate-current grid-voltage characteristic curve.

As a result of these combined conditions, it may be seen that a given tube will sustain undamped oscillations in an oscillatory circuit, provided the latter has a decrement smaller than a certain critical value. Now, in the case of Fig. 178, the oscillatory circuit continually loses energy by radiation into space. In fact, the greater the percentage of energy this circuit radiates, the better it performs its function as a transmitting antenna. It is thus seen that on account of the large amount of energy losses in the antenna circuit, the transmitting circuit of Fig. 178 requires, for satisfactory operation, a vacuum tube of large power capacity. The circuit will not operate satisfactorily with a low power tube and large antenna, and may even be entirely inoperative with an antenna circuit having too large a radiation for the tube in use. The power of the tube circuit may be increased by connecting several tubes in parallel, or else by using a single tube so constructed as to operate at a high plate potential, the maximum a.c. power being equal to the product of plate potential and current variations.

The radio telegraph transmitting circuit is of course not complete without a telegraph sending key. The function of this is to let the antenna circuit radiate energy at a certain predetermined wave length when the key is closed, and to stop such radiation when it is open. The two possible means of accomplishing this are to provide that the key, when open, will stop entirely the generation of the oscillations, or that it will simply detune the oscillatory circuit which will then radiate at a different wave length when the key is open, and therefore not energize the receiving circuit, which is tuned to the wave length corresponding to the closed position of the key.

The first method is best accomplished by placing the key at *a*, Fig. 178, in order to open the d.c. plate circuit of the tube and thus cut off the power supply. The oscillations may also be stopped by inserting the key at *b*, in the grid circuit of the tube. This, however, does not stop the electron flow from filament to plate, and merely disconnects and insulates the grid, which then accumulates a negative charge, as explained in a previous paragraph. When the key is then closed to start the oscillations, this charge must first leak off, and a certain lag occurs in the building up of the oscillations to their final value, which is objectionable as it introduces harmonics in the radiated wave train. This objection may be partly obviated by shunting the key contacts by a high resistance of the order of several megohms, to provide a leakage path for the charge from the grid to the filament while the key is open. The third possible location of the key is in the antenna circuit, at *d*, Fig. 178. This, however, is objectionable as the key contacts, when closed, may oppose a high resistance, which would prevent the oscillations from reaching their full amplitude. This is especially noticeable when operating at high speed, when the dots are made by a very rapid and light closing of the key.

The detuning method can be applied by connecting the key across one or a few turns of the antenna inductance, as shown in Fig. 179. The closing of the key then short circuits part of the antenna inductance, and thus alters the natural frequency and wave length of the antenna circuit. The antenna is thus seen to continually radiate energy, at one wave length when the key is open and the entire inductance in the circuit, and at a somewhat shorter wave length when the key is closed. Due to the sharp tuning of the receiving station, resulting from the use of undamped waves, the difference between the signaling and the detuned wave lengths need be only a few meters. In fact, both wave lengths are generally received, and, when the beat or heterodyne method of reception is used two notes are heard, one corresponding to the dots and dashes of the signals, and the other representing the intervals between. This does not prevent the signals from being received without difficulty, especially after some training of the receiving operator. It might be possible under certain conditions to adjust the frequency of the locally generated waves for heterodyne, to the same frequency as the incoming detuned wave so that there would be no note heard in the intervals.

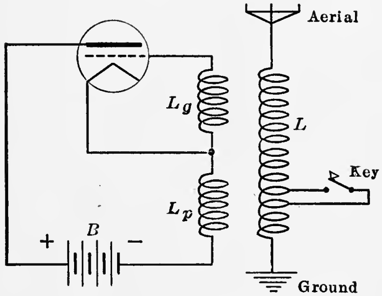


FIG. 179.

In order to overcome the tendency of the circuit of Fig. 178 to refuse to oscillate when used with a low power tube, a circuit using an intermediate oscillatory circuit as shown in Fig. 180, has been used successfully. Temporarily considering the antenna circuit as non-existent, the vacuum tube oscillator circuit is seen to be the same as that of Fig. 168. For a suitable value of the plate-to-grid coupling, strong undamped oscillations are then set up in the plate oscillatory circuit L_pC. As explained before, the oscillating current in the circuit L_pC may be many times the d.c. plate current. If now the antenna circuit is tuned to the circuit L_pC and coupled to it, as shown in Fig. 180, the current in the plate oscillatory circuit will induce undamped oscillations in the antenna circuit, and these oscillations, due to the large resonance currents in the circuit L_pC, will be of sufficiently great amplitude to operate the circuit despite the radiation of energy from the antenna.

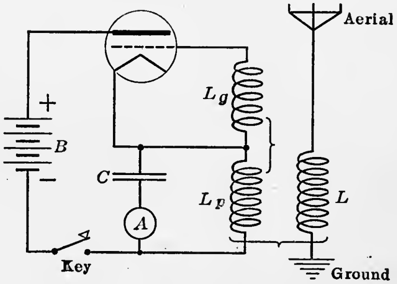


FIG. 180.

It may be of interest to describe the method of adjusting a circuit of this kind to radiate at a desired wave length. The antenna circuit is first opened, or its coupling so reduced that it will not be appreciably energized by the vacuum tube circuit and may be considered as absent. A hot wire ammeter A is connected in the plate oscillatory circuit L_pC, and the tube is made inoperative by opening the key, or cutting off the filament current. The circuit L_pC is then tuned to the desired wave length by exciting it with a wavemeter suitably set, and varying the condenser C or inductance L_p until resonance is obtained, as indicated by the ammeter A. The wavemeter is then removed, and the tube energized. The key is closed, and the vacuum tube circuit is adjusted for maximum oscillation generation. This is simply done by adjusting the grid-to-plate coupling until a maximum reading is obtained on the ammeter A. The antenna circuit is then fairly loosely coupled to the plate oscillatory circuit, and is tuned to that circuit by varying its inductance or capacitance.

When resonance conditions are obtained, it is known that the maximum current will be induced by the circuit L_pC in the antenna circuit. In other words, at resonance, the antenna circuit draws the greatest amount of energy from the plate oscillatory circuit, and the condition of resonance is thus indicated by a minimum reading on the ammeter A. The last step is now to obtain the coupling between the antenna and plate oscillating circuits for which maximum energy will be transformed from the latter to the former circuit. As was pointed out in Chapter III, and as may be easily understood, this corresponds again to the value of antenna-to-plate coupling which gives a minimum reading on the ammeter A. It is important to remember in making this adjustment that the coupling should never be made so close as to give a "double humped" resonance curve in the antenna circuit. This may be easily checked up for any value of coupling, by varying the wave length of the antenna circuit. A single minimum should be observed on the ammeter A, corresponding to a single maximum in the antenna circuit. The method above described is better than one using an ammeter in the antenna circuit, for the reason that it greatly facilitates the adjustments of the vacuum tube plate and grid circuits.

A circuit which has been successfully used is shown in Fig. 181, which is an adaptation of the circuit of Fig. 173 studied previously. The condenser C of the circuit of Fig. 173 has simply been made the antenna in the latter circuit, the antenna forming a condenser of large linear dimensions. The theory of both circuits is the same. The only change in the circuit is the addition of a second sliding contact W, which is used for adjusting the wave length of the antenna circuit. As may be seen from Fig. 173, the natural frequency of the oscillatory circuit L_gL_pC may be altered by varying either the condenser C or the value of the total inductance L_gL_p. In the case of the circuit of Fig. 181, the condenser C being the antenna, its capacitance cannot be readily varied, and it is therefore necessary to vary the inductance by means of the contact W. The filament sliding contact is then

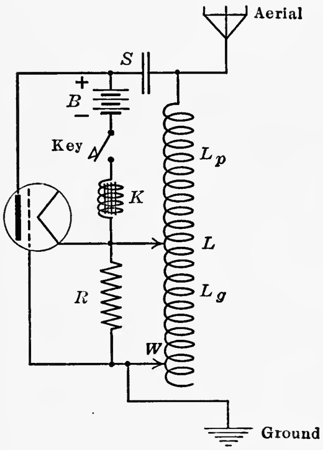


FIG. 181.

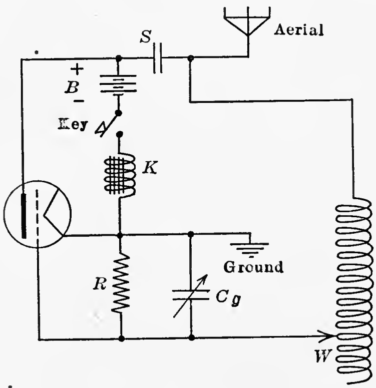


FIG. 182.

used, as before, to adjust the grid-plate coupling. The function of the condenser S need not be explained again, having been treated in connection with the circuit of Fig. 173.

As a last illustration, the circuit of Fig. 182 is shown as an application of the electrostatically coupled circuit of Fig. 172. The contact W performs the same function as in the previous example, while the variable grid condenser C_g is used to adjust the grid-to-plate coupling. The condenser S has been added to prevent a short circuit in case of accidental grounding of the aerial. This is a precaution which finds full justification in airplane radio sets where, as will be seen in Chapter XII, the trailing antenna frequently gets in a position to come into contact with the rest of the airplane which forms the counterpoise.

The above discussion of vacuum tube transmitting circuits can but give a very general idea of the subject. It may be seen that the number of circuits which may be used is too great to be covered in a book of this size and purpose. It is hoped, however, that the fundamental principles involved have been made sufficiently clear to enable the reader to intelligently analyze the various circuits which he may encounter in practice.

Radio Telegraph Receiving Circuits.—Three-electrode vacuum tube oscillator circuits have found wide application not only in undamped wave transmitting, but also in undamped wave radio telegraph receiving. Referring back to Chapter V, it was shown that one of the most efficient methods of reception of undamped waves was the heterodyne method, involving the generation at the receiving station of a high frequency alternating current, having an adjustable frequency of the order of that of the incoming undamped oscillations. This generation may be effected by means of a three-electrode vacuum tube connected as an oscillator, and used in place of the alternator shown in Figs. 124 and 126.

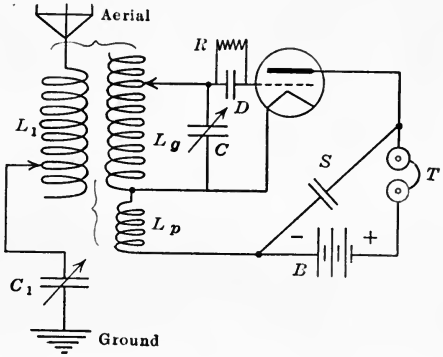


Fig. 183.

It is possible, however, to use a single vacuum tube for both detection of the incoming oscillations and for generation of the local oscillations. Thus, adapting the circuit of Fig. 169 to the present purpose, the circuit of Fig. 183 is obtained. This consists essentially of a three-electrode vacuum tube, the plate and grid of which are coupled to each other by means of the coils L_p and L_g. The locally generated oscillations are set up in the circuit L_gC, and the circuit is thus seen to be similar to that of Fig. 169. The plate circuit is energized by the battery *B* and comprises also telephone receivers T. The incoming signals are received on a tuned antenna circuit coupled to the grid circuit of the tube. This grid circuit also comprises the condenser D shunted by the resistance R. The operation of the tube is then as follows.

The antenna circuit is tuned to the frequency of the signal waves to be received. These waves then set up in this circuit strong oscillations of their own frequency, which by induction produce corresponding oscillations of the grid potential of the tube. But the tube being in an oscillating condition, locally generated oscillations are also set up. These are of the frequency of the grid oscillatory circuit L_gC, which is adjusted to a frequency slightly different from that of the antenna circuit. There are thus two alternating emf.'s of slightly different frequencies impressed upon the grid. As explained in Chapter V, the grid potential resulting from these two component has the nature of a "beat" current of a beat frequency equal to the difference of the component frequencies. For suitable values of grid, plate and fila-

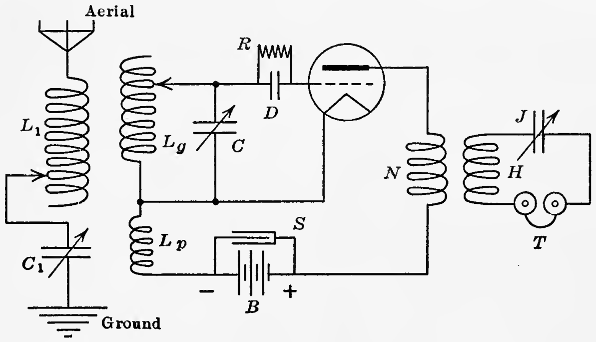


FIG. 184.

ment battery voltages, it is then possible to bring the d.c. operating point of the tube where it will act as a rectifiying detector for incoming oscillations while still oscillating: It follows that variations of the grid potential produced in this manner bring about audio frequency pulsations of the plate and telephone receiver current. The purpose of the condenesr S is to by-pass the locally generated high frequency oscillations which would otherwise be absorbed in the plate battery and telephone resistance. This method of self-heterodyne reception using a single tube for the operations of detection and oscillation generation is called the *autodyne* method of reception.

As was pointed out in Chapter V, the audio frequency pulsating current resulting from the use of the heterodyne method of reception has an almost sinusoidal alternating current component. It was also pointed out that the audio frequency, that is, the pitch of the sound in the telephone receivers, is directly dependent on the original oscillation frequency. By using in addition to the radio frequency circuits just described, an audio frequency tuned telephone receiver circuit, it is possible to increase greatly the intensity of the received signals, and to increase considerably the selectivity of the receiving circuit, as a result of the use of several tuned oscillatory circuits in succession. This is illustrated in Fig. 184, which differs from Fig. 183 only in that the telephone receivers instead of being directly connected to the plate circuits of the tube, are connected to an oscillatory circuit *HJ* coupled to the plate circuit, and tuned to the *beat frequency*, that is, to the difference between the frequency of the locally generated currents and of the incoming oscillations.

As an example of further modifications of the circuits, what might be termed a universal receiving circuit is shown in Fig. 185. This comprises an antenna circuit tuned to the signal oscil-

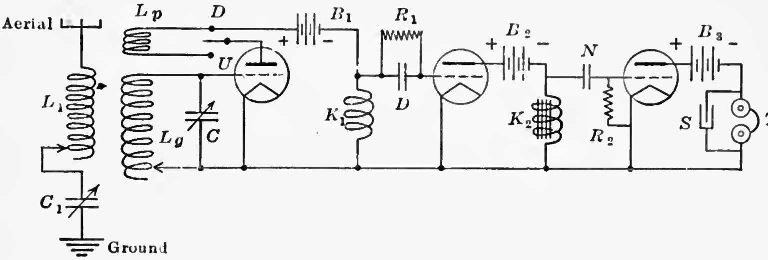


Fig. 185.

lation frequency, and three vacuum tubes in cascade. The first tube is coupled to the antenna circuit through its tuned grid circuit. Its plate circuit comprises a switch, which, when closed in the position *D* excludes the coil L_p from the plate circuit. For this connection the first tube acts as a high frequency amplifier. This position is used for the reception of damped or modulated wave signals. With the switch in the position *U*, the plate circuit is coupled to the grid through the coil L_p, frequently called a "tickler coil." If this coupling is very loose, then the first tube acts as a regenerative amplifier, for use with damped wave signals. If the coupling is close, the first tube generates oscillations, and is used for heterodyne reception of undamped waves. The second tube is a rectifying detector while the last tube is an audio frequency amplifier. The cascade connection here shown comprises choke coil coupled circuits, but transformer or resistance coupling may be used equally well.

Oscillator Circuits for Extreme Frequencies

Since the frequency of the oscillations maintained by a three-electrode vacuum tube in an oscillatory circuit is approximately equal to the natural frequency of that circuit, it is therefore possible with a suitable design of this circuit, to obtain alternating currents of any desired frequency.

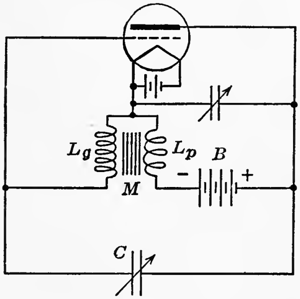


Fig. 186.

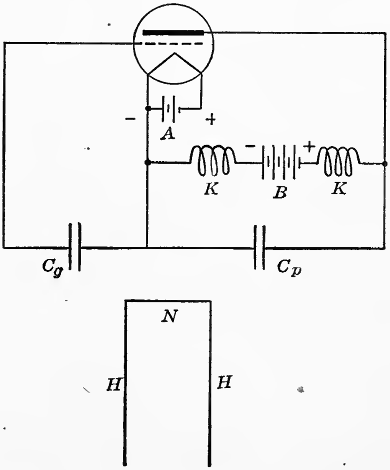


Fig. 187.

A circuit used for the generation of alternating currents of extremely low frequencies—as low as one-half cycle per second—is shown in Fig. 186.[1] This circuit is similar in principle to those previously studied but uses low frequency apparatus. Thus, the transformer coupling of the plate and grid circuits has an iron core, and the inductance coils and condensers have very large values. By means of a zero-center d.c. ammeter, it is possible to read instantaneous values of the current cycle, the period of which is greater than the inertia period of the ammeter needle.

For the generation of extremely high frequencies, the circuit of Fig. 187 has been used successfully to produce as high as 150 million cycles per second and even higher in some instances.[2]

[1] W. C. White, *General Electric Review*, 1916, p. 771.
[2] *Revue Générale de l'Electricité*, March 15, 1919.

In this circuit the condensers are of small capacitance, and the inductance used is that of the connecting wires, which are all made very short, and the length of which is varied when the frequency is to be altered. At such high frequencies, the plate-to-filament and grid-to-filament capacitances have an appreciable effect on the frequency of the oscillations and must be taken into account.

For certain designs of tubes, the grid coupling condenser C_g may even be omitted. The capacitance of the grid itself with respect to the other elements of the vacuum tube then supplies the necessary grid coupling capacitance.

The internal capacitance of the tube at extremely high frequencies results in an out-of-phase plate current being superimposed upon the ordinary in-phase plate conduction current. This out-of-phase current leads the in-phase current and causes the operating point of the tubes to describe in clockwise direction a closed loop having the normal grid-voltage plate-current dynamic characteristic curve as its axis. The area enclosed by this loop is proportional to the wattless energy component of the cycle.

As all electrostatic and electromagnetic induction effects are very considerable, the wire and other parts of the circuit must be laid out very carefully, for their relative positions materially affect the constants of the circuit. The choke coils *K*, which do not require more than a few turns to be effective, prevent the high frequency current from flowing through the battery. The wires connecting the battery terminals to these two coils thus do not form part of this a.c. circuit, and do not require any special attention as to their length or shape. They should, however, be as short as possible, and should be run, as far from the high frequency circuit as possible.

The wires *HH* are two long straight wires held parallel to each other, and connected at one end by a wire *N* in the proximity of the high frequency circuit. The mutual inductance of this wire *N* and the vacuum tube circuit is great enough to induce strong oscillations in the circuit *HNH*, and standing waves may easily be evidenced along the wires *HH*. As an illustration of the effect of the vacuum tube capacitance, it may be mentioned that certain tubes have too great an internal capacitance for extremely high frequencies. There is thus a limit for each specific design of tube to the frequencies which may be obtained.

CHAPTER X

RADIO TELEPHONY

General Underlying Principles

In undamped wave radio telegraphy, the transmitting circuit radiates energy continuously when the key is closed, at a constant and uniform rate, as a result of the constant amplitude of the alternating current in the transmitting antenna. There results, after simple rectification in the receiving circuit, an unvarying direct current, which, as was shown in Chapter V, simply deflects the telephone receiver diaphragm, without setting it into vibration, and therefore without producing any sound.

Now if the amount of energy radiated by the transmitting antenna is varied, for instance by varying the maximum amplitude of the alternating current flowing in the antenna, a correspondingly varying or pulsating unidirectional current will, after rectification, flow in the radio receiving circuit and telephone receiver. And if these variations of the current in the transmitting antenna are of audio frequency, such as those produced by a telephone transmitter actuated by the voice, a sound having the same frequency and characteristics as the variations of the transmitted energy will be produced in the telephone receivers of the radio receiving circuit, thus enabling the reproduction at the receiving station, under suitable conditions, of the sounds or speech producing these variations at the transmitting station.

A radio telephone transmitting circuit is thus seen to consist essentially of a radiating or antenna circuit in which oscillations are set up by some generator of undamped oscillations, and an arrangement of circuits whereby the amplitude of these oscillations may be varied or modulated by the human voice vibrations. The most important methods for generating undamped oscillations have already been studied; namely, the arc, the alternator and the three-electrode vacuum tube oscillator. The study of radio telephone circuits will therefore be mostly confined to that of modulating systems, and the methods of their connection or coupling to the oscillator systems previously studied.

A radio telephone receiving circuit, as explained above, is essentially the same as a damped wave radio telegraph receiving circuit and comprises a tuned antenna circuit, coupled or connected to a rectifying detector and telephone receiver circuit and such auxiliary tuned and amplifier circuits as may be required. Certain limitations and modifications of these circuits will be explained later, but it will be seen that they do not affect the fundamental characteristics just stated.

METHODS OF MODULATION

Generally speaking, the methods of modulating the high frequency alternating current flowing in the transmitting antenna may be divided into two general classes; namely, direct current modulation and high frequency modulation. This may be understood when it is considered that in the three methods mentioned above of generating high frequency oscillations—the arc, the alternator, and vacuum tube oscillator—there is simply a transformation of direct current into high frequency alternating current. Thus, in the oscillating arc, direct current is fed into the arc and alternating current obtained in the output oscillatory shunt circuit. In the high frequency alternator, direct current is sent into the field winding, and alternating current obtained in the output or a.c. armature circuit. In the three-electrode vacuum tube, it is the energy furnished by the plate battery which is transformed into alternating current. In all three cases, the alternating current power output is, at least within limits, proportional to the amount of direct current input.

One method of varying the amplitude of the generated high frequency alternating current is therefore to vary or modulate the direct current input before its transformation into alternating current. The other method is direct modulation of the alternating current output.

The first step in the modulation process is to transform the sound vibrations due to the voice into variations of the electric current in the transmitting circuit, proportional in amplitude to the sound vibrations, and of the same frequency and wave shape. This is generally achieved by means of the ordinary carbon microphone or telephone transmitter, although special transmitters have been built, as will be mentioned later, in order to overcome certain practical difficulties. One method also uses a condenser type transmitter, the principle of which is described in a later paragraph.

One difficulty which was encountered in radio telephone transmission is that for most effective modulation, the transmitted energy must be modulated or controlled over a large part of its maximum or steady value. That is, the best modulation will be that for which the transmitted energy may be made to vary all the way from its maximum or steady value to zero. Now, unless the range or distance of transmission is very short, the amount of power to be thus controlled is of the order of several kilowatts, and may be as great as 100 or 200 kw. when using an Alexanderson alternator. In view of the fact that the sound waves set up by the voice, upon which control and modulation of such a large amount of power must depend, have themselves a power of the order of only 0.00000001 watt, it is evident that they must first be amplified in some manner.

Three methods have been used principally to accomplish this purpose. They are generally designated as the microphonic method of control, the vacuum tube method, and the ferromagnetic method. The first of these has practically become obsolete since the introduction of the last two. The last mentioned is especially useful for handling very considerable amounts of power, while the second, although applicable to high power control, is economically used on low and medium power sets only. Combinations of the last two classes of control methods are also frequently used and are very efficient and flexible.

Microphonic Methods of Modulation.—The microphonic methods of control or modulation of the transmitted energy were probably the earliest methods used. They were employed principally in conjunction with circuits having an oscillating arc as the source of undamped oscillations. The arc circuits are therefore shown here in this connection.

Since an ordinary telephone transmitter will work satisfactorily at a current of not more than about 0.2 amp. and handle a maximum power of about 2 watts, a solution of the problem of handling greater modulating power was sought by connecting a number of similar microphones in parallel, in order to control a correspondingly larger power. The microphones were all mechanically and acoustically connected to a single mouthpiece in which the operator talked. This method of modulating the radiated energy, although apparently simple in principle, is not satisfactory. In general, the various microphones are of different resistances, which result in an uneven distribution of the current among them and produce excessive overloading and consequent heating of some of them. Also, the various microphones do not have identically the same period of vibration, inertia, and modulation characteristics, so that the voice vibrations are not reproduced synchronously or identically in the various transmitters, the result being a distorted modulation of the transmitted energy.

Attempts have been made to build a single microphone capable of carrying heavy currents, and provided with water or oil cooling systems, or in which the carbon granules were kept in constant motion or changed continuously. Microphones were tried also in which the resistance was no longer varied by a carbon contact, but by the variable volume of a water jet. All these various schemes, although more or less satisfactory in operation, are clumsy and little better than makeshifts. They are obviously inapplicable in small, portable radio telephone sets such as used in the field by military signal detachments, or on board any aircraft.

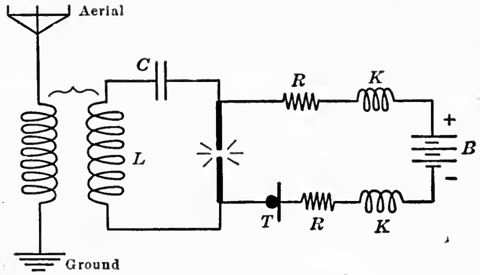


Fig. 188.

As mentioned above, these methods of control were mostly used in connection with the oscillating arc circuits, which are also entirely impracticable for field or aeronautical work. They are given here, however, in order to familiarize the reader with the history of the methods of modulation of the transmitted energy. All these circuits are shown using a single microphone as the modulating system. But it is evident that since the arc is a source of high frequency alternating current, just the same as the alternator or vacuum tube oscillator, the modulation methods described later are applicable to it, if suitably modified as required by the peculiar characteristics of the circuits.

The direct current method of modulation is illustrated by the circuits of Figs. 188 and 189. These show an arc connected for oscillation generation, as explained in Chapter V, and coupled to a tuned antenna circuit. The arc is supplied with direct current from a battery B in series with choke coils K and stabilizing resistances R. In series with this direct current supply circuit is the microphone telephone transmitter T which is either directly connected as in Fig. 188 or coupled as in Fig. 189. The resistance of the microphone is varied at audio frequency by the

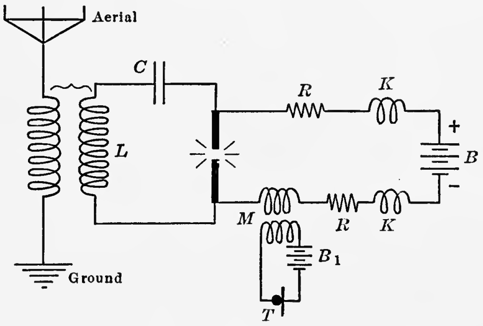


Fig. 189.

operator's voice vibrations, producing two simultaneous effects. Through the resistance of the d.c. supply circuit made variable by the transmitter, a correspondingly variable amount of d.c. power is fed into the arc to be transformed into alternating current. Also, as a consequence of the varying potential drop across the microphone, a varying d.c. potential is impressed across the arc, with a corresponding variation of the position of the d.c. operating point of the arc along its static characteristic curve. These two effects combine to produce variations or modulations of the high frequency alternating current output of the arc, and therefore of the radiated energy. The limits between which this modulation is without distortion will be studied in a later paragraph.

The frequency of the oscillations set up by the arc in the oscillating circuit may be considered roughly to be constant. It is determined by the constants of the oscillatory circuit. The effect of modulation is thus to vary the amplitude of these oscillations as represented by the curves of Fig. 190. The frequency of the modulation variations of the antenna current is of audio frequency and hence, of course, is very much lower than the radio

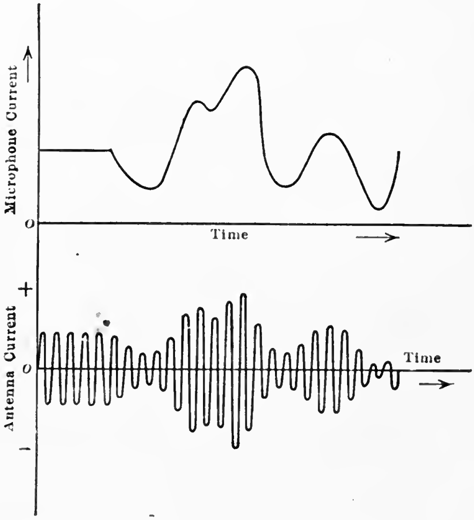


Fig. 190.

frequency of the oscillations themselves, as is clearly illustrated by the curves.

Some simple methods of high frequency modulation, that is, of direct modulation of the high frequency oscillations, are illustrated in Figs. 191, 192, 193 and 194. In these circuits, a con-

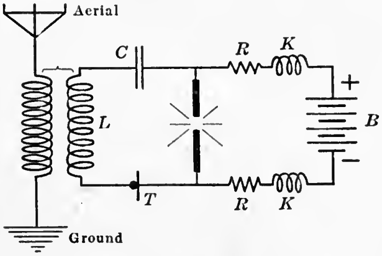


Fig. 191.

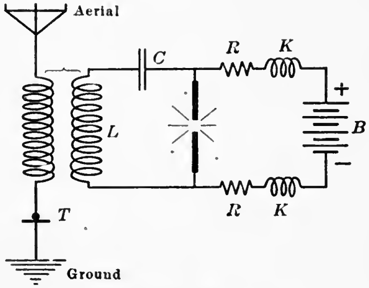


Fig. 192.

stant amount of d.c. energy is supplied to the arc by the battery *B*, and is transformed by the latter into a similarly constant amount of high frequency a.c. energy. This constant a.c. energy is used entirely in sustaining undamped oscillations in the system, since it supplies the energy losses at every cycle. That is to say, this energy is consumed entirely in the resistance of the oscillatory circuit. The two principal components of this resistance are the ohmic resistance of the circuit and the radiation resistance. Since the total energy loss in the resistance is constant, if one of these component resistances is varied, as is the

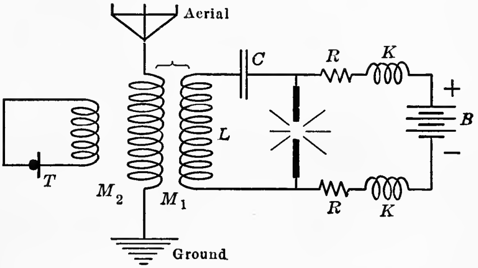


FIG. 193.

case with the ohmic resistance when the microphone is set into vibration by the voice, then the distribution of the energy between these two resistances will correspondingly vary. It follows that the energy spent in the radiation resistance of the system, which is the energy transmitted to the receiving station, may be varied or modulated by varying the ohmic resistance of the circuit by means of a microphone.

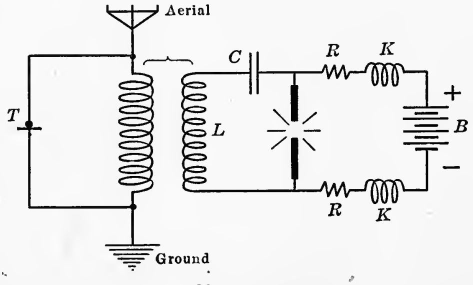


FIG. 194.

Thus, in the circuits of Figs. 191 and 192, the microphone is placed directly in series with one of the oscillatory circuits, and thus provides a means of varying its ohmic resistance. In the circuits of Figs. 193 and 194, the operation is slightly different, and is frequently referred to as the "absorption method" of modulation. In the case of Fig. 193, the constant d.c. energy supplied to the arc and transformed by it into a constant amount of high frequency a.c. energy, is transferred to the antenna through the inductive coupling M_1, and then uniformly and continuously radiated by the antenna. The microphone circuit T, however, being coupled to the antenna circuit, receives some energy from the latter. The amount of energy thus absorbed by the circuit T is directly proportional to the resistance of this circuit. A part of the energy supplied to the antenna circuit is thus locally absorbed in the microphone circuit, and the remaining energy radiated into space. By varying the microphone resistance with the voice, the amount of energy thus absorbed, and therefore the amount of energy radiated is varied or modulated in accordance with the speech or vibrations of the transmitter diaphragm.

Detuning Method of Modulation.—A method of microphonic modulation which is a departure from the above methods is known as the "Fessenden detuning method." This makes use of a condenser transmitter instead of an ordinary variable resistance microphone. The circuit is schematically illustrated in Fig. 195. The antenna circuit *ALCG* is coupled or connected to a source of undamped high frequency alternating current, and tuned to the frequency of the latter. In parallel with the antenna condenser or inductance coil a condenser-transmitter T is connected. This transmitter consists essentially of a condenser having one or several fixed plates and a like number of movable plates. The distance between the fixed plates and the movable ones may be varied by the voice vibrations, and this in turn varies the capacitance of the transmitter. By talking into the transmitter, the antenna circuit is thus detuned more or less with respect to the source of oscillations, and, as may be readily understood by referring back to the paragraph on resonance curves, the antenna current is correspondingly varied or modulated.

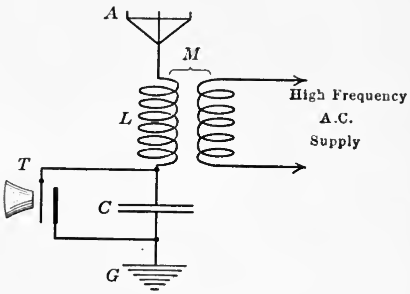


Fig. 195.

Vacuum Tube Methods of Modulation.—The use of three-electrode vacuum tubes has brought about improvements over the methods of modulation described above, both in the construction of the apparatus and in its operation. The essential function of the vacuum tube in a modulating system may be defined in a general manner as being to amplify the variable resistance reaction of a carbon microphone transmitter. The tube can be made to perform this function by connecting or coupling the microphone to the input or grid circuit of the tube, and connecting or coupling the output or plate circuit in place of the microphone shown in the previous circuits. On account of certain properties peculiar to the tube, however, this cannot be done in all of these circuits in a simple manner. The principal advantage of the vacuum tube modulator is derived when it is used in connection with a vacuum tube oscillator circuit, for then the modulator and oscillator tubes can be identical, when the modulating and oscillating circuits will have equal power capacities, which permits 100 per cent modulation.

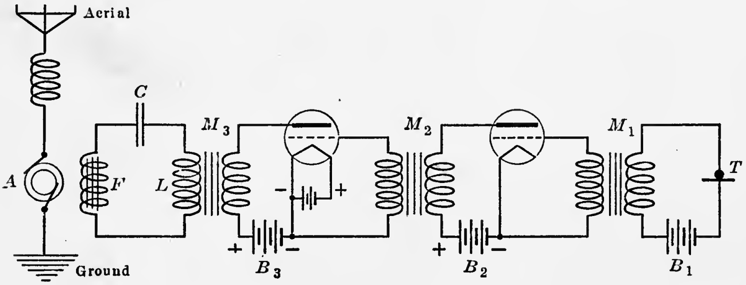


Fig. 196.

The application of the vacuum tube to direct current modulation may be simply done by inserting between the two coils M, Fig. 189, a cascade amplifier having its input circuit coupled to the microphone circuit, and its output circuit coupled to the d.c. supply circuit of the oscillator. An interesting application of this principle has been proposed for use with a high frequency alternator. It is illustrated in Fig. 196. It consists in connecting the field winding F of the alternator A to the output circuit of a vacuum tube cascade amplifier, the input circuit of which comprises a telephone transmitter T. When no speech is impressed on the transmitter, the plate current of the tubes is not varied, and there is no current in the field circuit. The alternator then generates no current whatever. If, however, the telephone current is varied by speaking into the transmitter, the plate current of the last tube will pulsate, and will induce in the field circuit *FLC* a current of varying amplitude following the modulations of the voice and considerably amplified as compared with the current in the microphone circuit. The existence of the varying field current will then cause a varying amplitude high frequency alternating current to be generated by the alternator. And since the alternating current generated is proportional (under suitable conditions) to the field current, the former will then also follow the voice modulations. The advantage of this method is that no radiation occurs when the operator is not talking, and there is thus no a.c. energy wasted.

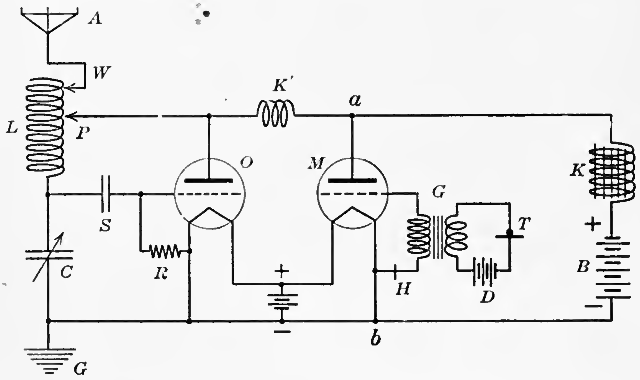


Fig. 197.

Very efficient methods of d.c. modulation also have been developed as mentioned above, when both the modulator and oscillator are vacuum tube circuits. One such method, known as the Heising modulation system, is schematically illustrated in Fig. 197. This circuit was developed by the Western Electric Co. and successfully used during the war by the U. S. Signal Corps on land and airplane radio telephone work. The battery *B* forms the supply of direct current energy and is connected in series with a choke coil *K* of large inductance, so that the direct current supplied by the battery to the plates of the tubes is kept constant even if the resistance of the external circuit is varied at audio frequency. In other words, the system *aKBb* forms a constant current source of supply. Across this are connected in parallel the plate circuits of two preferably identical three-electrode vacuum tubes M and O, so that the direct current supplied by the battery B divides between the two tubes in inverse proportion to their respective internal d.c. plate resistances.

The tube M is used as a modulator, and its grid circuit is therefore coupled to a telephone transmitter T through a telephone transformer G. By talking, into the transmitter, the direct current flowing from the battery D is varied and a varying potential is induced between the grid and filament of the tube M. This in turn alters the internal plate resistance of this tube, and therefore the distribution of the total constant current supplied by the battery B to the two tubes M and O. The tube O is connected for oscillation generation and coupled to the antenna oscillatory circuit. In the case of Fig. 197, the grid and plate circuits of the oscillator tube O are coupled for oscillation generation through the inductance L, grid coupling condenser C, and the aerial-to-ground condenser which forms the plate coupling condenser. The functions of the grid stopping condenser S, grid leak resistance R, plate coupling switch P, and wave length adjusting switch W were explained in connection with the study of vacuum tube oscillators, Chapter IX. The choke coil K', which is of low ohmic resistance, and of only a few turns, serves the purpose of preventing high frequency currents generated by the tube O from flowing through the modulator tube M, thus isolating the latter from the radio circuits.

The operation of the circuit just described is then as follows: The oscillator tube O generates undamped oscillations in the antenna circuit of frequency or wave length determined by the constants of the antenna circuit. The amplitude of these oscillations is proportional[1] to the amount of direct current supplied to the oscillator tube by the battery B. If speech is impressed then on the transmitter T, the grid potential and internal resistance of the modulator tube M are varied, with a corresponding variation of the amount of current which can flow from the plate to the filament. The current supplied to the oscillator tube O is then also varied, since it is always equal to the difference between the constant total current supplied by the battery B and the plate current of the modulator tube.

This system is very effective because the two tubes are identical and hence have the same power capacity, and the amplitude of the oscillations can thus be varied from zero to twice their normal

[1] Within limits established later.

(non-modulated) value. This extreme modulation is of course obtained only when the grid potential of the modulator tube can be varied by means of the telephone transmitter between sufficiently wide limits. When using certain types of high power tubes, this is difficult to achieve with a single telephone transformer *G*, and it becomes necessary to insert between the transmitter and modulator tube an audio frequency amplifier tube *A*, Fig. 198, which then amplifies the potential variation due to the speech vibrations of the transmitter *T*.

On the other hand, if the grid variations of the modulator tube are so great as to entirely cut off the plate current through that

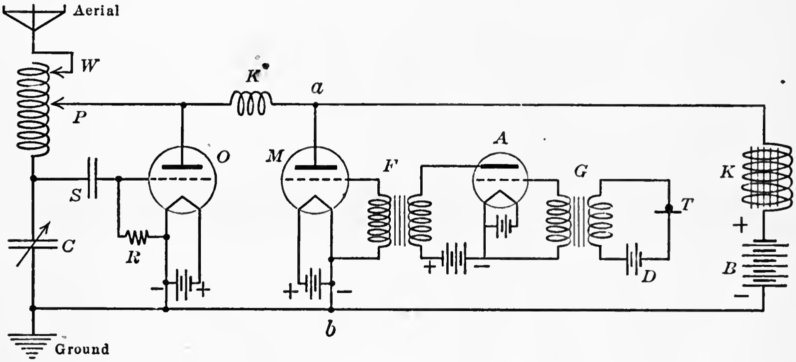


FIG. 198.

tube, or to bring the operating point in the upper flat part of the characteristic curve, part of the speech modulation will be imperfectly reproduced or entirely lost. This is commonly known as "blocking" of the modulator.

Direct current control with tubes may be obtained by connecting the transmitter to act directly on the grid circuit of the oscillator tube as shown schematically in Fig. 199, where the tube connected for oscillation generation comprises in its grid circuit a telephone transformer. The principle of this method is to impress on the grid of the tube a varying potential by talking into the transmitter *T*, thus altering the internal plate resistance of the tube. This in turn alters the amount of direct current supplied by the battery *B* to the tube and transformed into alternating current.

This method is not a very good one, although simple, as may readily be understood by referring to the preceding chapter where the effect of grid potential on the generated oscillations was studied in detail.

The absorption method of high frequency modulation represented in Fig. 193 has also been used successfully with vacuum tubes. As illustrated in Fig. 200, undamped oscillations are set up in the antenna circuit *ALG* by means of a high frequency alternator, which may also be an oscillating arc or a vacuum tube circuit suitably coupled or connected. Coupled to the antenna circuit is the plate circuit of the modulator tube *M*, the coupling being effected through the coils *L* and L_p. The grid circuit of the tube is coupled to a telephone transmitter circuit. Generally speaking, the function of the modulator tube circuit inserted

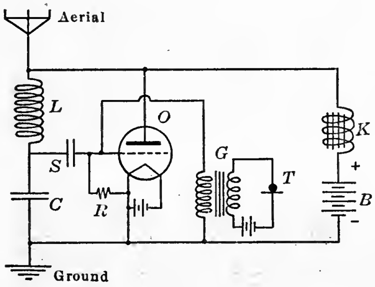


FIG. 199.

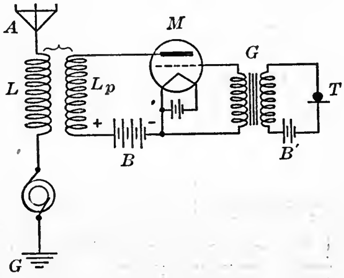
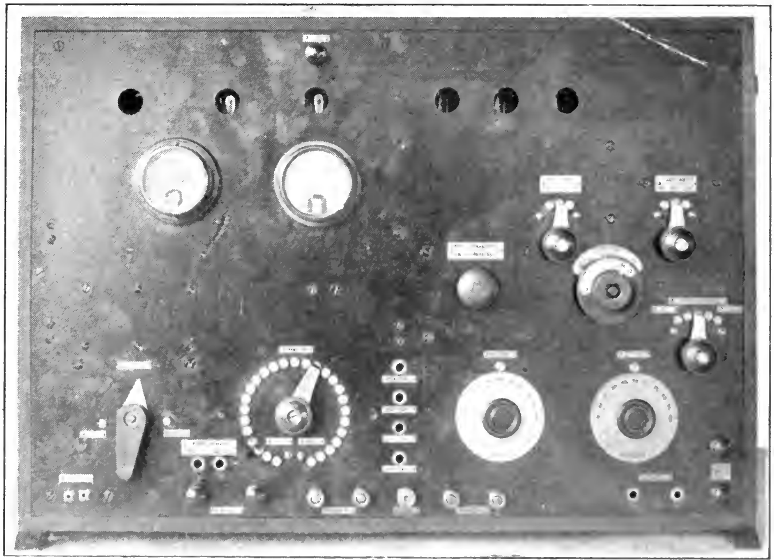


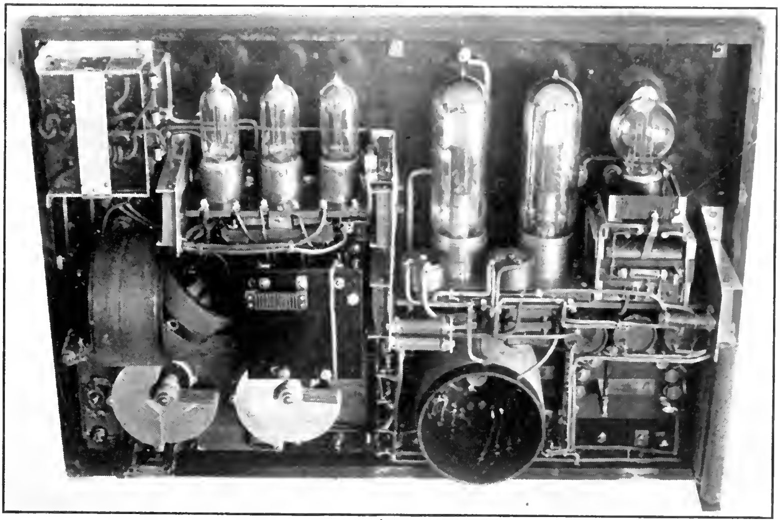
FIG. 200.

between the antenna and microphone circuits is to amplify in the antenna circuit the small variable resistance reaction of the microphone circuit. The actual operation, however, is somewhat more complex than that of the ordinary amplifier and may be explained as follows:

When no speech is impressed on the telephone transmitter *T*, Fig. 200, the grid potential has a fixed value, which in the case of the figure is zero, but which is generally made negative by means of a battery connected in series with the grid in order to reduce or prevent the flow of current in the grid circuit. Under such conditions of unvarying grid potential, the current flowing in the plate circuit is the resultant current due to the emf. of the battery *B*, and the high frequency alternating emf. of constant maximum amplitude induced in the coil L_p, by the antenna current generated by the alternator and flowing in coil *L*. The plate current is then pulsating, and each of its d.c. and a.c. components has an unvarying amplitude.



(A)



(B)

Plate 9.—A high power two-way radio telephone set. From left to right the vacuum tubes are a detector, two receiving amplifiers, a transmitting oscillator, modulator and audio frequency transmitting speech amplifier. Signal Corps set type SCR–109.

(*Facing page* 242.)

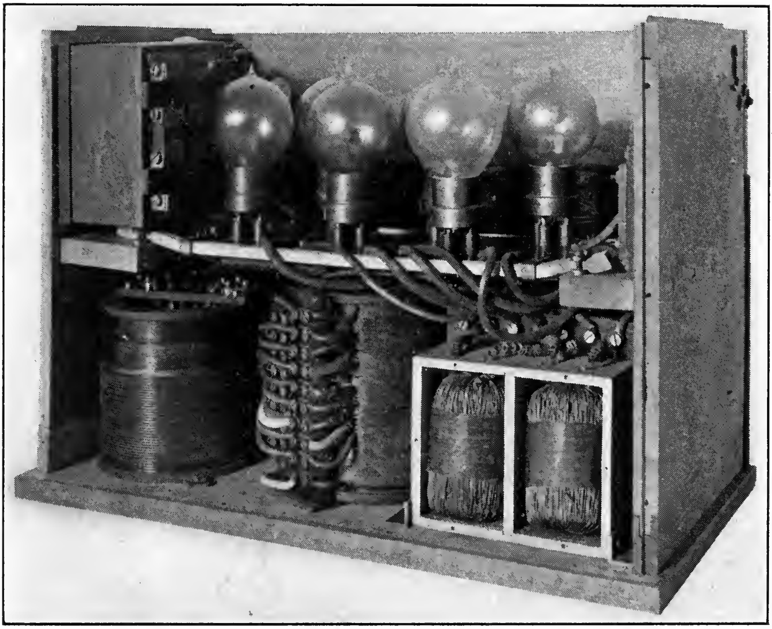


Plate 10.—A typical undamped wave two-way set of French design. Note that the six vacuum tubes used as detector, oscillators and amplifiers are all identical. The amplifier iron core transformers are seen at the lower right hand corner, and at the left are the transmitting and receiving inductances.

Representing these conditions graphically in Fig. 201, the grid-potential plate-current characteristic curves of the tube are given for plate potentials of 200, 250, 300, 350 and 400 volts. Assuming that the alternating emf. induced in coil L_p produces sinusoidal plate potential variations between the limits of 250 and 350 volts, the operating point P of the tube with a plate battery emf. of 300 volts will alternately move up and down between the points A and B of the ordinate corresponding to a grid potential of zero volts. The resulting pulsating plate current is plotted against time at the right of the characteristic curves for the period of time Ot_1. During the same period of time the grid potential remains zero, since no speech is impressed on the microphone. This is plotted downward against time.

Suppose now that the operator talks into the microphone T, Fig. 200, and thus produces variations of the grid potential, as represented in function of time by the curve C of Fig. 201. This will correspondingly shift the ordinate along which point P is oscillating. The resultant plate current is shown by the solid line curve D between the time t_1 and t_2. From this curve it is seen that the effect of the grid potential variation due to the speech is primarily to vary the direct current component of the plate current. On the other hand, the variation or modulation of the alternating plate current component, which is finally the only one effective in modulating the antenna output, is due to the fact that the distance AB does not remain constant as the grid potential is varied. It is therefore largely a function of the plate-voltage plate-current characteristic of the tube.

The component of the antenna current which is spent in the radiation resistance of the antenna and is therefore representative of the radiated energy, is the difference between the total current generated by the alternator, Fig. 200, and the alternating current component of the plate current of the modulator tube M. This radiation resistance component is represented in the lower right hand, curve of Fig. 201, which may also be taken as representing the current in the receiving antenna, by substituting a suitable antenna current scale. This received current, after rectification by an ordinary detector, will then have an envelope similar to the microphone current in the transmitting circuit, and will thus reproduce the speech in the telephones of the receiving circuit.

This absorption method of modulation, although simple and effective, is not as good as Heising's direct current method of modulation studied previously. One difference between the two is that in the direct current modulation methods, the high frequency antenna current is always symmetrical, which is not the case in the absorption method, as may be seen from Fig. 201.

In general vacuum tube oscillator and modulator circuits are of low or medium power, so that some method of amplification is required for long distance transmission. The best method is probably to generate undamped oscillations by means of a low-

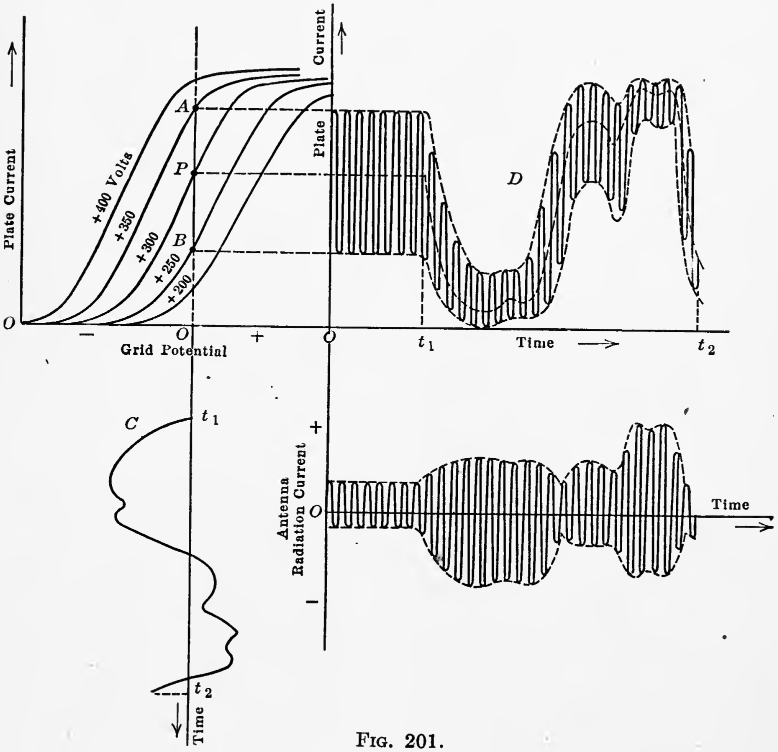


Fig. 201.

power vacuum tube, modulate these oscillations by means of a similarly low-power vacuum tube modulator, such as described above, and then amplify the modulated high frequency oscillations by a vacuum tube cascade or pyramid radio frequency amplifier before impressing them upon the antenna circuit. The important factor here is obviously that the oscillations be amplified without distortion. This consideration is taken up in a later esction of this chapter.

Ferromagnetic Methods of Modulation.—As mentioned previously, vacuum tube modulation although theoretically practicable for any amount of power, becomes quite cumbersome when the amount of power to be controlled is large, for example when the antenna is excited by a high frequency alternator of 50 to 200 kw. capacity for long distance work. The defect of the vacuum tubes here is that since each tube handles only a small amount of power, it becomes necessary to connect them for pyramid amplification and to use a large number of tubes.

For long distance work then, another type of modulator known as the ferromagnetic modulator is used. This name is derived from the fact that use is made of the magnetic properties of iron. Only two of the several makes of magnetic modulators in use will be described here, as illustrating the principles involved; namely, the Alexanderson magnetic amplifier and the Hartley system, the latter being derived from the first. Like the vacuum tube modulator the ferromagnetic modulator is used primarily to amplify the variable resistance reaction of an ordinary microphone telephone transmitter, which is thus made to control considerable amounts of power.

The Alexanderson magnetic amplifier was primarily designed for use with the Alexanderson high frequency alternator, the principle of which is given in Chapter V.[1] A schematic diagram of a radio telephone transmitting circuit comprising this apparatus is given in Fig. 202. This shows a high frequency alternator A, connected in the usual manner in series with the transmitting antenna in which an undamped high frequency alternating current is thereby established. Shunting the alternator is an impedance made up of two coils L_1 and L_2. These are wound in opposite directions on two iron cores, and connected in parallel. Their magnetic fields are therefore at each instant in opposite directions, as shown by the arrows, which condition prevents any high frequency currents from being induced in the coil L_3, the

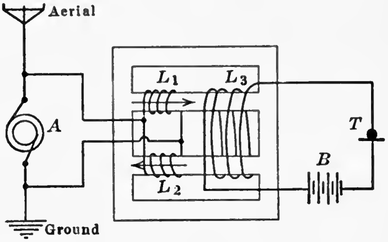


Fig. 202.

[1] See *Radio Telephony* by A. N. Goldsmith for full theory of the Alexanderson amplifier.

function of which is explained later. It is thus seen that the antenna circuit and the coils L_1 and L_2 are connected in parallel across the alternator terminals, so that the alternating current generated by the alternator divides between these various circuits in inverse proportion to their respective impedances.

The impedance of the antenna circuit, determined by its constants, is itself constant; and if the circuit is tuned to the alternator frequency, this impedance will simply be the resistance of the antenna circuit. If then the impedance of the coils L_1 and L_2 is in some manner altered, the distribution of the alternator current between these coils and the antenna will be correspondingly changed. This is effected by means of the third coil L_3, wound over the two iron cores of the coils L_1 and L_2, and connected as shown, to a battery B and a microphone T, which may be inserted in the circuit either directly or through a vacuum tube amplifier. The amplifier, however, is not required in many cases. This coil L_3 is made up of a large number of turns, so that even a small current of the order of 1 amp. furnished by the battery B will create a strong magnetic field in the iron cores. The effect of talking into the microphone T is then to vary the direct current set up by the battery B in the coil L_3 and therefore to vary the magnetization of the iron core. And since the permitivity of iron is a function of its magnetization, especially about the magnetic saturation point at which the device is operated, it follows that the reactance of the coils L_1 and L_2 is correspondingly varied, and modulation of the alternator output is accomplished.

If properly designed, this device can be made to have a very large amplification factor and the small amount of power controlled in the coil L_3 can be made to control a considerably larger amount of power in the coils L_1 and L_2. This can be readily understood by considering that the reactance of the coils L_1 and L_2 is directly proportional to the permitivity of the iron, which is in turn controlled by the ampere-turns of the coil L_3. There is thus a multiplication of the effect of the microphone in its control of the reactance of coils L_1 and L_2. This reactance variation combined with the alternator characteristics affords a considerable degree of amplification. It has thus been possible by producing a variation of 0.2 amp. in the current of coil L_3 to vary the antenna power from 5.8 kw. to 42.7 kw.[1]

[1] A. N. Goldsmith, *Radio Telephony*, p. 203.

This result was obtained, however, with an improved type of magnetic amplifier, illustrated in Fig. 203, which differs from the one of Fig. 202 by the addition of the four condensers shown. These condensers are both for the purpose of better adapting the amplifier to the alternator characteristics, thus giving greater stability to the modulating system, and also of providing greater sensitiveness and more effective control. Thus, the introduction of condenser C_1 in series with the windings L_1 and L_2 partly compensates for the inductive reactance of these coils. The impedance variations due to the control field variations in coil L_3 then become more of the nature of an "in-phase" or resistance variation. The function of the condenser C_4, which shunts the coils L_1 and L_2 is to form with the latter an oscillatory circuit of near resonance frequency. The variations of the permitivity of the iron cores of coils L_1 and L_2, by varying the inductances of the coils, produce then a very much greater resistance reaction variation, since the operation is effected along the steep slope of the resonance curve peak.

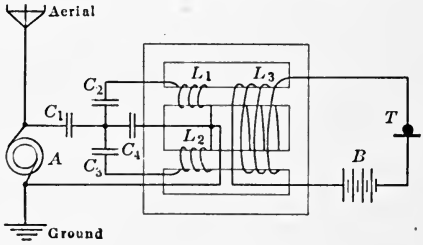


Fig. 203.

Finally, the condensers C_2 and C_3, each of which is in series with one of the coils L_1 and L_2, have a capacitance small enough to introduce a high audiofrequency (or speech frequency) reactance, though a low radio frequency reactance. Their use is necessary because in the amplifier of Fig. 202, the coils L_1 and L_2 are short circuited and in series, whence the current variations in coil L_3 could induce audio frequency currents in coils L_1 and L_2 and thus prevent effective control of the alternator current.

An objection to the ferromagnetic modulator of the type described and to other similar modulators, is the distortion of the modulated high frequency alternating current flowing in the coils L_1 and L_2. It is due to the hysteresis in the iron cores and may be quite considerable at radio frequencies. In order to obviate this defect, a modification of the above modulation scheme has been devised and is schematically represented in Fig. 204.[1] The basic principle, previously made use of in the magnetic detector

[1] U. S. Pat. 1287982 issued to R. V. L. Hartley.

(see Chapter IV), is that the hysteresis effect, which is due to a lag of the iron molecules in moving under the effect of an alternating field, is partially or totally destroyed if the molecules are kept in constant vibration or motion.

To this end, the modulator of Fig. 204 which is a modification of that of Fig. 202, is provided with an additional cross-magnetizing winding. This consists of a number of turns of fine wire wound within the iron core as shown in the figure, and connected to an auxiliary high frequency alternator A' which may of course be a vacuum tube oscillator, since only a small amount of power is

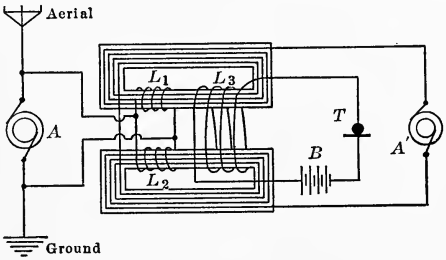


Fig. 204.

required. It is thus seen that the current in the coils L_1 and L_2 is flowing at right angles to that in the cross-magnetizing coil. The iron molecules are thus kept in constant motion by two independent alternating magnetic fields acting 90 degrees apart with the result that hysteresis is considerably reduced. The operation of this modulator is similar to the Alexanderson modulator just described. The microphone T may of course be inserted in the control circuit L_3 through a vacuum tube amplifier if large control currents are needed in that circuit.

RADIO TELEPHONE RECEPTION

In the previous paragraphs it was shown that the current in the radio telephone transmitting antenna is a high frequency alternating current of varying maximum amplitude, the variations following the current variations of the microphone or telephone transmitter circuit when actuated by the voice. This modulated high frequency antenna current sets up in space similarly modulated alternating electromagnetic and electrostatic fields, which in turn induce in the receiving antenna a modulated high frequency current which is a proportional reproduction of the current in the transmitting antenna. An idea of the shape of the received current may be had from the lower curve of Fig. 190.

If this current is rectified by a suitable detector, such as a crystal or vacuum tube, only the negative or the positive half-cycles will be permitted to flow in the receiving circuit. The current thus rectified will have an envelope and an average value which vary in proportion to the microphone current at the transmitting station. The received rectified current will therefore reproduce the speech when made to flow through telephone receivers.

It follows from the above that in a radio telephone receiving circuit, the only modification of the received current required in order to make the signals audible in a telephone receiver, is rectification by means of a suitable detector. Damped wave radio telegraph circuits making use of such detectors can therefore be used without alteration for radio telephone reception.

It is important to note, however, that radio telephone signals do not permit of very sharp tuning of the circuits. This is due to the fact that, through the combination of radio and audio frequencies, the transmitting antenna does not radiate a wave of constant frequency, but sends out a wave the frequency of which varies above and below the radio frequency by an amount equal to the impressed audio frequency. This may be explained mathematically as follows:[1] Let

$$I \cos 2\pi ft$$

represent the radio frequency alternating current generated in the transmitting antenna when no speech is impressed on the microphone. To simplify the explanation, suppose that a pure note of frequency F is impressed upon the microphone, thus producing a variation

$$kI \cos 2\pi ft$$

in the amplitude of the antenna current. The resultant antenna current is then equal to the product

$$(1 + \sin 2\pi Ft)\ (I \cos 2\pi ft) = I \cos 2\pi ft + \tfrac{1}{2} kI \sin 2\pi\ (f + F)t - \tfrac{1}{2} kI \sin 2\pi\ (f - F)t$$

[1] E. B. Craft and E. H. Colpitts, *Proceedings American Institute of Electrical Engineers*, February, 1919.

which is also proportional to the current in the receiving antenna. This shows a component of radio frequency, $I \cos 2\pi ft$, which is inaudible, and two components of respective frequencies $(f + F)$ and $(f - F)$. In practice, the audio frequency F of the speech seldom exceeds 3000 cycles per second. The transmitting and receiving circuits must therefore be capable of oscillating between the limits $f + 3000$ and $f - 3000$ cycles per second. It is thus seen that tuning will be sharper, the greater the frequency of the radio or "carrier" wave used since the per cent variation will then be smaller.

A consequence of the broad transmitted wave is that the resonance curve of the tuned receiving circuit is rather flat—more so than in radio telegraphy, where only one frequency is involved. This permits, therefore, the use of receiving oscillatory circuits having a comparatively high resistance. It will be shown in the next section that such a resistance is not only permissible but is, within limits, a requisite for satisfactory radio telephone reception. This of course does not apply to radio telegraph reception.

Broad tuning is not the only limitation of radio telephony as compared to radio telegraphy. Another consideration is that for a given distance between the transmitting and receiving stations, a greater amount of power is required for radio telephony. This is readily explained. In radio telephony, the antenna current is modulated or varied above and below its normal value. Thus, suppose the antenna current flowing in the transmitting antenna is just sufficient to produce barely audible radio telegraph signals in the receiving circuit of the distant station. Now if the same current is modulated above and below its normal value in transmitting speech, all parts of the speech which correspond to currents below the normal value will be too weak to be heard at the receiving station, and the message will not be understood.

Distortion of Speech.—All of the above radio telephone circuits, both for transmitting and receiving, are limited in their range of operation in that they will give a true reproduction of the speech over only a limited range of the possible values of current and voltages available in the sets. Beyond this range, the speech wave is distorted or even entirely obliterated.

When speech is impressed on the telephone transmitter, the diaphragm vibrates under the effect of the sound waves impinged upon it, and these diaphragm vibrations are of the same frequency and wave shape as the sound waves, and of a proportional amplitude.[1] The diaphragm vibrations in turn produce corresponding variations of the microphone resistance, and therefore variations of current and potential in the microphone circuit which are true, undistorted electrical reproductions of the sound waves set up by the voice.

In order to obtain a perfect speech reproduction, it is necessary that the currents set up in the telephone receiver circuit at the receiving station be exactly and in all details proportional to the current in the microphone circuit at the transmitting station. It is therefore essential that no distortion whatever of the microphone current variations be permitted to occur in the amplification of the microphone current, in the modulation of the generated oscillations, in the transfer of these modulated oscillations to the transmitting antenna circuit, in the transmission of the waves between the transmitting and receiving stations, in the transfer of energy between the receiving antenna and the detector and amplifier circuits coupled thereto, and finally in the rectification and amplification of the received signals. If at any one of these stages a distortion of the speech is allowed to take place, the sound at the receiving station will not be a true reproduction of the voice at the transmitting station, and the speech will be distorted to a greater or less extent and may even become entirely unintelligible. The engineering which is necessary at each stage is therefore taken up briefly in the following paragraphs.

Amplification of the microphone current variations before modulation of the oscillations was shown to be necessary only in case the modulating system requires a comparatively large amount of control energy for complete modulation of the undamped oscillations. This occurs when the modulating system, which is essentially an amplifier, has too small an amplification factor of its own. Taking the case where a vacuum tube amplifier is thus used before modulation, the grid potential variations which are produced by the microphonic currents, should not be so great as to bring the operating point of the amplifier beyond or even in the vicinity of the upper and lower bends of its characteristic curve. For if grid potential variations occurred beyond the

[1] This of course is on the assumption that the diaphragm has only a very dully defined resonance period, and it is only true below a certain value of sound intensity.

upper bend, for instance, where the grid-potential plate-current curve is almost horizontal, there would be no corresponding plate current variations, and the plate current would then not be a true reproduction of the microphonic current. It is thus seen that the operating point of the amplifier must remain in the central part of the curve which may be considered as a straight line and where the output is proportional to the input. Similar considerations hold for amplification of the received signals at the receiving station.

In a similar way, the impedance variations of the modulator, both in the case of a vacuum tube modulator and of a ferromagnetic modulator, must be proportional to the microphonic currents. This relation is, however, somewhat altered if the system generating the oscillations to be modulated has not a strictly straight line characteristic.

Considerations of another order govern the distortion of speech during the transfer of energy between coupled circuits, that is, between the generating-modulating circuit and the transmitting antenna circuit, between the transmitting and receiving antenna, and between the receiving antenna and the circuits to it. As a general rule, it may be said that where speech-modulated energy is being transferred from one circuit to another, there will be no distortion of the speech as a result of the transfer if the circuit receiving the energy has a decrement greater than, or, as a limiting case, equal to that of the circuit radiating the energy. Suppose that the transmitting antenna is excited by modulated high frequency oscillations set up in a primary circuit coupled to it, and that this primary oscillating circuit has a decrement greater than that of the antenna circuit. The distortion which would result may be explained as follows:

When speech is impressed by the voice on the telephone transmitter diaphragm, the undamped oscillations generated at the transmitting station by the oscillator system become modulated, as previously explained. In other words, the oscillations instead of being undamped and of zero decrement, increase and decrease in amplitude, and are thus given a decrement (positive or negative), the instantaneous value of which varies according to the speech modulations.

In Chapter II it was shown that when a damped oscillating current flowing in a primary oscillatory circuit is made to induce a current in another (secondary) tuned circuit coupled to it, the induced or secondary oscillating current has the same decrement as the primary current provided the secondary circuit has a decrement greater than or equal to that of the primary oscillation. If the secondary circuit, however, has a smaller decrement, then the current induced in it will decrease at a slower rate than the primary current, and the envelope of its time-current curve will have a smaller slope. Applying these principles to radio telephony, it follows that if at any time the undamped or carrier oscillation is so modulated by the voice that its instantaneous decrement is greater than the decrement of the circuit to which energy is being transferred, the modulated alternating current induced in the latter will not follow the original modulations, but will have a maximum decrement equal to that of the circuit; that is, smaller than that of the original modulated current. This will cause a smoothing out of the induced current variations and a corresponding distortion in the reproduction of the voice articulations.

The practical conclusion is that in order to avoid this type of distortion, it is necessary to make the various transmitting and receiving circuits all of the same decrement. Or, if this is not possible, to make them of increasing decrement, as taken in succession in the order they are reached by the modulated radio frequency energy wave. Also, the minimum decrement allowable for any one of the circuits should be greater than the greatest instantaneous decrement of the speech modulated wave.

Now it was shown that the logarithmic decrement of a circuit is equal to $\pi R\sqrt{\frac{L}{C}}$, where R, C and L are respectively the resistance, capacitance and inductance of the circuit. In order to have a high decrement, one may then increase the resistance R of the circuit. This, however, is objectionable as it increases the energy losses in the circuit. The other solution is to decrease the inductance L and increase the capacitance C. This must be done in such proportions that the product LC will remain unchanged, in order to maintain the circuit in tune for the signal frequency. Also, this cannot be done indefinitely, as an excess of capacitance will produce an effect opposite to that of the inductance, resulting in reverse speech distortion. Furthermore, the use of high decrements prevents sharp tuning of the circuits. All these considerations must then be balanced against each other when designing a circuit.

Finally, during the travel of the waves through the ether from the transmitting to the receiving station, voice modulations of high frequency or pitch are absorbed to a greater extent than those of lower frequency, with the result that the speech in the telephone receivers will be of unequal volume and may be hard to understand. It is therefore helpful in radio telephone communication to speak in an even and rather low tone of voice.

The above discussion on distortion is far from complete and merely points to several of the more important sources of speech distortion. It is impossible in a book of this scope to enter into greater details, but this subject is one requiring careful consideration in designing actual circuits.

MULTIPLEX RADIO TELEPHONY

The impossibility of obtaining sharp tuning in radio telephony for the reasons which have been explained, makes it difficult to carry on without serious interference a large number of radio telephone conversations in a restricted area, since the wave lengths of the various stations must differ widely, even to a greater extent than with damped wave radio telegraphy. Certain requirements arise, however, where the working of a large number of sets in a limited area is highly desirable, for example under such conditions as were experienced during the war in squadron work of submarine chasers. This need brought out the development of a method of multiplex radio telephony.[1]

This consists essentially in having several stations transmit at the same wave length, which must be chosen short, of the order of 150 meters, corresponding to an oscillation frequency of 2,000,000 cycles per second. At each station this radio frequency wave is modulated, not at audio or speech frequency, but at another radio frequency different for each station and lower than the common radio frequency. Thus, with say three stations radiating at 150 meters (2,000,000 cycles), the first station will modulate this 150-meter wave at a frequency of 25,000 cycles, the second station at 35,000, and the third at 45,000 cycles per second. These modulation waves, although of lower frequency than the carrier wave, are still of a frequency beyond that of audibility. At each station, the individual sta-

[1] E. B. Craft and E. H. Colpitts, *Proceedings Amer. Inst. of Elect. Engrs.*, March, 1919.

tion waves may therefore be themselves modulated by the voice at an audio frequency. In other words, at each station the speech vibrations are made to modulate radio frequency oscillations of a frequency characteristic of the station, which are in turn made to modulate oscillations of a higher frequency, the latter frequency being common to a number of stations.

This may be achieved in many different ways, one method

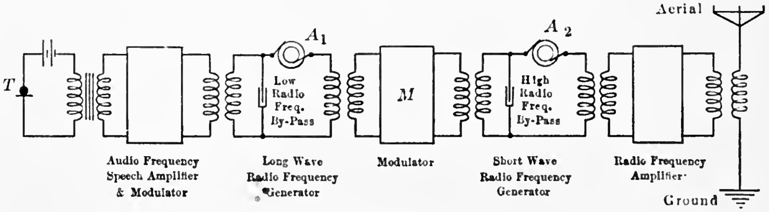


FIG. 205.

being illustrated in Fig. 205. The telephone transmitter T is made to modulate in the usual manner the radio frequency alternating current generated by the alternator or vacuum tube oscillator A_1. The frequency of these oscillations is, in the present example, 25,000, 35,000 or 45,000 cycles per second. The oscillations thus modulated are then made to operate the control circuit of a modulator M which modulates the high frequency

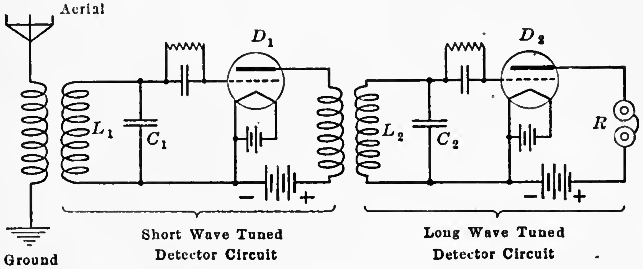


FIG. 206.

alternating current generated by the alternator or vacuum tube oscillator A_2. In the present case the frequency of these oscillations is 2,000,000 cycles per second. These double modulated high frequency oscillations are then delivered to the antenna circuit and radiated into space.

A receiving circuit, having an antenna and detector circuit tuned to 150 meters will then receive all three stations. Thus, if the antenna circuit of Fig. 206 and the tuned grid circuit L_1C_1

of the detector tube D_1 are tuned to 150 meters, the detector tube, in rectifying the 2,000,000-cycle incoming waves, will give a pulsating current in its plate circuit, the alternating-current component of which will be a 25,000, 35,000 or 45,000 cycle voice modulated current. By coupling this plate circuit to a second detector tube D_2 having a suitable tuned input circuit L_2C_2, it will then be possible to pick out any one of these alternating current components, which, upon rectification, will reproduce in the telephone receiver R the telephone conversation spoken at the corresponding transmitting station.

Two methods of adapting the circuit of Fig. 205 to vacuum tube circuits are illustrated in Figs. 207 and 208. These represent only two special cases of a large number of possible arrangements.

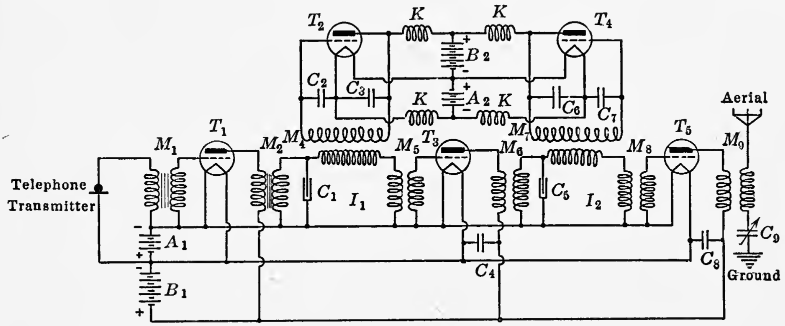


Fig. 207.

In the circuit of Fig. 207, use is made of vacuum tube oscillators and of vacuum tube modulators connected for modulation of the radio frequency oscillations by the absorption method. The speech impressed on the telephone transmitter, Fig. 207, causes resistance variations of the latter and sets up corresponding fluctuations of the current flowing in the primary winding of the transformer M_1. Amplified reproductions of these current variations are set up in the plate circuit of the amplifier tube T_1, and therefore in the primary of the transformer M_2. These are in turn made to vary the grid potential of the tube T_3 after being transferred through the intermediate circuit I_1 and transformer M_5. By setting up audio frequency variations of the grid potential in the tube T_3, the result of the telephone transmitter speech vibrations is to vary in a corresponding manner the internal plate resistance of the tube.

On the same intermediate circuit I_1 are impressed the long wave radio frequency undamped oscillations generated by the tube T_2 suitably coupled by the condensers C_2 and C_3, and coupled to the circuit I_1 through the oscillation transformer M_4. These oscillations are then transferred from the circuit I_1, to the grid of the modulator tube T_3 through the air core transformer M_5, producing corresponding radio frequency pulsations of the plate current of the tube T_3. The amplitude of these pulsations, however, depends upon the position of the operating point of the tube T_3 on its characteristic curve, or, which amounts to the same thing, upon the internal plate resistance of the tube T_3. This internal resistance is varied at audio frequency as a result

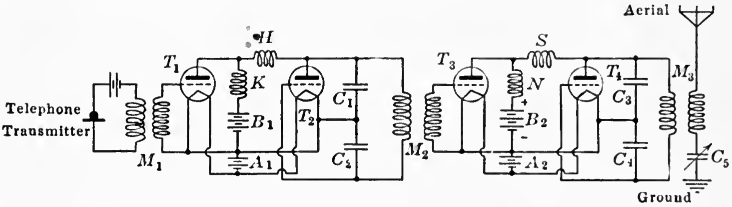


FIG. 208.

of the audio frequency grid potential variations due to the telephone transmitter vibrations. The radio frequency alternating current component of the plate current of tube T_3 is thus of the frequency (approximately) of the oscillations generated by the tube T_2, and the amplitude of this current is modulated in accordance with the audio frequency speech variations of the telephone transmitter.

This modulated long wave radio frequency alternating current is made to set up corresponding grid potential variations in the tube T_5, after transfer to the intermediate circuit I_2. Through this same intermediate circuit the short wave radio frequency oscillations generated by the tube T_4 are impressed on the grid of the tube T_5. The alternating current component of the plate current of the tube T_5 will then be (approximately) of the frequency of the oscillations generated by the tube T_4 in the circuit $C_6C_7M_7$, and of an amplitude which is varied or modulated by the potential oscillations due to the oscillator tube T_2. These potential oscillations are of long wave radio frequency, and are furthermore modulated at audio frequency by the telephone transmitter speech vibrations. It follows that the alternating plate current of tube T_5, and therefore the antenna current and radiated oscillations, are short wave radio frequency currents, twice modulated.

The condensers C_1, C_4, C_5 and C_8 are radio frequency by-pass condensers offering low impedance shunt paths for the high frequency oscillations, which would otherwise be choked out by the high impedance of the transformers M_2 and M_6, or the resistance of the battery B_1. It will be noticed also that the d.c. filament and plate circuits are respectively in parallel, which reduces the number of batteries required to energize the circuit. The two oscillator tubes T_2 and T_4 are fed from common batteries, and the choke coils K prevent the oscillations generated by either tube from flowing through the battery, or from affecting the other tube circuit.

The circuit of Fig. 208 is another and perhaps better adaptation of the same principle. Use is made in this circuit of the d.c. power modulation method already explained. The long wave radio frequency oscillations generated by the tube T_2 are modulated at audio frequency by the telephone transmitter vibrations through the medium of the modulator tube T_1 connected in parallel with the tube T_2 across the source of direct current energy B_1. These modulated oscillations are then impressed through the transformer M_2 upon the modulator tube T_3, which correspondingly modulates the energy input of the short wave radio frequency oscillator tube T_4. The modulated oscillations are then transferred to the antenna through the oscillation transformer M_3. The coils K and N are audio frequency choke coils, that is, coils of large inductance. The coils H and S are, respectively, long and short wave radio frequency chokes. The functions of these various coils were explained in a previous part of this chapter.

CHAPTER XI

DIRECTIONAL RADIO AND LOOP ANTENNÆ

Underlying Principles

It was shown in Chapter III that an antenna of suitable shape will radiate or receive energy in certain directions more strongly than in others, whereby it is possible, to a certain extent, to direct the waves between two stations. The same object may be attained by other methods which are of especial interest because they provide very pronounced directional effects and make use of transmitting and receiving circuits which may be of very small linear dimensions. The basic principle for transmitting is the creation in various parts of the radiating circuit of high frequency currents having such phase relations as to set up electric fields which add in certain directions of space and neutralize each other in other directions. For receiving, the same principle of phase difference is made use of.

In order to understand the underlying principles, a method of transmission now little used, which is known as the Blondel system, will first be studied. Consider two single-wire vertical antennæ AB and $A'B'$, Fig. 209, of equal length and therefore of identical electrical constants, oscillating at the same wave length λ, and separated by a distance d equal to $\lambda/2$. Temporarily neglecting the antenna $A'B'$, it will be recalled that the electric field set up directly around the antenna AB by the current flowing in it, is in synchronism with the field at all points at a distance from the antenna equal to λ, or a multiple of λ, that is, 2λ, 3λ, etc. Also, as a consequence of this, the field at a distance $\lambda/2$ (or $\frac{3}{2}\lambda$, $\frac{5}{2}\lambda$, etc.) from AB is at every instant 180 degrees out of phase with the field at the wire AB. Then if the alternating current in the antenna $A'B'$ is always 180 degrees out of phase

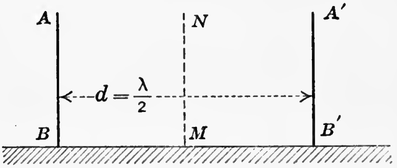


Fig. 209.

with that in the antenna AB,[1] it will produce in the plane $ABB'A'$ a field which is always in phase with the field produced by the current in AB. It will then be seen that under such conditions, the fields of the antennæ will add at all points of the plane containing the two antennæ. On the other hand, at all points equidistant from the two antennæ, that is, at all points of the plane at right angles to that of the two antennæ and located midway between them, the fields of the two antennæ will be exactly equal and opposite, and will neutralize each other. Energy is thus radiated by the system with maximum amplitude in the plane of the two antennæ, and with zero amplitude in the plane at right angles to this and midway between the antennæ.

Now if the two antennæ are separated by a distance d smaller than $\lambda/2$, but kept oscillating 180 degrees out of phase, similar effects will take place, with the difference, however, that the resultant maximum field will be less than in the previous case. This may be demonstrated mathematically as follows:[2]

Assume an antenna MN, Fig. 209, to be located midway between the two antennæ AB and $A'B'$. This antenna would produce at point P of the plane $ABB'A'$ an electromagnetic field proportional to the hypothetical current $I_0 \sin 2\pi ft$ flowing in it. The field set up at point P by the antenna $A'B'$, which is nearer to it by the distance $d/2$ is then proportional to

$$I_0\, sin\left(2\pi ft + \frac{\pi d}{\lambda}\right)$$

and the field due to AB is proportional to[3]

$$-\,I_0 \sin\left(2\pi ft - \frac{\pi d}{\lambda}\right).$$

The field at point P resulting from the currents in the antennæ AB and $A'B'$ is then proportional to their algebraic sum

$$I_0\left[\sin\left(2\pi ft + \frac{\pi d}{\lambda}\right) - \sin\left(2\pi ft - \frac{\pi d}{\lambda}\right)\right] = 2I_0 \sin\frac{\pi d}{\lambda}\cos 2\pi ft$$

[1] This phase relation of the currents in the two antennæ may be achieved by insulating their lower ends from the ground and connecting them through a wire connected or coupled to the oscillation generating circuit, and comprising an inductance coil of such value as to produce a lag of the current of 180 degrees.

[2] See C. Tissot, *Manuel Elémentaire de Télégraphie sans Fil*, Paris, 1918.

[3] The currents in AB and $A'B'$ are 180 degrees out of phase as explained before, and hence are of opposite signs at any given instant. This explains the minus sign before the next expression.

This expression, and hence the field, is a maximum when $d = \frac{\lambda}{2}$, and gradually decreases as d decreases or increases from this value.

Now consider the closed rectangular metallic circuit or loop $ABB'A'$, Fig. 210, in which a high frequency alternating current is made to flow. At every instant, the current in the vertical branch AB flows in a direction opposite to that in the branch $B'A'$. The condition is therefore similar to that of two vertical antennæ in which the currents are 180 degrees out of phase, and it follows, as was shown above, that such a closed loop circuit will have a directional effect giving maximum energy radiation in the plane of the loop, and zero radiation in a plane at right angles to that of the loop. The radiation will however be rather small, as the loop is generally of small dimensions as compared with the wave length. That is, the ratio $\frac{d}{\lambda}$ is in general a small one. It is seen, however, that radiation increases as the wave length decreases, since then $\frac{d}{\lambda}$ increases and d tends to become more nearly equal to $\frac{\lambda}{2}$.

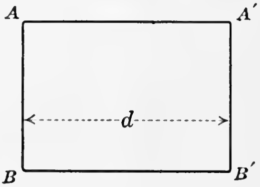


Fig. 210.

The most common method of setting up a high frequency alternating current in a loop of the kind just described is to use the loop as part or all of the inductance of an oscillatory circuit. Fig. 211 represents a simple damped wave loop transmitting circuit. The oscillatory circuit is made up of the loop L, condenser C and spark gap G, and is excited by connecting it in the usual way to a source of high alternating or pulsating potential. In the circuit here shown, the condenser C is charged periodically by the secondary S of an induction coil, the primary winding P of which is energized by the battery B when the key K is closed, the primary current being rapidly interrupted by the vibrator V. The condenser D serves merely to quench the spark at the vibrator, and thus bring about a quicker break of the primary current.

The circuit of Fig. 212 is an undamped wave loop transmitting circuit. This is essentially the same as the circuit of Fig. 172, and represents a three-electrode vacuum tube, coupled electrostatically for oscillation generation. The inductance L of the oscillatory circuit is at the same time the loop antenna.

In actual practice the loop for transmitting purposes is made of one or a few turns of heavy wire, generally less than ten turns. Its linear dimensions seldom exceed a few feet and its shape may obviously be square, rectangular, triangular, circular, or any other closed shape that may be convenient. The great advantages of such loop antennæ are their small linear dimensions, sharp directional characteristics and ease of orientation. It is simply necessary to turn the loop to change the direction of trans-

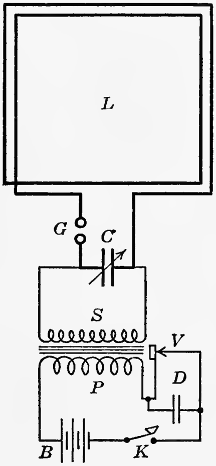


FIG. 211.

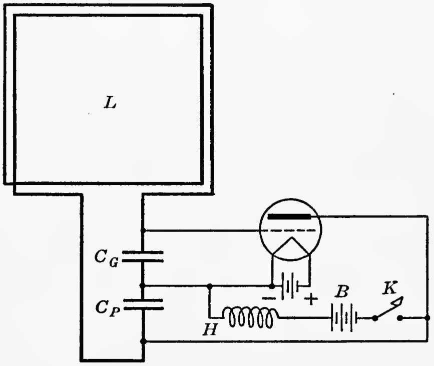


FIG. 212.

mission of the waves. The disadvantage is that short waves are generally required for efficient operation for the reasons previously given, and this may reduce the range of transmission because of the greater absorption losses of short waves in their travel through the air.

It should be noted that the field of a loop antenna is different from that of the two associated vertical wire antennæ of Fig. 209, in that the condenser of the loop oscillatory circuit is of such small physical dimensions that the electrostatic field extends out into space over a very short distance only, and radiation is therefore effected almost entirely by the electromagnetic field of the circuit. This must be taken into account when designing a receiving circuit which is to receive signals specifically from loop transmitting circuits. In that case the receiving circuits must preferably be highly sensitive to electromagnetic field variations rather than electrostatic field variations.

Loop antennæ may be used equally well or even better for receiving signals, and their use in this connection has rapidly increased since the introduction of receiving vacuum tube amplifying circuits. A simple and common form of damped or modulated wave loop receiving circuit is shown in Fig. 213. It consists simply of a loop antenna *L*, forming the inductance of an oscillatory circuit which is tuned to the proper wave length by means of a variable condenser *C* connected across the loop terminals.

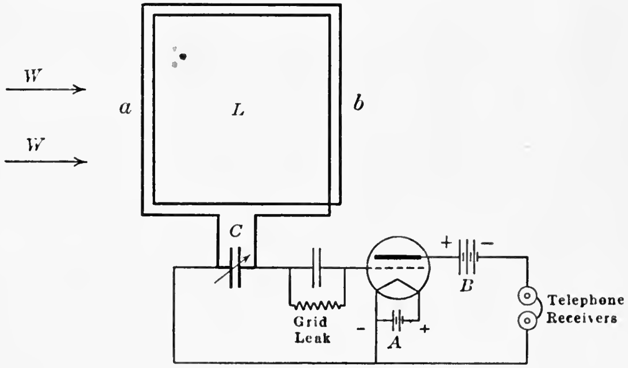


Fig. 213.

These terminals are connected to a detector and telephone receiver circuit. In the circuit here shown, a vacuum tube detector is used, although in general a multi-stage amplifier is required to obtain signals of sufficient intensity, especially if a loop of small dimensions is used.

A loop antenna thus connected may be used to receive signals sent out by either an ordinary antenna or a loop antenna. In either case, however, it responds almost exclusively to the electromagnetic field variations set up by the transmitting antenna and is only faintly sensitive to the electrostatic field variations.

The loop antenna has directional receiving characteristics which are similar to its transmitting characteristics; that is, signals are received with a maximum intensity when the plane of the receiving loop is parallel to the direction of travel of the waves, and the signals are received with zero or minimum intensity when the plane of the loop is at right angles to that direction.

To explain this, suppose the loop L, Fig. 213, to be in a plane parallel to the direction of motion of the waves to be received, which direction is shown by the arrows W. Then, the side a of the loop will be reached by the wave before the other side b, and all reversals of the field due to the transmitting antenna will occur at a a fraction of time before they take place at b. In other words, the emf. induced across the conductor a will be out of phase with, and lead the emf. induced in b, thus setting up a current in the loop circuit. The effect which is dependent on the difference of phase due to the space displacement of the two sides of the loop, is greater, the greater the distance between a and b (as measured along the direction of travel of the waves) and the shorter the wave length received.

Now if the plane of the loop is turned so as to make an angle with the direction of travel of the waves, the difference of time between the arrival of the wave at a and b will be smaller, and the current set up in the loop will be correspondingly reduced. If the loop is perpendicular to the direction of travel of the waves, then all the points of the loop will be reached by the wave at the same time and no current will be induced in it. No signal is then received. These directional properties are made use of to determine the direction of travel of the waves and thereby locate transmitting stations. When used for this purpose, the loop is known as a "goniometer" or radio compass.

The small linear dimensions of loop antennæ make it generally necessary to use high power multi-stage amplifiers in order to obtain legible signals. By means of such an amplifier and a small loop, it is possible to receive signals from distant stations just as easily as when using a large antenna of the ordinary open (or condenser) type.

Design of Receiving Loop Antennæ

In order to study the various factors affecting the design of a receiving loop antenna, it is well to express mathematically the effect of the incoming waves on the loop circuit.[1] The emf. e induced in the loop by the alternating flux Φ set up about the loop by the current in the transmitting antenna is, at each instant, expressed by the relation

$$e = \frac{d\Phi}{dt}$$

[1] See *Electrical World*, Vol. 73, No. 10, March 8, 1919, pp. 464–467.

from which it follows that the effective emf. induced in the loop is

$$E = \sqrt{2}\,\pi fNAH \qquad (36)$$

where N = number of turns of the loop
A = cross-sectional area of loop
H = effective field intensity
f = frequency

This alternating emf. will set up a current I in the loop circuit and if this circuit is tuned to the frequency f, its impedance will reduce to the effective resistance R of the loop, and the current will be expressed by the equation

$$I = \frac{E}{R} = \frac{\sqrt{2}\pi fNAH}{R} \qquad (37)$$

As previously explained, the energy thus set in motion in the receiving loop circuit is made to act upon a telephone and detector circuit connected across the tuning condenser C, Fig. 213. In the case of a vacuum tube detector connected as shown, the intensity of the received signals may be indicated by the effective potential E_c produced across the condenser by the oscillating current I flowing in the loop circuit. If λ is the wave length of the received signals, this condenser voltage is equal to

$$E_c = \frac{mNAL}{\lambda^2 R}, \qquad (38)$$

when m is a constant.

From this formula it is seen that, if a given loop antenna is used for receiving signals of different wave lengths, the condenser voltage E_c, and therefore the intensity of the signals, will tend to increase as the wave length is made shorter. On the other hand, the effective resistance R of the loop increases with the frequency, and therefore increases with decreasing wave length, thus having an effect on the signal strength tending to offset that of the reduced wave length. It follows that the condenser voltage and signal strength will pass through a maximum value as the wave length of the signals is gradually reduced. These relations are clearly illustrated by the curves of Fig. 214, which do not require further explanation. In general, the maximum signal intensity occurs at a wave length of two to three times the fundamental wave length λ_0 of the loop. This is the natural wave length of the loop closed upon itself without any tuning condenser, and oscillating by reason of its inductance and distributed capacitance. It is therefore the general practice when receiving signals with a loop antenna to use an antenna having a fundamental wave length λ_0 not exceeding ½ or ⅓ the signal wave length to be received.

An examination of the above formula (38) will also show that the signal strength is greater the greater the area, number of turns and inductance of the loop. However, increasing the area decreases the inductance, and increasing the number of turns, while increasing the inductance, also increases the resistance and distributed capacitance by requiring a greater length of wire.

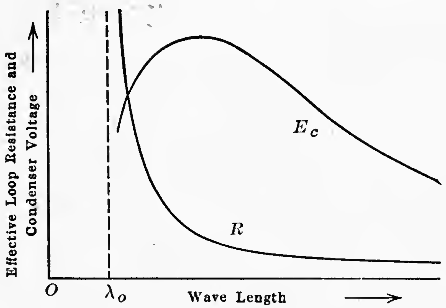


FIG. 214.

All these factors must therefore be carefully considered in the design of a loop, and the range of wave lengths desired must also be considered. In general, it may be said that a loop for short wave reception should preferably have a large area and very few turns, say not more than three or four, thus having low resistance and small inductance and distributed capacitance. On the other hand, a loop for long wave reception should be made smaller and with a greater number of turns, say at least 20 or 30.

For a given loop, the inductance and distributed capacitance increase when the turns are wound close and also when the number of turns is increased. The turns of a loop antenna should therefore be so spaced as to give a suitably high inductance without giving too great a distributed capacitance, which would limit the use of the loop to the longer wave lengths. As a rough rule, it may be said that for a given length of wire in the loop coil, the fundamental wave length of the loop can be kept the same for different loop dimensions and turn spacings. The following table is given to illustrate this rule.

Length of side of a square loop (feet)	Number of turns	Spacing of wires (inch)	Inductance (microhenries)	Capacitance (micro-mfd.)	Fundamental wave length (meters)
8	3	½	96	75	160
6	4	¼	124	66	170
4	6	¼	154	55	174
3	8	⅛	193	49	183

It is seen that the various loops described in the table have approximately the same fundamental wave length, and therefore approximately the same working range of wave lengths. The first loop, however, being the largest, would probably be preferable in view of its sharper directional characteristics, despite its smaller number of turns.[1]

Transmission Formulæ.—The above considerations covering the design of loop antennæ do not take into account the distance range over which communication is possible when using this type of radiating system. Some of the formulæ worked out by the Bureau of Standards and published by the U. S. Signal Corps which make it possible to estimate the distance of transmission and permit comparison between loop antennæ and other types of radiating circuits, are given here.

Sending with a vertical wire antenna, and receiving with a loop antenna, the current in the receiving loop is given by the relation

$$I_r = \frac{1184 h_s A N I_s}{R_r \lambda^2 d}, \tag{39}$$

where s and r refer to the sending and receiving circuits respectively, I being the current in amperes, h the height of the vertical wire antenna, A, N and R the area, number of turns and resistance of the receiving circuit, λ the wave length, and d the distance between the two stations.

Sending with a loop antenna and receiving with a vertical wire antenna

$$I_r = \frac{1184 h_r A N I_s}{R_r \lambda^2 d} \tag{40}$$

[1] See paper by Dr. J. H. Dellinger, *Proceedings A. I. E. E.*, Oct., 1919, for more exhaustive study of loop antennæ and transmission formulæ.

Sending and receiving with loop antennæ

$$I_r = \frac{7450 A_r A_s N_r N_s I_s}{R_r \lambda^3 d} \tag{41}$$

These formulæ, although readily established, are not derived here.

RADIO GONIOMETRY

Single Loop Goniometer.—Because of its directional property and the ease with which it may be rotated through 360 degrees, the receiving loop antenna has found extensive application as a direction finding apparatus. It is used to locate radio transmitting stations and thus to furnish valuable information both in times of war and peace. The general directional characteristics of a loop antenna have already been outlined in previous paragraphs. It was shown that such an antenna will receive signals with an intensity depending on the position of the loop with respect to the direction of travel of the waves, other things being equal. From the fact that the effective emf. E and therefore the current in the tuned receiving loop oscillatory circuit are essentially proportional to the electromagnetic flux variations in the loop, it follows that this emf. is directly proportional to the cosine of the angle of the plane of the loop with the direction of travel of the waves. Thus,

$$E = a \cos \alpha$$

where a is a constant and α the angle defined. This may be represented in polar coördinates by the two tangent circles of Fig. 215, where the length of the vector OA is proportional to the received current for the various positions of the loop. The current is thus shown to be a maximum when the angle α is zero, and equal to zero when α is equal to 90 degrees.

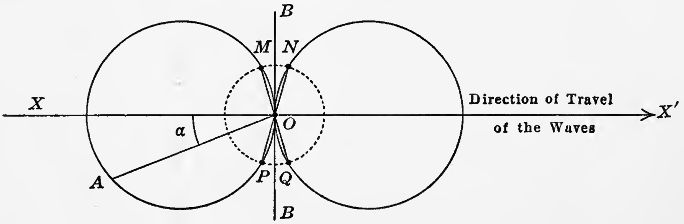


Fig. 215.

Having then a loop receiving circuit, such as is shown in Fig. 213, with provision for rotating the loop antenna about a vertical axis, it will be understood that by turning the loop (after having tuned the circuit) into a position such that the signals are heard with maximum intensity, the plane of the loop will coincide with the direction of travel of the waves, and will point directly toward the transmitting antenna. If the position of the loop is then recorded and a similar operation is repeated with signals from another station of known location, then it is possible by simple triangulation to locate the position of the unknown transmitting antenna. Or, as in the case of a ship or aircraft, if the loop is successively directed toward three land stations of known location, it is possible for the ship to locate its own position.

It will be noted, however, that for a given variation of the angle α, the cosine of this angle, and therefore the signal intensity, varies by a greater amount when the angle is around 90 degrees than when it is approximately zero. That is, the device will be more sensitive if adjusted for zero signal instead of maximum signal. To this greater sensitivity of the apparatus, is added the greater sensitivity of the human ear in determining the existence or non-existence of a signal, rather than the maximum intensity of the signal. For these reasons, the general practice is to turn the loop into a position for which the signals are not heard at all, this permitting a more accurate adjustment. The position of the loop is then normal instead of parallel to the direction of travel of the waves.

This last statement is only true, however, for the hypothetical case in which the detector and telephone receivers have perfect sensitivity. Actually, a certain very small current is required to operate the detector and telephone receivers, and if the received current is less than this, no sound will be heard in the telephones. If the vector OM, Fig. 215, represents this minimum current producing a sound in the telephones, it is evident that no signals will be heard when the loop occupies a position between OM and ON, or between OP and OQ. In other words, if the loop is rotated from its initial position OX, the signal will gradually decrease in intensity and will entirely fade out when the loop comes to the angular position OM. No signal will then be heard between the positions OM and ON. The sound will then reappear when the loop passes the position ON, will increase, pass through a maximum for the position OX', decrease to zero for the position

OQ, reappear for the position *OP*, and grow to the original maximum for the position *OX*. In order to determine the absolute zero position then, since it can be located only within the limits *MN* or *PQ* by ear, the positions *OM*, *ON*, *OQ* and *OP* of the loop for which the signal fades away are noted. The mean position between these four readings is then taken as that corresponding to the absolute zero position *BB* of the loop.

From the preceding discussion, it is seen that the smaller the minimum current required to produce a sound in the telephone receivers, the closer the four loop positions *OM*, *ON*, *OQ* and *OP* to the actual position *BB*. The minimum current is made smallest by increasing the sensitiveness of the device. This can best be done by the use of a high power, multi-stage amplifier. Vacuum tube amplifiers giving radio frequency amplification before rectification, and audio frequency amplification after detection have been used with great effectiveness and have resulted in an accuracy of the direction finder of very high order.

Another departure from the ideal curve represented in Fig. 215 is a distortion of the curve, as shown in Fig. 216, due to the pres-

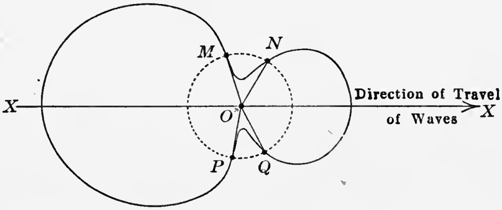


FIG. 216.

ence in the vicinity of the loop of asymmetrically disposed conducting bodies which absorb energy in certain parts of the space surrounding the loop. An abnormal distribution of the field about the loop is thus produced resulting in a correspondingly distorted curve. Under these conditions, the four positions of the loop for which the signals become inaudible are no longer symmetrical with respect to the direction of travel of the waves. This is not especially objectionable when using a very sensitive detector, amplifier and telephone receiver, but may seriously impair the accuracy of the direction measurements when the detector telephone circuit is of lesser sensitiveness.

It is possible to correct for the distortion of the loop directional characteristic by a suitable counterbalancing of the disturbing conducting masses in the vicinity of the loop. A simple means is illustrated by the circuit of Fig. 217, where the two terminals of the loop are each connected to one plate of an adjustable condenser, the other plate of which is grounded. By suitably setting these condensers *A* and *B*, the distortion may be eliminated or greatly increased, and the loop made unidirectional. The loop will then determine the absolute direction of travel of the waves.

The simplest form of a radio compass for land stations makes use of a square or triangular loop measuring between 2 ft. and 6 ft. square, pivoting about a vertical axis, and connecting to a suitable tuning condenser and detector, amplifier and telephone receiver circuit. By the simultaneous use of several stations of known location, simple triangulation processes permit the accurate and rapid location of hidden, unknown or distant radio transmitting stations. The operation of such compasses was so perfected during the war that the location of enemy radio stations was commonly made and battery fire directed to the spot. Similar stations, especially equipped but operating on the same fundamental principle, succeeded in locating enemy airships and submarines which betrayed their presence when sending out signals. It was a comparatively easy matter for two goniometric stations on the lookout to then follow the enemy airship in its travel, even if it were at a considerable distance from the stations.

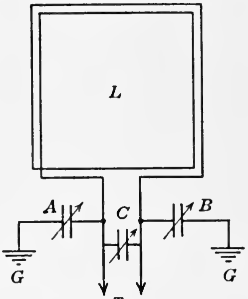


FIG. 2.17

A loop antenna design has also been adapted for use on submarines. This is more fully described in the chapter following.

Bellini-Tosi Goniometer.—Although the form of direction finder just described is probably of the simplest possible construction, it has certain deficiencies, as follows. In order to enable the rapid and easy operation of the set, it is necessary to make the loop winding comparatively small, say 10 ft. or 12 ft. square at largest. This small size results in the loop intercepting only an extremely small part of the total field set up by the transmitting antenna, and requires then the use of a powerful amplifier in order to produce sounds of sufficient strength in the telephone receiver to be audible. When such amplifiers are not available, it becomes necessary to use a larger loop antenna. But as this is cumbersome to rotate, the following scheme was originated by Bellini and Tosi to overcome this difficulty:

Two identical stationary loop antennæ, L_1 and L_2, Fig. 218, generally triangular in shape and of large linear dimensions, are set up at right angles to each other, each connected in series respectively with a variable condenser C_1, C_2 and a coil M_1, M_2. These two coils M_1 and M_2 are similar to each other, are fixed at right angles, and comprise several turns of wire. Each is in the same plane as its loop L_1, L_2, although this is not an essential condition. Each of the circuits $L_1C_1M_1$ and $L_2C_2M_2$ is then tuned to the wave length of the signals to be received, and the incoming waves produce in these circuits oscillating currents proportional in amplitude to the cosine of the angle of the plane of each loop L_1, L_2 with the common plane of the receiving and transmitting stations. These currents, flowing in the coils M_1 and M_2, produce about them electric fields which are proportional to the amount of energy received by their respective loops L_1 and L_2. These fields combine and add vectorially so that the resultant field of the coils M_1M_2 is proportional in amplitude to the received energy, and is parallel to the direction of travel of the waves, that is, parallel to a line joining the transmitting and receiving aerials. If, then, a small flat coil P is rotated around the common vertical axis of the coils M_1 and M_2, the emf. induced in this coil P by the resultant field of the two coils will be a maximum when the plane of the coil P is parallel to the line of the two stations and zero for a position normal to this line.

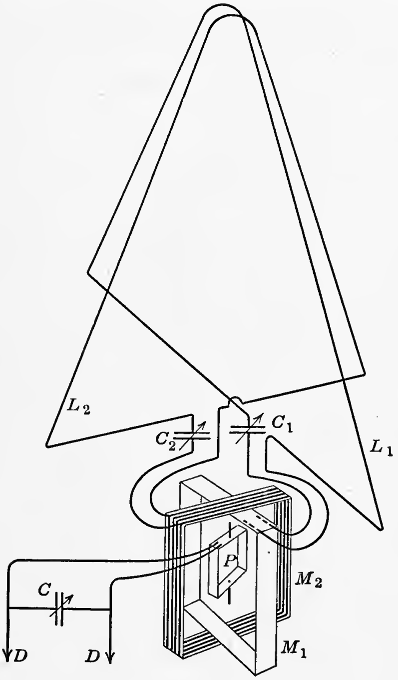


Fig. 218.

The coil P is connected to a tuning condenser C, and a detector, amplifier and telephone receiver circuit, and is made to rotate over a graduated dial to indicate its direction. As in the case of the single loop radio goniometer, the coil P is orientated so that the signals are received with minimum or zero intensity, to insure greater sensitivity. The coil P then does not point toward the transmitting station, but 90 degrees away from it. The accuracy of the system depends on the perfect symmetry of the loops L_1L_2, coils M_1M_2, and other parts of the circuit. It is in general necessary to take several readings with the coil P rotated 180 degrees each time.

The form of goniometer just described may be used for transmitting by connecting the coil P to a suitable exciter circuit instead of a receiving circuit. The direction of transmission may then very simply be changed by rotating the coil P about its vertical axis.

Airplane Radio Compass.—Both the single loop and double loop goniometers described above, when used as a direction finder or radio compass were shown to have maximum sensitivity when set for zero signal. That is, the loop or coil is turned until no signals are received, when the plane of the loop is normal to the direction of travel of the waves. When such a direction finder is used on an airplane, it is found that only the strongest signals can be heard, on account of the engine and wind noises, and the setting of the goniometer in the position of zero signal with any degree of accuracy becomes impossible. On the other hand, the setting for maximum signal has the advantage of being readily audible, but the maximum signal reading is quite inaccurate on account of the small difference in the intensity of the signals for relatively large angular rotations of the loop about its true position for maximum signals. This situation made it desirable to devise a method which would have the accuracy of the zero signal setting with the goniometer on maximum signal.

This has been successfully achieved by means of the circuit of Fig. 219, which illustrates the basic principle of the method used. The device comprises two vertical loops L_1 and L_2 permanently mounted at right angles to each other. They may be rotated simultaneously about some vertical axis. When the double pole double throw switch S is closed to the left, the two loop windings are connected in series, and their two extreme terminals are shunted by a variable tuning condenser C_1 and by some detector, amplifier and telephone receiver circuit. In the case of the figure, a vacuum tube detector D is shown, as this is generally used on account of its great sensitivity, and because a crystal detector would not operate satisfactorily under the vibration of the airplane engine. When the switch S is closed to the right, the loop L_2 is disconnected from the circuit and is not used, but a condenser C_2 is connected in its place. The loop L_1 is then shunted by the two parallel connected condensers C_1 and C_2.

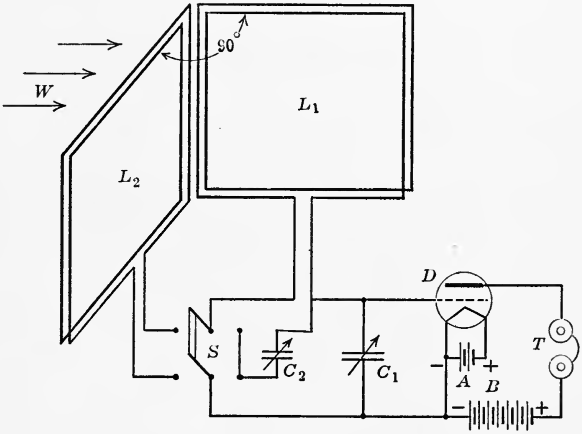


FIG. 219.

The tuning of the circuit to the wave length of the station to be received by the airplane and on which the latter will direct its course, is generally done on the ground before the flight, by exciting the circuit with a wavemeter emitting the desired wave length. The circuit is first tuned with the switch S closed to the right, by adjusting the condensers C_1 and C_2 until maximum response of the receiving circuit is obtained as evidenced by a maximum sound intensity in the telephone receivers. Then, closing the switch S to the left, the condenser C_2 is disconnected and the loop L_2 connected in its place. The result will be decreased capacitance and increased inductance of the receiving oscillatory circuit. If the condenser C_2 was given a suitable setting before throwing the switch, the product LC of the total inductance and capacitance of the circuit and therefore its natural wave length, may be made to remain unchanged when the inductance of the loop L_2 is substituted for the capacitance of the condenser C_2, and the circuit thus be made to be in tune for both positions of the switch S.

With the circuit thus adjusted and tuned, its use and operation on an airplane may be described as follows. The problem is to direct the airplane by having certain radio stations on the ground emit prearranged signals at certain wave lengths. The radio goniometer on the airplane is then used to indicate the direction of these stations of known position, and it is then possible for the navigator to locate his position on the map. Other uses of the airplane direction finder will be shown later. They all involve the orientation of the receiving loops, which is done in the following manner.

With the airplane flying, and a ground radio station emitting signals, the switch S, Fig. 219, is first closed to the right, thus disconnecting the loop L_2, and making use only of the antenna L_1 for the reception of the signals. The circuit having been tuned on the ground before the flight, it is simply necessary to turn the system around its vertical axis until the signals from the ground station are heard loudest. This will bring the loop L_1 to point toward the emitting station, that is, in a position parallel to the direction of travel W of the waves. At the same time, the loop L_2, which is mounted at 90 degrees to the loop L_1 will be normal to this direction W, and the waves will therefore induce no emf. in it. Now if the switch S is closed to the left, the oscillatory circuit will remain in tune and no change will be observed in the intensity of the received signals. If, however, as is generally the case, the loop L_1 is not exactly in the direction W, the loop L_2 will be somewhat away from its position corresponding to a zero induced emf., and since the loop L_2 is then around its most sensitive position, this small deviation from the actual zero position will manifest itself by a sharp increase or decrease in the intensity of the signals when the switch S is closed from right to left. By this method it is seen that great accuracy is obtained despite the fact that the setting is made for maximum instead of zero signal intensity.

Two methods have been used to construct and install the loop antennæ on an airplane. In one, the loops are mounted permanently on the wings and vertical struts of the airplane, and in order to turn the loop antennæ, it is necessary to turn the entire airplane. The other method makes use of two small loops wound on rectangular wooden frames which are held together at right angles to each other and mounted on a common vertical axis within the fuselage. They may then be rotated over a suitable graduated dial.

The wing loops are used mostly for direct flying toward or away from a given station. The airplane is pointed directly in line with the radio station, in which course the signals are heard with maximum intensity. The navigator then throws the switch *S* alternately to the right and left, and the airplane is kept in such a direction as to give no change in sound intensity when operating the switch. The fuselage loops are used when it is desired to find the direction of a ground station without altering the course of the airplane.

In using loop antennæ, and especially on board aircraft, sensitive and rugged detectors are used, and multi-stage high power amplifiers, such as are described in Chapter VIII.

ELIMINATION OF INTERFERENCE

Loop receiving circuits have of late been used with success in eliminating static interference. The general principles of this application will be explained briefly here.

In the reception of radio signals, it frequently happens that oscillating currents are produced in the receiving circuit which are due to natural electrical disturbances and result in noises more or less loud, in the telephone receivers. These noises sometimes seriously interfere with the reception of the signals. This is especially true of circuits using ordinary aerials of large linear dimensions, and on the other hand this type of interference is frequently absent in circuits using small loop antennæ. The causes of these natural electrical disturbances are not very definitely known. Electrical storms are only one of many probable sources of disturbing electromagnetic waves. The most serious sort of disturbance, frequently called "grinders" seems to be due to highly damped polarized waves, apparently traveling vertically, that is, normally to the earth's surface, as if generated under foot or overhead. Due to their high damping, these waves excite an oscillatory circuit at its own frequency, so that the use of tuned circuits does not help much in coping with this trouble. Also, they act simultaneously on circuits separated by several thousand feet. Another variety of "static" or natural disturbance often called "click," is produced by highly damped waves traveling approximately parallel to the earth's surface.

The elimination of these various disturbing oscillations with the retention of the signaling oscillations in the receiving circuits is the problem which appears to have been solved successfully by the methods described below.[1]

The fundamental principle of one class of circuits which deal successfully with the problem of interference consists in the use of two receiving antenna circuits so disposed and located in space with respect to the waves that the static or disturbing waves will set up in them currents of approximately equal amplitude and in phase, while the signal waves will produce in them currents differing in phase, or amplitude, or both. By suitably coupling these receiving antenna circuits to a single detector and telephone receiver circuit it becomes possible to eliminate the disturbing currents in the latter circuit, without destroying the signal current.

Another method, which is a development or adaptation of the one just explained, consists in the use of one or more antenna circuits so disposed in space that they will set up in a common circuit to which they are coupled or connected, currents due only to the disturbing waves, while the signal current is entirely neutralized or absent. Another antenna system in which both the disturbing and the signal currents are allowed to act is then coupled to this common circuit. By suitable adjustments of the coupling, decrements and other circuit constants and factors, it is possible to eliminate the interfering currents, leaving only the signaling current.

Thus, consider first the circuit of Fig. 220. This consists of two identical loop antennæ L_1 and L_2, each of which is connected to a suitable tuning condenser C_1, C_2. Each loop circuit comprises also an adjustable resistance R_1, R_2, and is coupled to a common tuned intermediate circuit through the inductive coupling M_1, M_2. This intermediate tuned circuit is then coupled to a suitable detector and telephone receiver circuit, which in practice comprises also a high power amplifier such as a multi-stage radio and audio frequency cascade vacuum tube amplifier. The two loops, L_1 and L_2 are set up at a distance from each other equal, for best results, to one-half the wave length of the signals to be received, and are so situated that their common plane will pass through the transmitting station sending the signals.

Now if the waves emitted by the transmitting station travel

[1] Roy A. Weagant, *Proc. Inst. Radio Engrs.*, June, 1919.

along the direction W, Fig. 220, the loop antenna L_1 will be reached by the wave one-half period before the loop L_2, on account of the separation of the two loops, and the currents thus set up in them by the signaling wave will be one-half cycle, or 180 degrees, out of phase. On the other hand, a disturbance such as a grinder, traveling vertically as shown by the arrows S, will affect both loops simultaneously, and will therefore set up in them currents in phase, of approximately the same amplitude, and of a frequency equal to the natural frequency of the

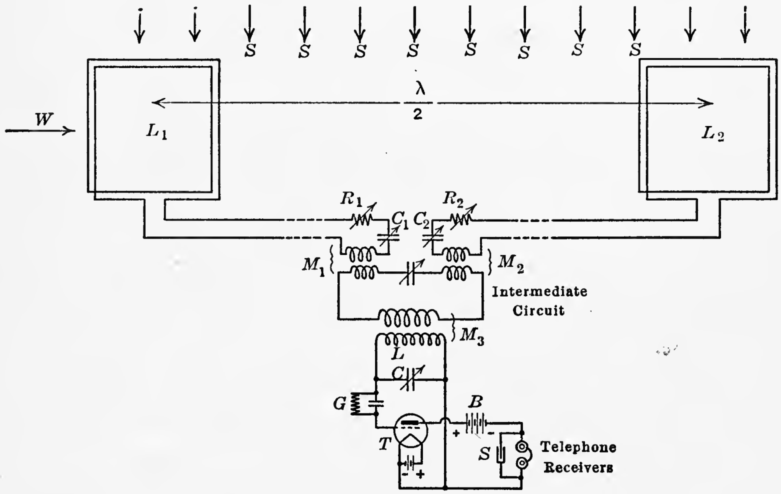


FIG. 220.

tuned loop circuits. By a suitable adjustment of the resistances R_1 and R_2, and by using the proper values and polarities of the couplings M_1 and M_2, it is possible to make the signal currents of the two loop circuits add their effects in the intermediate tuned circuits and at the same time, have the disturbing currents neutralize each other in the same circuit. Conversely, by simply reversing the polarity of either one of the couplings M_1, M_2, the signal current will be reduced to zero and only the disturbing current will be retained. The circuit of Fig. 220 thus enables the elimination of either the disturbance or the signal.

A defect of this circuit is that, for the reception of long waves such as are used for instance in Trans-Atlantic work, it is necessary to separate the two loop antennæ L_1 and L_2 by a long distance, of the order of one or two miles, or even more. This in turn requires the use of long connecting lines between the loop antennæ and the coupling and detector circuits, which introduce distributed capacitance and inductance in the circuit and may thus complicate the operation of tuning. Also these long lines may become receiving aerials for other wave disturbances and increase the difficulties of the problem. This is in part avoided by modifying the circuit as shown in Fig. 221, where the lines

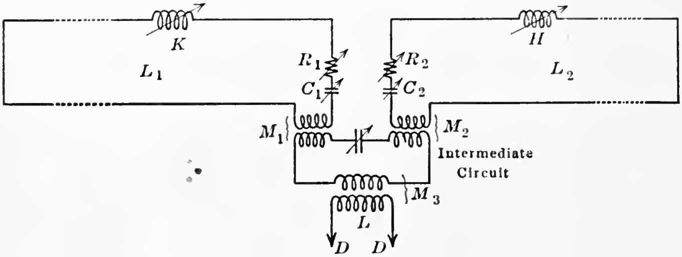


FIG. 221.

have been spread apart, and themselves form the receiving loop aerials. In this arrangement, however, the current and voltage distribution along the loops make it necessary to insert the tuning inductance coils *H* and *K* in the upper wire of the loops, midway between their ends. In view of the dimensions of the loop these coils may be several hundred feet away from the central receiving station.

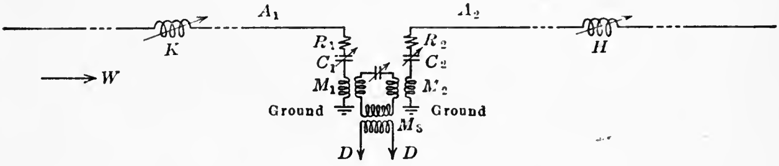


FIG. 222.

It will be noted in the two circuits of Figs. 220 and 221 that the elimination of interference is due to the action of the disturbing wave in setting up in the circuit approximately equal and in-phase currents while the signal wave produces currents about 180 degrees out of phase in the two antennæ. The circuit of Fig. 222, where two inverted "L" antennæ are substituted for the loops, has the added advantage that the antennæ are not simply directional in their planes, but are unidirectional, as was explained in Chapter III. Thus for a signal wave moving in the direction W, the antenna A_1 is in the position corresponding to minimum signal strength while the antenna A_2 is in that of maximum signals. On the other hand both antennæ A_1 and A_2 are of equal sensitiveness to "grinders" or disturbing electromagnetic waves. The neutralization of static with the retention of the signal is thus greatly facilitated and improved.

As was explained before in connection with the circuit of Fig. 220, it is possible by using suitable values of the coupling coefficients M_1 and M_2, to obtain in the intermediate circuit either the signaling current alone, or the disturbing or static current alone. In the latter case, the double antenna system

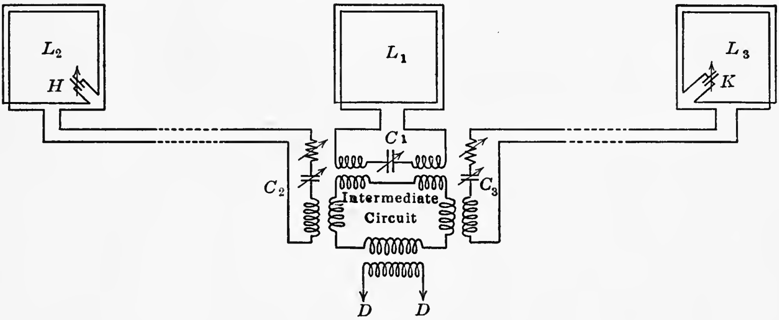


FIG. 223.

may be considered as a source of static disturbances, supplying these to the intermediate circuit. Considering then the circuits of Fig. 223, suppose the loops L_2 and L_3 to be adjusted to supply to the intermediate circuit static disturbing currents at the exclusion of signal currents. The loop L_1, on the other hand, being an ordinary loop receiving antenna with no other loop to counterbalance its received current, will supply to the intermediate circuit both the signal current and the disturbing current. By suitable adjustments, it is then possible to make the disturbing currents supplied from the system L_2L_3 exactly counterbalance those due to the system L_1, and leave the signal current alone undisturbed. The advantage of this system over that of Fig. 221 is that the click form of static is taken care of equally as well as the grinder form, which was not the case in the circuit of Fig. 221.

The various circuits above, while all effective in the elimination of natural disturbances, have the common disadvantage of being of large linear dimensions, covering many hundreds of feet, and therefore, frequently unwieldy, cumbersome, and in many cases impracticable. To meet this objection, the circuit of Fig. 224 has been devised and permits of successful interference elimination with apparatus of quite small dimensions. The loop L is an ordinary receiving loop antenna, which is tuned by means of the condenser C and coupled to a vacuum tube detector and telephone receiver circuit through an untuned intermediate circuit. In actual practice, the circuit also comprises

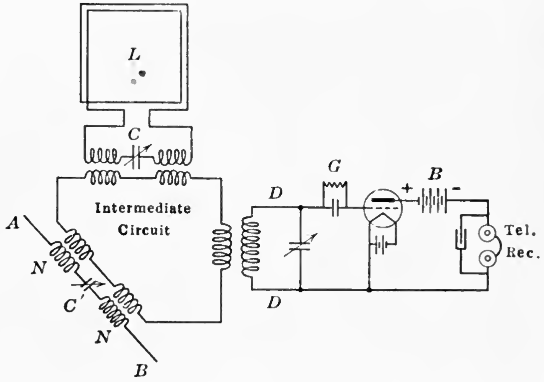


FIG. 224.

a multi-stage amplifier. Coupled to the same intermediate circuit is the conductor AB. This is a conductor about 9 ft. long held along a straight line by means of insulators fastened to a straight support. The conductor comprises at its center suitable coupling and tuning coils N and a condenser C'. It thus forms a receiving antenna, and provision is made to enable its being directed toward any point of space. By referring to Chapter I it will be remembered that currents may or may not be induced in such a conductor by a current flowing in another circuit (such as the transmitting antenna circuit), depending upon the relative positions of the two circuits. This is found to hold true here, and a position is found for the conductor AB for which no signal current is set up. For the same position, however, static disturbing currents will be induced in AB, due to the different direction of travel of the disturbing waves from that of the signal waves, and thus the conductor AB will supply to the intermediate circuit disturbing currents which, by suitable coupling adjustments, may be made to exactly counterbalance the static currents collected by the loop L, leaving the signal current of the latter undisturbed.

A method based on a somewhat different principle, and for the main purpose of eliminating interference from other stations, has been developed by Alexanderson under the name of "barrage receiver." This is diagrammatically illustrated in Fig. 225. Two long and low inverted L antennæ A_1 and A_2, installed in line with each other are used, each having a length which is preferably equal to one-quarter the wave length to be received.

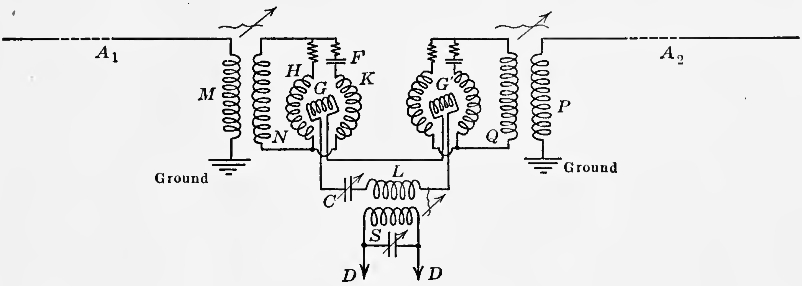


FIG. 225.

Each of these antennæ is individually coupled to a phase shifting device through the coils MN and PQ, respectively. Such a device consists essentially of two stationary coils HK connected in parallel, one of which (K) is connected in series with a condenser F. The function of the condenser is to establish a phase difference between the currents flowing in the two coils H and K when an electromagnetic wave energizes the antenna. A coil G may be rotated between or within the two stationary coils and may thus be coupled to either one to a greater or less extent. By connecting the two rotor coils G and G' in series with a coil L and tuning condenser C, and suitably turning each of these two rotor coils, signal currents may be drawn from each antenna A_1 and A_2 at such phase angles that interfering currents, which generally occur at different phases, may be eliminated or neutralized. The circuit $GG'LC$ is tuned to the signals to be received, and coupled to a tuned secondary circuit S which is in turn connected to a suitable amplifier, detector and telephone receiver circuit.

CHAPTER XII

MISCELLANEOUS APPLICATIONS OF RADIO CIRCUITS

Airplane Radio Apparatus.—Airplane radio apparatus does not differ fundamentally from ordinary radio apparatus. It must be especially designed, however, to meet a number of special conditions existing on an airplane, which are to a great extent absent in ground radio.

A primary consideration is that all apparatus shall be built of carefully selected material to insure minimum weight and bulk, as the carrying capacity and available space of the airplane is limited and heavy auxiliary apparatus is not thought of kindly by the airplane design engineer. The design and construction of all parts must also have great ruggedness in order that the apparatus may operate properly and without changing its adjustments under the continual vibrations of the airplane due to the engine, wind, etc. For this reason, all adjusting handles are usually locked in position after the set has been tuned and adjusted. Vacuum tube sockets are mounted on soft sponge rubber cushions to absorb the shocks, and the set box itself is frequently mounted on shock absorbers.

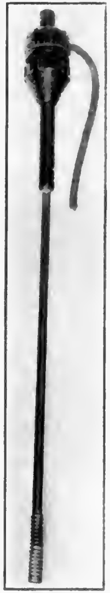
In addition to these mechanical considerations of the installation, great care must be taken in preventing all possibility of producing an open spark, as the danger of fire on an airplane is extremely great. This will be well appreciated when it is noted that the "dope" or varnish painted on the linen or cotton fabric of the wings and fuselage is a highly inflammable composition. The fire hazard is also increased by the likelihood of gasoline vapors being present in the cockpits. For these various reasons, all wiring on an airplane is made with high insulation and is securely attached along the airplane struts so that it will not shake loose. Special telegraph keys are used, having enclosed contacts, and all spark gaps are either totally enclosed or covered with wire gauze.

Damped and undamped wave radio telegraph sets and also radio telephone sets were developed and successfully operated on airplanes by the various allied governments during the war. In all the later airplane sets, except of course the damped wave transmitting sets, vacuum tubes were used for both transmitting and receiving. The circuits are in general the same as for ordinary sets designed for ground use. An interesting question, however, has been as to the best method of supplying power for the vacuum tube filaments and for the plate circuits, etc. Storage batteries have been used to quite some extent for certain circuits where high voltages were not required. But the limited time during which such batteries can be depended upon to supply power of uniform voltage without recharging, makes it essential on account of the long duration of many flights to derive the required energy from a small generator. Most of the generators thus used are enclosed in a streamline casing and mounted either on the wing or one of the landing gear struts and usually in the air stream of the propeller on tractor type planes. They are driven by a small propeller or airfan which rotates at very high speed on account of the high velocity of the wind. In order to obtain constant output from such generators, special regulating devices are used. These include variable pitch airfans, which drive the generator at approximately constant speed despite variations in the airplane or wind velocity; special regulating vacuum tubes, or iron wire resistances, etc. These various devices, together with several types of generators used in the U. S. Army during the war have been described in the technical press and are not treated further here.[1]

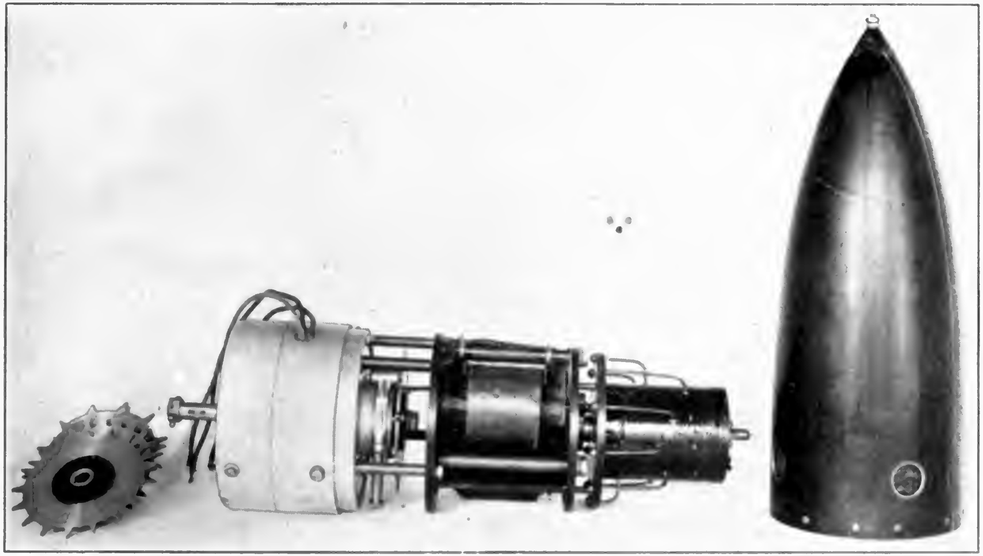
A problem bearing more directly on radio communication is that of the radiating or antenna circuit used on airplanes. Both the loop and the "open" type of antenna are used successfully. As regards the loop type, reference should be made to a section of Chapter XI in which the loop radio receiving circuits used on airplanes were fully described. These are mostly used for direction finding purposes, but may of course be used also for receiving messages in the usual way. Other types of loop antennæ permit both transmitting and receiving from an airplane, and have been used, although only to a limited extent, on airplanes flying in squadron formation.

The type of open antenna which has probably found the widest application is that illustrated in Fig. 226, called a "trailing antenna." It consists of a 300-ft. length of bare wire which

[1] See *Electric Journal, Electrical World, Proceedings of the A. I. E. E.*, etc.



(*A*)



(*B*)

Plate 11.—(*A*) One type of antenna fairlead with a piece of lead-in wire attached for connection to the radio set box. (*B*) Partially dismantled damped wave airplane radio telegraph transmitting set with synchronous rotary spark gap. From left to right, the self-excited alternator, spark gap, power transformer and oscillation transformer are integrally mounted and enclosed by the micarta streamline casing seen at the right. Extra rotary electrodes for the spark gap for changing the spark frequency are seen at the extreme left. Signal Corps set type SCR-73.

(*Facing page* 284.)

Plate 12.—An induction coil type of spark set operated from the storage battery at the left. Designed and used extensively by the British in airplane fire control work. The antenna tuning inductance is the flat spiral coil mounted on top of the set box. The set and battery are fastened on a sliding shelf for quick removal from the fuselage. Note the antenna reel at the right, outside the cockpit. Present practice is to put the reel inside the fuselage. British Sterling set. Signal Corps set type SCR–65.

is attached to a reel at one end while the other or trailing end is weighted by means of a small lead weight shaped often like a fish and called by that name. The antenna wire is led out of the airplane through a tube called the "fairlead" which is mounted in the floor or wall of the fuselage. The antenna wire

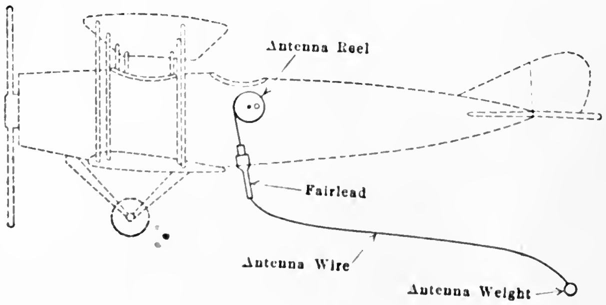


Fig. 226.

bears against the metal tube extending through the center of the fairlead, which is connected to the radio set box, and makes electrical contact. The antenna wire is thus free to slide through the fairlead and may be reeled in to avoid breakage before the airplane comes down for landing.

The trailing wire forms one side of the radiating circuit and is equivalent to the aerial of the ordinary ground system. The counterpoise is made up of the metallic parts of the airplane—the engine, stay wires, etc., which are all carefully bonded together. The radiation characteristic of a trailing antenna is shown in the diagram of Fig. 227, which illustrates the directional properties of this antenna. It is seen that the airplane, when sending or receiving, must fly toward the point with which it is desired to communicate for greatest range and loudest signals.

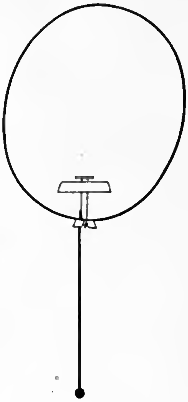


Fig. 227.

It may be noted from Fig. 227 that the directional characteristic is not quite symmetrical with respect to the longitudinal axis of the airplane. This is because the antenna reel is generally mounted on the right or left of the fuselage, instead of the center. In the case of the figure, the antenna wire is led out of the airplane from the left hand side of the fuselage, which results in an angular displacement of the electrostatic axis of the system, giving maximum directional effect ahead of the airplane and a little to the right.

The construction of the fairlead is shown in Fig. 228, which represents a cross-sectional view. The fairlead is made up of a tube of insulating material, such as bakelite or micarta, which is clamped to the fuselage wall or floor. Inside this tube is a metal sleeve or pipe, through which the bare antenna wire passes. The metal sleeve thus makes electrical contact with the wire and a binding post attached to the sleeve furnishes a ready means of connecting the antenna to the radio set box.

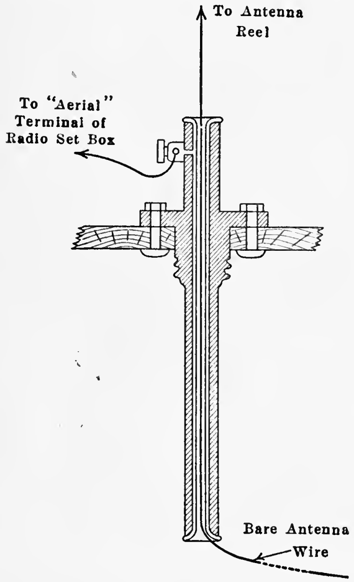


Fig. 228.

The length of antenna wire used depends on the wave length of the signals. As a rough rule, and using a Curtiss JN-4 airplane, the natural wave length in meters is approximately equal to the length in feet of the trailing antenna. It should be noted that the natural wave length of such a trailing antenna depends not only on the length of the trailing wire but also on its position relative to the counterpoise and the airplane. It follows that the antenna constants, and therefore the tuning of the airplane radio set, change whenever the airplane alters its direction of flight, or whenever the wind alters the position of the trailing wire. When using damped wave sets, this results in temporary fading out of the signals due to detuning. When using undamped wave sets with heterodyne reception, it results in a variation of the pitch of the signal note.

Another type of antenna is shown in Fig. 229, and is known as the double trailing antenna. The diagram is self-explanatory. This antenna is not quite as directional in front of the airplane as the single trailing antenna, as shown in Fig. 230, but the waves spread out more on both sides of the machine so that it is better suited for squadron flying or intercommunication between airplanes. Also, each of the trailing wires is only 150 ft. long, instead of the usual 300 ft. as in the case of the single wire antenna. This permits a greater liberty of flying, and enables the airplane to make sharper turns, or even to loop the loop without as great danger of getting the antenna entangled with the tail of the airplane. This is obviously a danger common to all trailing wire antennæ. To obviate this, the antenna of Fig. 231 has been devised.

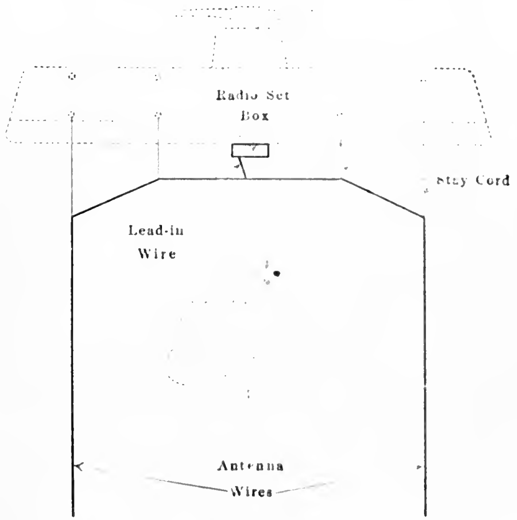


FIG. 229.

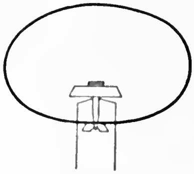


FIG. 230.

It consists of a triangular antenna stretched out in a horizontal plane between insulators above the airplane, while the stay wires and engine form the counterpoise. The objections to

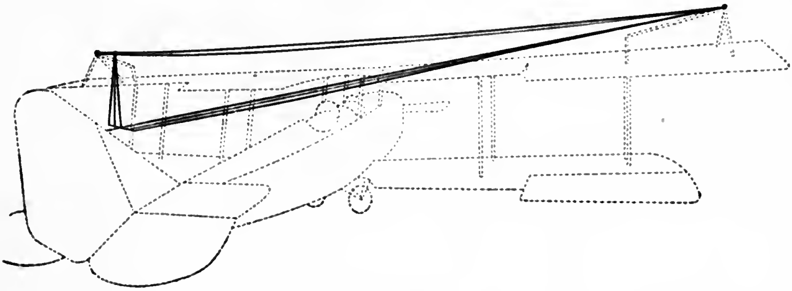


FIG. 231.

this type of antenna are its low radiation and high head resistance when flying.

Submarine Radio Apparatus.—The submarine is itself only a few years old and the use of radio telegraphy in conjunction with it is of quite recent date. There is consequently only comparatively little data to be had for publication at the present time. As pointed out in the preceding chapter, a loop antenna has been used successfully in this country for submarine radio work and it will be briefly described here.[1]

Referring to Fig. 232, the submarine loop antenna *ABEF* is shown to be made up of a wire *AB*, insulated at end *A* by means of the insulator *I*, and grounded at the lower end *B* to the hull of the ship, which forms part of the loop circuit. Point *B* is thus connected to point *E* through the metal body of the sub-

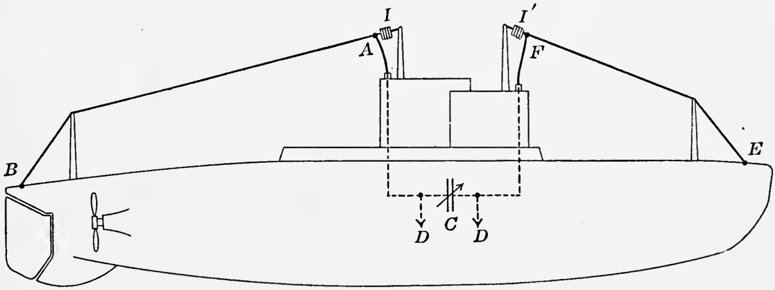


Fig. 232.

marine. *EF* is another wire similar to *AB*. Both these wires are of copper, covered with a heavy insulation. Lead-in wires similarly insulated, are connected at *A* and *F*, and brought inside the ship to a variable tuning condenser *C* and a suitable transmitting or receiving circuit connected at *D*. The loop circuit is thus electrically the same as any ordinary loop circuit. Its simple construction does not require any additional mast and therefore does not interfere with the submerging of the boat.

The operation of the loop when the submarine is running on the surface of the water is of course no different from that of an ordinary loop used on any ship or ground station. Reception and transmission of signals can furthermore be effected without special difficulties when the boat and loop are totally submerged. During tests made by the Bureau of Standards off the Atlantic Coast, the following important data among other were obtained.

When the submarine is submerged, any North American or European station can be received as distinctly as when it is on

[1] The data on this loop antenna was furnished by the Bureau of Standards, Washington, D. C.

the surface. The maximum depth at which signals are readable, however, is dependent upon the wave length of the transmitting station. This depth increases with the wave length, being about 21 ft. (for the top of the loop) when receiving signals of 10,000 meters wave length. The receiving apparatus used in these tests comprised a vacuum tube detector and a three-stage resistance-coupled vacuum tube cascade amplifier.

Transmission of signals from below the surface of the water can also be effected successfully, the range decreasing as the submarine submerges deeper below the surface. With a 1-kw. spark set and a 952-meter wave length, the range observed in the Bureau of Standards tests was 10 to 12 miles with the submarine and loop just below the surface, while it reduced to 2 or 3 miles for a depth of 8 to 9 ft. measured from the top of the loop to the surface. When the submarine is running awash or on the surface, the range becomes at least 100 miles even in very stormy weather. Grounds or short circuits caused by heavy seas which render the ordinary antenna inoperative in stormy weather, have no effect on the closed loop. A further use for the submarine loop radio is as a direction finder, whether submerged or on the surface, in the manner explained for ordinary loops in the preceding chapter.

The above results are of especial interest since they are among the few observations made of the propagation of radio waves below the surface of the sea. The success of the efforts to communicate from below the surface is due in large measure to the availability of the extremely sensitive vacuum tube amplifiers used for receiving the signals. The propagation of the electromagnetic waves in sea water is not essentially different in principle from the propagation in air, despite the considerably greater electrical conductivity of salt water. The underlying principle was given in Chapter I, where current conduction in a metal or other conducting body was studied. In the light of these principles, an explanation of the process of wave propagation in sea water may be given as follows:

Suppose the submarine loop to be entirely submerged and used for transmitting signals. The high frequency alternating current which is sent through the loop circuit during transmission, sets up about the loop a similarly alternating electric field. The effect of this field is twofold. First, it sets into motion the free electrons contained in the water, thus creating conduction or eddy currents which permanently absorb a part of the energy of the field and represent a loss of energy similar to that encountered in the case of ionized air. The second effect is to produce a molecular distortion of the water, which distortion was shown to be the agent of the field wave propagation. The theory thus outlined may be worked out more fully by a direct application of the principles given in Chapter I.

Radio Frequency Current Generation.—Three-electrode vacuum tubes are coming into use as a source of alternating currents of any frequency desired for special laboratory purposes. This was previously pointed out, and only a few special applications will be cited here. Any of the vacuum tube oscillator circuits described in Chapter IX may of course be used for generating undamped oscillations for laboratory purposes. However, it is better to adapt the circuit used to the particular object in view.

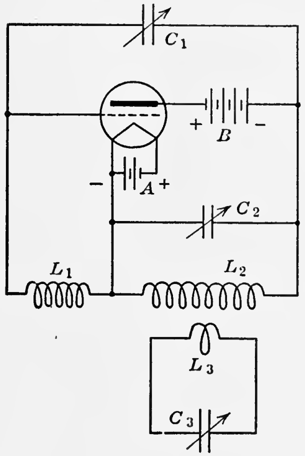


Fig. 233.

The circuit of Fig. 233 has been used successfully to obtain heavy alternating currents of any desired frequency. The circuit consists of a three-electrode vacuum tube, the plate and grid circuits of which are coupled for oscillation generation through the coils L_1 and L_2 and condenser C_1. The frequency of the oscillations is set to the desired value by means of the condenser C_2. The coil L_2, in which the oscillating current flows, is then coupled to a coil L_3 which is itself closed upon a condenser C_3 so adjusted as to tune the circuit L_3C_3 to the frequency of the oscillations generated by the tube. In order to obtain a current of considerable amplitude in the circuit L_3C_3, the circuit must be of low resistance and is therefore made of heavy copper ribbon. Also, and this is the essential feature of the circuit, the transformer L_2L_3 is a stepdown transformer, built upon the same fundamental principles as welding transformers. The secondary L_3 of this transformer is of a single turn, and consists of a heavy copper tubular sleeve over which is wound the coil L_2, insulated from it and comprising many turns. A consequence of this construction is that, due to the small inductance of coil L_3, the condenser C_3 must be of large capacitance in order to bring the circuit $L_3\,C_3$ in tune. This in turn favors the setting up of large currents in the circuit. Currents as high as 75 to 100 amperes r.m.s. have thus been obtained at the Bureau of Standards with a similar circuit.

The circuit of Fig. 234[1] is one of many used for the production of high alternating voltages of high frequency. It consists of a three-electrode vacuum tube coupled inductively through the coils L_1 and L_2 to the oscillatory circuit CL_1L_2. The grid-to-plate coupling of the tube may be varied by means of the contacts

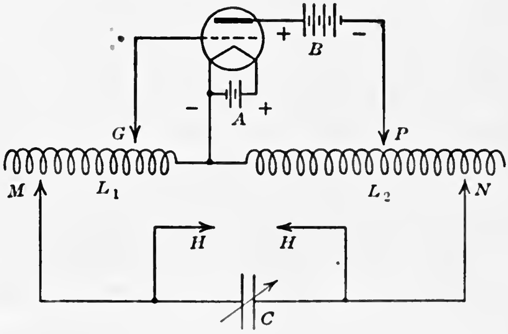


Fig. 234.

G and *P*, while the oscillatory circuit may be given the desired natural frequency by means of the contacts *M* and *N*. The condenser *C* is not used for tuning purposes, as will be explained presently. As a result of the oscillations, the potential difference across the condenser plates alternates at the circuit frequency. Its maximum value, however, is dependent on the capacitance of the condenser, being greater the smaller this capacitance (see Chapter II). It follows that the potential difference at points *HH*, may be simply regulated by adjusting the condenser *C*.

The great advantages of vacuum tube methods of obtaining undamped high frequency alternating currents are the ease of varying the frequency, the good wave shape of the oscillations, which for most purposes can be considered sinusoidal, and the absence of movable parts such as the rotor of the high frequency alternator. To these should also be added the great constancy of the frequency of the oscillations. Under suitable conditions,

[1] *General Electric Review*, Vol. 20, 1917, p. 636.

it has been possible to check by actual tests that the frequency was maintained exactly the same for a period of minutes with a variation of only one cycle in 500,000.

It may be of interest to describe these tests in a few words, as conducted by the Bureau of Standards. Use was made of the cathode ray oscillograph or Braun tube.[1] Two similar three-

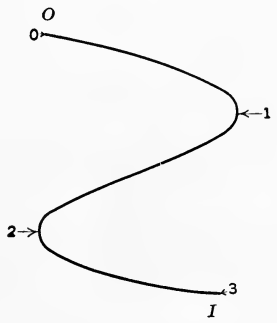


FIG. 235.

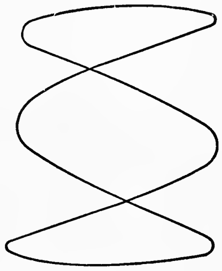


FIG. 236.

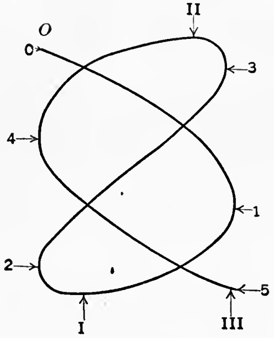


FIG. 237.

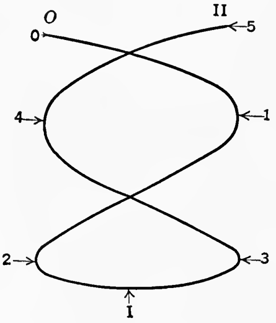


FIG. 238.

electrode vacuum tube oscillator circuits were set up and connected so as to act simultaneously upon the cathode stream of the oscillograph tube, the potential across one circuit deflecting the stream horizontally while the current through the other circuit deflected it vertically. The circuits were then tuned to definite frequency ratios, and the resulting Lissajou's figures photographed, as traced by the cathode stream. There was no

[1] This device is not described here. It is a standard piece of laboratory apparatus. Detailed descriptions are given by J. P. Minton in the *General Electric Review.*

connection between the two circuits except the slight electrostatic coupling due to the deflecting elements of the cathode ray tubes. The effect of this coupling was, however, successfully compensated for by suitable electrostatic shielding of the circuits. The figures given here are tracings from actual photographs, which were too dim to be reproduced satisfactorily. Fig. 235 shows a frequency ratio of 3 to 1 between the two circuits. Fig. 236 shows the same ratio with a constant difference of phase of 90 degrees. Figs. 237 and 238 show frequency ratios of 3 to 5 and 2 to 5, respectively. The perfect constancy of the ratio during operation is denoted by the smoothness of the curves. Any slight variation of their ratio would immediately have altered the shape of the curves very appreciably.

The use of this Braun tube method of picturing high frequency currents makes possible an accurate comparison of radio frequency oscillations in two circuits. It also permits of an exact measurement of the distributed capacitance of a coil with a degree of accuracy far beyond that of other methods previously used.

Use of the Vacuum Tube for Sustaining Mechanical Oscillations.—The three-electrode vacuum tube may be used to sustain mechanical oscillations in any system possessing inertia and elasticity, in a manner similar to that of sustaining electrical oscillations. It is simply necessary that the mechanical system be started oscillating and made to vary the grid potential of the tube in such a way that the resulting plate current variations will be of suitable magnitude and phase to sustain the mechanical oscillations. This subject has been given a great deal of space in the technical press of Europe recently, but seems to have little practical value in its present status.

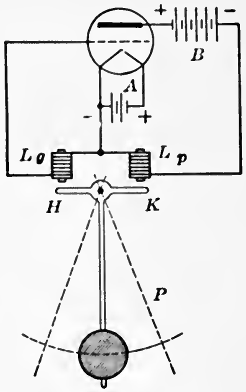


Fig. 239.

As an illustration, consider the system of Fig. 239, which shows how a three-electrode tube may be used to maintain undamped oscillations of a pendulum *P*. Two coils L_g and L_p are respectively inserted in the grid and plate circuits of the tube and placed in front of a small iron armature *HK* which is integral with the pendulum. As the pendulum is swung out of position it oscillates back and forth, moving the end *H* of the armature alternately toward and away from the grid coil L_g. This induces an alternating potential between the grid and filament of the tube, which in turn varies the current in the plate circuit and plate coil L_p. There results a correspondingly varying attraction of the coil L_p on the end *K* of the armature which, for suitable magnitude of the currents and proper polarity of the connections, has such a phase relation with respect to the oscillation cycle of the pendulum as to sustain its motion continuously. The energy expenditure from the plate battery is thus seen to compensate for the friction losses which in the absence of the vacuum tube device would damp out the oscillations of the pendulum or bring it to rest.

Another example is given in Fig. 240, where undamped vibrations of a tuning fork *T* are obtained by means of grid and plate coils L_g and L_p, disposed on either side of the tuning fork. The explanation is quite similar to that just given for the pendulum. It will be noted that in both cases the plate and grid circuits are coupled magnetically to a common mechanical system possessed of a natural period of vibration or oscillation of its own. This is identical with the case of electrical oscillations, where the tube circuits were inductively coupled to a common oscillatory circuit having a natural period of electrical oscillation.

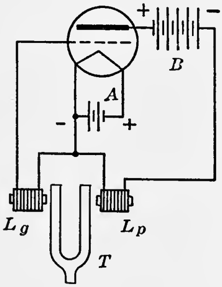


Fig. 240.

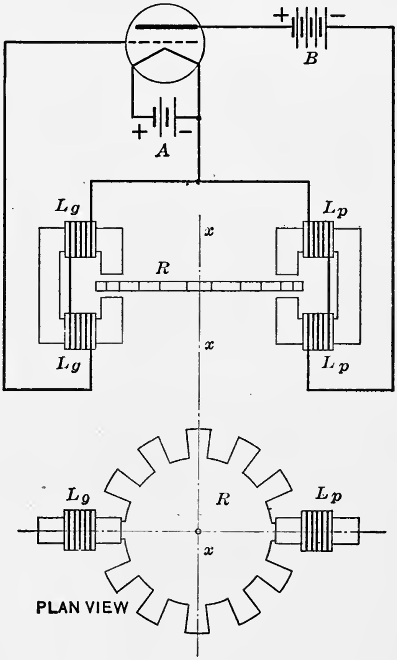


Fig. 241.

A somewhat different case is that of Fig. 241, which, however, is merely a different application of the same fundamental principle. The grid and plate coils L_g and L_p are wound over iron cores having a gap in which is placed an iron disc R which is free to rotate around the axis XX. This disc is provided with a number of teeth. When set in motion, the teeth and slots of the disc alternately pass the iron core yoke of the grid coils, inducing an alternating grid emf. which in turn synchronously varies the current in the plate coils L_p. The attraction of the latter on the rotor teeth thus varies synchronously with the motion of the rotor between a maximum and minimum. With a suitable angular position of the coils L_g and L_p around the disc and proper polarity of the connections, these variations will occur at such times that the rotor is kept in continuous motion.

INDEX