RADIO ENGINEERING PRINCIPLES

RADIO ENGINEERING PRINCIPLES

BY

HENRI LAUER, B. S.

AND

HARRY L. BROWN, B. E. E.

SECOND EDITION

McGRAW-HILL BOOK COMPANY, INC.
NEW YORK: 370 SEVENTH AVENUE
LONDON: 6 & 8 BOUVERIE ST., E. C. 4
1928

PRINTED IN THE UNITED STATES OF AMERICA

THE MAPLE PRESS COMPANY, YORK, PA.

PREFACE TO SECOND EDITION

This second edition of "Radio Engineering Principles" is a revision and amplification of the first edition, which was written after the conclusion in 1919 of the authors' radio work as Second Lieutenant and Captain, respectively, in the Signal Corps, U. S. A.

The new edition follows the same general plan and purpose as the first, but the development of the art since the World War, particularly with respect to the science and practice surrounding the three-electrode vacuum tube, has made considerable additions desirable.

The book is intended to serve as a general textbook on radio. It is devoted in large part to a 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 mainly to center. But the principles involved in the older forms of radio apparatus are also treated with sufficient fulness to round out the student's information and to cover all the essential principles of wireless communication.

In the detail development of the principles involved, 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 applications 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 apparatus.

THE AUTHORS.

December, 1927.

POSTSCRIPT

As this Edition is about to go to press word has come to me of the sudden death of my co-author and friend Harry L. Brown. His loss is keenly felt by his associates in many varied fields of the electrical industry to which his activities were devoted. The preparation of the present edition of our book had been the object of careful attention on his part, and no later than a few days ago did he put the final touches to the new Preface of the book.

In remembrance of his kind cooperation and as a tribute to his memory, I most respectfully dedicate this second edition to his widow, Mrs. Harry L. Brown, with the expression of my sincere grief and deep sympathy.

HENRI LAUER.

November 11, 1927.

CONTENTS

Chapter X. Radio Telephony

Chapter XI. Miscellaneous Applications

RADIO ENGINEERING PRINCIPLES

CHAPTER I

UNDERLYING ELECTRICAL THEORY

Electricity and Matter.—Electricity may be developed in material bodies in a variety of ways, such as by contact, friction, or heating. Thus, two pieces of matter rubbed or simply placed in close contact with each other and then separated, acquire each the property of attracting or repelling other bodies. This permits a distinction between two kinds of electricity, called positive and negative, like charges repelling each other while opposite charges attract each other. When coming into contact, opposite charges combine and the bodies supporting them revert to a neutral condition.

Electricity is thus normally present in material bodies and may be brought forth by placing these under suitable conditions. When materials are not placed under such special conditions, they do not display electrical properties and are said to be in a neutral state.

Now the study of chemical and other properties of matter having shown every substance as made up of a large number of atoms or molecules, all similar for a given substance, there are two possibilities as to where electricity may exist in store within the substance: it may be between the atoms, or within the atoms themselves. In fact, both theories must be merged into one in order to account properly for such phenomena as electrical conduction, emission of negative electricity by heated bodies, etc., which have gradually permitted a better and closer understanding of the atomic structure of matter.

An atom is thus considered as having a mass or charge of positive electricity at its center, 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 polarities exactly neutralize each other, and no exterior electrical force is observed. All these negative charges are equal, and are called *electrons.* They constitute the smallest possible fractions of electricity obtainable, and have a diameter which is estimated as 1×10^{-13} cm. (The hydrogen atom is about 60,000 times larger.) The entire structure resembles a miniature solar system in which the planets, represented by the electrons, are all equal in size. The difference between atoms of various substances is in the number and arrangement of the electrons revolving around the central positive charge, which accounts for the atomic weights and other characteristic properties. Thus, while an atom of hydrogen probably has only two electrons, an atom of mercury has several thousands of them.

Another similarity with a solar system is the existence of electrons in the atoms of certain sustances, such as the metals, which follow very eccentric orbits, like the comets of our solar system. Such electrons, being thus projected at comparatively great distances from the atom center, may even go so far as to escape the action of one atom and enter the system of another one, and are therefore called "free electrons." They may easily be torn from any one atom system under the attraction (or repulsion) of some electric charge external to the body and will then travel between the atoms in a common general direction, creating within the body a drift or current of electricity. This *electric current* will be studied later in greater detail.

It is not within the scope of this book to study the forces acting on the revolving electrons within the atoms, such as the attraction of the positive center charge on the electrons, the mutual repulsions between different electrons of the atoms, the centrifugal forces attending to the rotation of the electrons, etc. It should be stated, however, that the atomic structure just described is a stable one, owing to the action of these internal forces an attempt to shift the electron ring with respect to the central positive charge encountering their opposing action. This tends to restore the original shape of the atomic structure when the external distorting forces are removed.

In view of this electron structure of matter, substances may be divided into two classes, according to whether they contain free electrons or not. In the latter case they are called insulators or dielectrics, and in the former case, conductors. Actually, all materials contain at least a few free electrons, but some contain them in such small numbers per unit volume that they may be considered substantially as insulators.

Coulomb's Law.—The action of electric charges upon each other takes place in vacuum as well as in material media, and irrespective of whether the charges are free or bound to material bodies. Thus consider first *two charges concentrated at two points, having no material support and placed in empty space.* The respective values of the charges, that is, the quantities of electricity constituting them being m and m', and r designating the distance between the charges, each charge will exert a force f upon the other along the straight line mm' numerically equal to

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

as was discovered experimentally under conditions approximating the above. The product mm', and hence the force f, is positive or negative depending respectively on whether the charges have the same or opposite polarities. In the former case the force f tends to set the charges in motion away from each other. In the latter, the force produces mutual attraction of the charges and tends to make them move toward each other.

Field Intensity.—Applying this to the definition of the unit of electric charge or quantity of electricity, a unit charge is that which, placed at a unit distance from an equal charge, is repelled by the latter with unit force. In the practical unit system, 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 is then equal to

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

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

The field intensity of a charge m isolated in space is thus the same for all points of the sphere of radius r having the charge m at its center. The direction of the force at each point of the surface is normal to the latter along the radius 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 an infinite distance from m is the *potential* at point A. The *difference of potential* between two points of the field is hence equal to the work done in moving a unit charge from one point to another. If the two points are a unit of length apart, and the work accomplished in 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 a unit potential difference. In the practical unit system this 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 follows that all points of that sphere are at the same potential.

Electrostatic Field.—Coulomb's equation expresses the effect of a stationary charge m on any other charge. That is, it expresses the peculiar condition around the charge m whereby a force will be exerted on any other charge which may be present, which force produces work when allowed to move this other charge. An amount of latent or *potential energy* is thus stored and distributed in the space surrounding the charge m, part of which is converted into work whenever some other charge placed in the field moves under its action.

The amount of potential energy available at any point, being expressed by the value of potential at that point, may be represented graphically by indicating (*a*) the direction of the force at each point, and (*b*) the potential at each point of the field. The electrostatic field of a single charge m is thus represented in Fig. 1. The direction of the force at any point is along the straight line joining that point to the charge m, as shown for various points by the dotted radial lines. The potential is shown by means of equipotential surfaces, obtained from the value of potential gradient along a radial line of the field, for instance mX, as expressed by Coulomb's equation.

Dielectric Constant.—If, instead of considering electric charges in a vacuum, they are placed in a material insulating medium, such as air, the conditions differ from the preceding ones in that the space surrounding the charges is not empty but contains a number of molecules or atoms, or in other words, a number of sets of small positive and negative charges, each set being arranged in a stable structure as described heretofore to form an atom. Through its electrostatic field the negative charge m will then, in accordance with the above, attract the positive central charge of each atom of the insulating medium and repel the negative electrons, without however attracting or repelling the atom as a whole, the forces exerted by the charge *m* on its positive and negative component charges being equal and opposite. The charge *m* will hence simply produce a distortion of the atomic structure, which may be visualized as a shift of the elec-

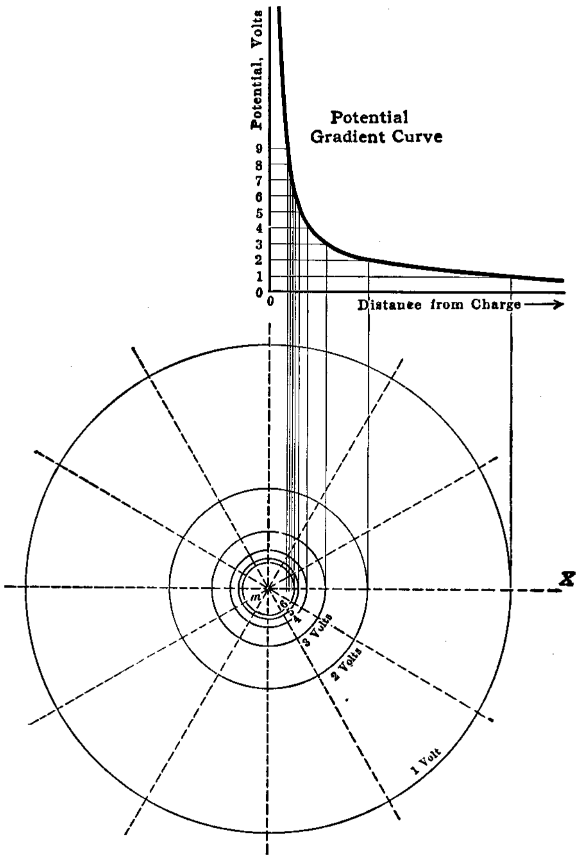


FIG. 1.

tron ring with respect to the positive atom center. This displacement of the component charges of the atoms represents a certain amount of work on the part of the electrostatic field of the charge *m*, and a consequent decrease of its potential energy, as explained in the preceding section, which correspondingly reduces the radii of the equipotential spheres representing the field of the charge *m*. In order to restore the field to the value which it would have in a vacuum, it is then necessary to increase the charge m to some value $m.''$ The ratio

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

obviously depends on the nature of the dielectric substance which surrounds the charge and is called its *dielectric constant* or *specific inductive capacity*. To fix ideas, this is approximately equal to 1.0006 for air, 2.3 for paraffin, 6.6 for mica, 10 for glass, and 80 for water.

In a material medium, Coulomb's relation therefore takes the form

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

also valid for vacuum, for which k is equal to unity.

Electrostatic Induction.—Consider a charge m (Fig. 2) of, say, negative polarity, setting up an electrostatic field as represented

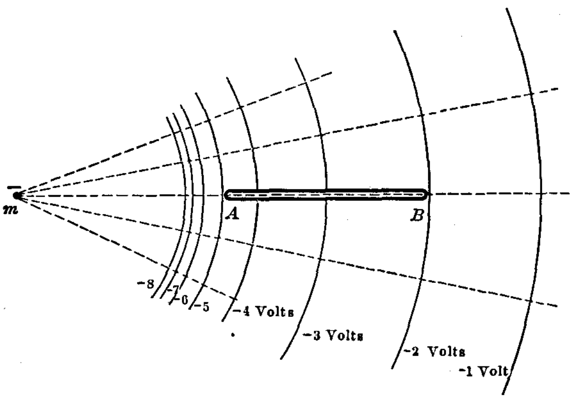


FIG. 2.

in the figure by its equipotential spheres. Let AB represent a rod of a material containing free electrons, a metal for example, placed with its ends lying in different equipotential spheres, for instance, along one of the lines of force of the field. A difference of potential being thus established between the ends A and B, the free electrons are, under the repelling force of the charge m, driven toward the end B of the rod which, being furthest away from the charge, has the weakest negative potential. This causes a deficiency of electrons at the end A of the rod and a corresponding excess of electrons at the end B. These ends therefore become, respectively, positively and negatively charged, since in the end A the sum of the central positive charges of the atoms now slightly exceeds that of the negative electrons remaining in that part, while the reverse is true of the other end B.

The charges thus created at the ends of the rod AB produce electrostatic fields in their vicinity, which combine respectively with that of the charge m, increasing the negative potential around the end B, decreasing it around the end A, and thus tending to neutralize the potential difference originally set up by the field of the charge m between the ends A and B. When this neutralization is complete, the transfer of free electrons ceases within the rod and a condition of equilibrium is reached.

Condenser, Capacitance.—This property is made use of in devices called condensers for storing electrical potential energy in a dielectric (or insulating) medium.

Such condensers consist essentially of two conductors placed some distance apart and insulated from each other. In most cases these conductors are given the shape of parallel plates A and B (Fig. 3). If these plates are connected to a source of continuous potential, such as a battery or direct-current generator G, a potential difference will be established between the two plates, making, for instance, the plate A positive and the plate B negative. This is accomplished by the battery or generator removing some of the free electrons from the metal of the plate A and placing a like number of electrons on plate B, the process constituting a temporary flow or current of electricity along the metallic circuit AGB as in the preceding case. When the electrostatic field created between the two plates by the presence of the charges on them exactly counterbalances the electromotive force of the generator G, the transfer of electricity along the circuit ceases and the condenser is said to be charged.

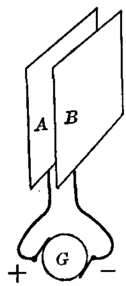


FIG. 3.

The value of the electrostatic field in a given condenser being directly proportional to the charges on its plates (as follows from Coulomb's law), the charge of a condenser is, conversely, directly proportional to the potential difference between its plates, this potential difference being a measure of the electrostatic field. The ratio M/V of the charge of a condenser to the potential difference between its plates has, hence, a constant value C, characteristic of the condenser, called its capacitance or electrostatic capacity. In the practical unit system the unit of capacitance is the farad. This is the capacitance of a condenser which, for a potential difference of 1 volt between its plates, will store a quantity of electricity of 1 coulomb. Subdivisions of this unit frequently used in radio are the microfarad (mfd.) and micro-microfarad (micro-mfd.) which are respectively the millionth part of a farad and of a microfarad.

Since a charge M produces, in a vacuum, the same field as a charge kM in an insulating medium of dielectric constant k, as shown above, it follows that if $C = M/V$ is the capacitance of a condenser in a vacuum, its capacitance in an insulating medium of dielectric constant k will be

$$C' = \frac{kM}{V} = kC.$$

This, incidentally, gives a method for measuring k in measuring the capacitance of a condenser in a vacuum and then of the same condenser having as a dielectric the material under test, the ratio of the latter value to the former giving the value 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 centimetres of one side of one conducting plate, d the distance between the plates in centimetres, and k the dielectric constant of the insulator separating the plates.

Condensers may be built with a fixed or adjustable capacity. The former are generally made up of sets of alternate metal foil and paper, mica films, or glass plates, all even and odd numbered metal foils being commonly connected respectively to form the two "plates" or armatures of the condenser. The latter generally consist of a set of fixed parallel metal plates and a set of movable plates. The movable plates are separated from the fixed ones by air or oil and mounted on a shaft or on rails so that a rotation or a sliding motion will permit a variable portion of the movable plates to be inserted between the fixed plates, thus varying the capacitance of the condenser. Depending on the shapes of the plates, the capacitance variation may be made almost any desired function of their relative motion.

Energy Stored in a Condenser.—The charging process of a condenser consisting, as explained, in the transfer of a charge M from one of its plates to the other, may be considered as effected through the successive transfer, from the one plate to the other, of a large number n of small equal charges m, and

$$m = \frac{M}{n}.$$

The transfer of each charge m then raises the potential difference between the plates by a same amount v so that the final potential difference of the fully charged condenser is

$$V = vn.$$

From the definition of a potential difference, a certain amount of work is performed when transferring any one of the charges m from the one plate to the other, and the energy required for this work is substantially equal (if n is large and hence m and v small) to the product of the charge m by the potential difference existing between the plates at the time of its transfer. The total energy spent for charging the condenser is thus substantially equal to

$$\begin{aligned} W &= v \cdot m + 2v \cdot m + 3v \cdot m + \cdot \cdot \cdot + nv \cdot m \\ &= v \cdot m(1 + 2 + 3 + \cdot \cdot \cdot + n) \\ &= \frac{V}{n} \cdot \frac{M}{n}(1 + 2 + 3 + \cdot \cdot \cdot + n), \end{aligned}$$

and since, as known,

$$1 + 2 + 3 + \cdot \cdot \cdot + n = \frac{n(n+1)}{2},$$

the expression becomes

$$W = \frac{VM}{n^2} \cdot \frac{n(n+1)}{2} = \frac{VM}{2} \cdot \frac{n(n+1)}{n^2}$$

and n being large, the expression $n(n + 1)/n^2$ may be considered as equal to unity, so that

$$W = \frac{VM}{2}.$$

Finally, since, from the preceding section

$$M = CV,$$

the energy spent in charging the condenser is equal to

$$W = \frac{CV^2}{2},$$

which is stored in the condenser in the form of potential energy. This may then perform work when the condenser is allowed to discharge.

Field of a Uniformly Moving Charge.—The charging process of a condenser leads to the study of the conditions attending to the motion of an electric charge, and the flow of electric current.

Consider a single electric charge m (Fig. 4) traveling along a straight line $X'X$ at an unvarying speed s, carrying its electrostatic field along with it. No driving or accelerating force need then be applied to the charge, a force being by definition the cause of a variation in the state of motion(or of rest), and there being no resistance to motion in empty space. Designating by P a stationary point outside the path $X'X$, the radial lines of force of the electrostatic field will, as the charge m moves along $X'X$, sweep in succession past point P with a speed v equal to

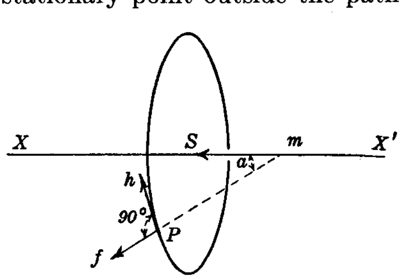


FIG. 4.

$$v = s \cdot \sin a,$$

where a is the angle of $X'X$ with the line mP joining the point P to the position of the charge at the instant considered. The speed v is simply the component of the speed s in a direction normal to the line mP.

As a result of this motion of the electrostatic lines of force past point P, a new force is found to arise at this point in addition to the force of the electrostatic field. This is called a *magnetic force*, for it will cause the deflection of a freely pivoted magnet needle (compass needle) when placed at point P. The *direction* of this magnetic force at point P, which is the direction taken by the pivoted needle under the action of the force, is normal to the plane of mP and $X'X$. It is particularly normal to the direction of the electrostatic force at the point P considered and to the direction of motion of the charge. The *intensity* h of the magnetic force at point P is directly proportional to the electrostatic force intensity f at this point and to the speed of motion v of the electrostatic lines of force past point P. It may be expressed

$$h = fv = \frac{mv}{r^2}$$

where m is the value of the charge and r the distance mP at the instant considered. Thus, the magnetic field intensity, like the electrostatic field intensity, varies inversely as the square of the distance.

The line $X'X$ being an axis of symmetry for the figure, this may be rotated around it, and point P then describes a circle having its center on $X'X$ and its plane normal to $X'X$. At any given instant the electrostatic and magnetic forces are the same, respectively, for all the points of this circle. The electrostatic force is directed along the straight line joining m to the point considered (line of force of the electrostatic field of the charge m) while the magnetic force h is tangent to the circle, which thus constitutes a *line of force of the magnetic field of the moving charge m.*

If any of the factors affecting the magnetic field intensity vary, such as, for instance, the speed or value of the moving electric charge, the circles constituting the magnetic lines of force, and along each of which the magnetic field strength has a constant value, correspondingly shrink or expand.

The magnetic field of the moving charge m represents an ability, on the part of the charge, to perform work by acting on magnetic masses (displacing a pivoted magnet needle, for example). In other words, it represents for the charge m a store of magnetic energy, in addition to its store of electrostatic energy, equal to the amount of energy expended at the time the charge was set into motion. This energy will in turn perform work against retarding or accelerating forces applied to the charge when the state of motion of the latter is to be altered. The magnetic field of the charge thus gives to the charge an effective inertia, which tends to oppose changes in the state of motion.

Summarizing the foregoing, an electric charge possesses an electrostatic field through which it exerts an action on other electric charges. When in a state of motion, it possesses, in addition, a magnetic field through which it can exert an action on magnetic masses with respect to which it is moving.

Conversely, this action of an electric charge on a magnetic mass with respect to which it is moving may be considered as an action of the magnetic mass or field on the electric charge. This reciprocity of interaction of moving electric charges and magnetic masses or fields is a consequence of the relativity of their state of motion. More specifically, the magnetic field of the moving electric charge and the magnetic field with respect to which the charge is moving, add to each other at some points and oppose and neutralize each other at other points, causing forces to arise which tend to draw the electric charge toward the points of reduced magnetic field intensity.

Thus consider, for instance, a negative charge m (Fig. 5), moving at constant speed between the poles of a magnet in a direction Xm perpendicular to the lines of magnetic force H set up by the magnet. As explained before, the lines of magnetic force developed by the charge m as a result of its motion are circles enclosing the path Xm of the charge. One of these circles is shown in the figure, from which it is seen that the magnetic force developed by the charge is, at some points, in the same direction as the lines of magnetic force H of the magnet (points located, in the figure, around the upper portion of the circle), and at other points (the lower portion of the circle) in opposite direction. In accordance with the preceding statement, the charge will be drawn toward these points of weakest magnetic field by a force E directed perpendicularly to both the trajectory Xm of the charge, and the magnetic lines of force H and proportional to (a) the value of the charge m, (b) the intensity of the magnetic field, and (c) the speed of the charge. The charge, being thus deflected from its original direction of travel, and moving in a new direction, its magnetic field will again combine with the fixed field H, and a new deflecting force arises, as may be seen by reasoning as before. The charge, ultimately, describes a circle.

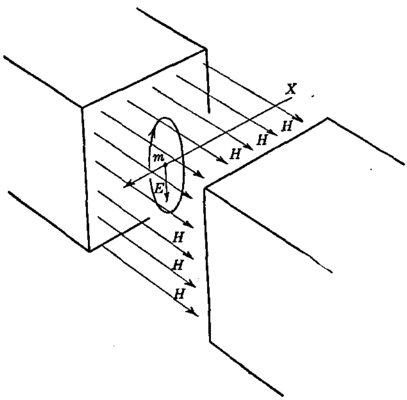


FIG. 5.

Electric Current Intensity.—Referring back to Fig. 3, an electric current was seen to flow in a metal rod having its ends placed in different equipotential surfaces of an electrostatic field, that is, having a potential difference established between its ends. Such an electric current is temporary, however, the transfer of electricity along the rod accumulating electric charges at its ends which neutralize the original electrostatic field.

If by some means the original potential difference is sustained, for instance, by removing the neutralizing charges as they accumulate at the conductor ends, the electric current will flow in a continuous, steady manner. Such a constant potential difference may be established between the conductor ends A and B by inserting the conductor in a closed conducting circuit containing a battery or generator G (Fig. 6). The resulting potential gradient along the circuit, setting the free electrons into motion at all points of the circuit in a common direction, creates a continuous electric current flow. The intensity of this electric current, as measured by the quantity of electricity passing through one cross-section of the wire during a unit of time, is obviously the same for all points of the circuit. In the practical unit system the unit of current intensity is called the ampere.

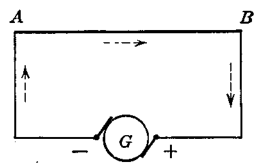


FIG. 6.

Before the application of the electron theory, the electric current was arbitrarily defined as flowing from the positive to the negative end of the conductor. The free electrons being negative, charges move from the negative to the positive terminal. This distinction should be remembered when speaking of electric current and electron current.

Resistance.—During the process of electric conduction through a metal (or other substance) the free electrons, moving in a common general direction to constitute an electric current, enter into frequent collisions with the atoms of the metal (or substance) which happen to be in their path. This tends to reduce the velocity of the electrons, as if a retarding force were applied to them. The number of collisions per unit of time, and hence the equivalent retarding force being directly proportional to the speed of the electrons (since this is equal to the conductor length covered by the moving electrons in a unit of time), is also directly proportional to the intensity of the current flowing in the wire, as defined in the preceding section.

In order to maintain constant the average speed of the electrons and current intensity I, it is then necessary to apply to the electrons a driving or electromotive force E, equal and opposite to the retarding force due to the collisions. As just shown, this force is proportional to the current intensity, and this may be written

$$E = RI,$$

which is the familiar Ohm's law.

The factor R is the *resistance* of the conductor. It is a constant determined by the substance and shape of the conductor, since it depends on the number of free electrons present and on the space arrangement of the atoms within the substance. If r is the resistivity of the material (resistance across opposite faces of a unit cube of the material), the resistance of a conductor of length l and cross-sectional area a is thus

$$R = \frac{rl}{a}.$$

Energy Lost through Resistance.—Work is done whenever a force displaces its point of application along its own direction, equal to the product of the force by the distance traveled by this application point. A steady electric current in a wire being produced by the application of an electromotive force E to its free electrons, which then travel in it at a constant average speed s, work is done during every second by this force E in dragging each electron through the wire, equal to

$$p = sE.$$

If N is the number of moving electrons per unit length of the wire, the total work done per second, or power expended, is then

$$P = Np = NsE.$$

The product Ns being the number of electrons which pass, during one second, through one cross-section of the wire, is by definition equal to the current intensity I, so that

$$P = NsE = IE,$$

and since from Ohm's law $E = RI$, the expression becomes

$$P = IE = RI^2 = \frac{E^2}{R}.$$

The applied electromotive force being, as explained in the preceding section, solely required for the purpose of neutralizing the retarding force due to the collisions between moving electrons and the atoms of the conductor, the energy P thus spent during every second represents the energy gained by the electrons during their periods of free travel, when they accelerate under the effect of the applied electromotive force, and surrendered by them to the atoms when their speed is decreased by a collision. In this manner the atoms receive energy and enter into vibration about their average position of equilibrium, raising the temperature of the conductor substance. The energy expended is thus completely transformed into heat, at the expense of the source of electromotive force.

Inductance.—Since the electromotive force applied to a conductor exactly counterbalances the retarding force due to the conductor resistance, the condition of motion of the electrons in the conductor at a constant average speed may be likened to that of uniformly moving electric charges in a vacuum, as considered in a preceding section.

For points outside the conductor however, the electrostatic (radial) field of these electrons being, as known, neutralized by the positive atom centers of the conductor substance, there only remain the magnetic fields of the moving electrons, the lines of force of which are circles enclosing the electron trajectory. The magnetic field which thus surrounds the conductor when carrying an electric current, being the sum total of the magnetic fields of the individual moving electrons which constitute the current, is therefore directly proportional to the current intensity. This may be expressed by the relation

$$\Phi = LI,$$

where Φ is the magnetic field flux, I the electric current intensity, and L a constant called the *inductance* of the circuit element considered.

The inductance is thus equal to the magnetic 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 unit system, this is called the henry.

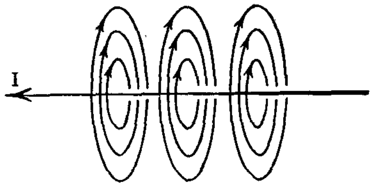


FIG. 7.

FIG. 8.

The inductance thus depending on the same factors as affect the magnetic field surrounding the wire, depends on the shape of the circuit and the nature of the medium surrounding it.

This dependence on the *shape* of the circuit may be readily understood when it is considered that the magnetic fields surrounding various parts of a circuit combine with each other, adding or subtracting according to their relative positions in space. Thus Fig. 7 represents the magnetic field about a straight wire carrying an electric current I. If this wire be bent over itself like a hairpin (Fig. 8) the current will flow up one branch and down the other, and the magnetic fields about the two branches being in opposite directions, neutralize each other. This combination has therefore substantially no inductance.

Similarly, in winding a wire into a coil, the successive turns will carry current in the same direction, creating a considerable magnetic field about the coil, which may thus have a very great inductance. For a single-layer solenoid coil, the inductance may be calculated from the formula

$$L = 4\pi nNS,$$

where n is the number of turns of wire per centimeter length of coil, N the total number of turns, and S the cross-sectional area (in square centimeters) of the coil.

The dependence of inductance on the nature of the surrounding medium is due to the fact that, similar to the electrostatic field, the magnetic field produces an atomic distortion of the medium. This introduces in the formula a multiplying factor, called the *permeability* of the medium. While this is substantially a constant for most substances, it is for some substances, like iron, steel, or nickel, a function of the magnetic flux. The inductance of a coil having, for example, an iron core is therefore variable, and use is made of this property in a number of radio devices.

Magnetic Energy.—As in the case of a single moving electric charge, the magnetic field of an electric current represents a store of energy accumulated during the period when the free electrons were set into motion and the current established.

Let i be the current intensity in a circuit of inductance L at a certain instant of the transient period, and Φ the corresponding magnetic flux around the circuit. When the current increases by an amount di the magnetic 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 its final intensity I, the total energy stored in the magnetic field is

$$W = \int_{t=0}^{t=t_1} Lidi = \tfrac{1}{2}LI^2.$$

Electromagnetic Induction.—Consider a straight wire AB (Fig. 9) in which an electric current of intensity I is flowing from A to B. This current sets up a magnetic field, proportional to its intensity I, the lines of force of which are circles enclosing the wire, as previously explained.

Suppose a negative charge m to be located at a fixed point P outside and in the vicinity of the wire AB, and let H be the magnetic force at that point (directed tangently to the circular line of magnetic force passing through P). If now the current intensity I is varied, say increased, in a continuous manner, the lines of magnetic force, corresponding each to a given magnetic field strength, will expand. The magnetic force H then displaces itself in the direction D perpendicular to it (H thus remaining parallel to itself) at a speed which depends on the rate of change of the current intensity I. The stationary charge m located at point P hence finds itself in a magnetic field the lines of force of which, parallel to H, move in the direction D. It may, therefore, be likened to a charge moving in a direction opposite to D in a fixed magnetic field—a case already studied in connection with Fig. 5, from which it is seen that, here too, an "induced electromotive force" E finds itself applied to the charge m, perpendicular to the directions H and D.

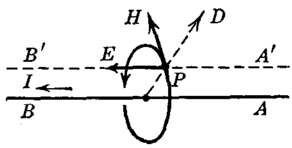


FIG. 9.

The *magnitude* of this force E is directly proportional to that of the magnetic force H and to its displacement speed in the direction D. It is therefore directly proportional to the intensity I of the current and to its rate of change, and is also inversely proportional to the square of the distance between the point P considered and the wire AB.

The *direction* of the force E depends on the polarity of the charge m, on the direction of the force H and on the direction of motion of this force H. In the figure it has been drawn for a negative charge m and for a motion of the force H away from the wire AB corresponding to expanding lines of magnetic force due to increasing current intensity I. It would be reversed if the current were decreasing, or if the current should increase while flowing in a reverse direction, that is, from B to A.

If the electric charge m is one of the free electrons in a wire $A'B'$, the force E will set it into motion along this wire, as well as all the other free electrons of the wire, which are all under similar conditions. If this wire is connected in a closed circuit, an electric current will flow along it which, by definition, is directed oppositely to the electron current. The induced electric current in the wire $A'B'$ is thus seen to flow in a direction opposite to that of the inducing current in the wire AB. This inducing current is the current which, in the wire AB, must be added to or subtracted from the current of intensity I when the latter is varied.

In particular, if the wires AB and $A'B'$ are neighboring portions of successive turns of a coil (Fig. 10) and a varying current is sent through the coil, the current flowing in the portion AB induces, in the portion $A'B'$, a current tending to oppose the current variation. Extending this to the whole coil, a current variation in the coil thus induces in it an electromotive force of such polarity that the resulting induced current tends to oppose or neutralize the inducing current variations. This is a so-called self-induced counter-electromotive force, which is proportional to the rate of change of the current. Thus,

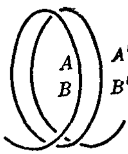


FIG. 10.

$$E = -L\frac{di}{dt}.$$

The factor L is the inductance of the coil, as defined previously. The minus sign expresses the fact, that the electromotive force is directed oppositely to that producing the current change.

Mutual Inductance.—As stated above, induction takes place, as well, in a circuit entirely separate from the inducing circuit. It is simply necessary that a part of the varying magnetic field of the first (inducing) circuit be made to link with the second circuit. The induction effect is of course smaller the greater the distance between the two circuits, since the field intensity varies in inverse proportion to the square of the distance.

If a unit current is made to flow in a circuit A, then the flux M linking with a 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 with circuit A.

From what has been said before, 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 secondary 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 electromotive force or current in the circuit. In this case the mutual inductance M is zero.

Electromagnetic Radiation of Energy.—Consider a straight wire AB (Fig. 11) carrying an electric current of intensity I flowing from A to B. Let H and K represent two lines of magnetic force due to this current, the line K being of somewhat greater diameter than the line H (and hence corresponding to a weaker magnetic force). Suppose the current intensity to be increased to some value I'. The magnetic field about the wire being increased in proportion, the lines of force along which the magnetic force has the same values, respectively, as along the original lines H and K are now circles H' and K' of greater diameters than before.

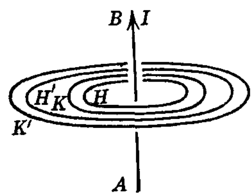


FIG. 11.

Irrespective of how slowly or quickly the current intensity has been made to increase from the value I to the value I', the same amount of time as required for this increase is also required for the lines H and K to expand into, respectively, the lines H' and K'. But the line H will *start* to increase in size, and also will *cease* to expand a little before the line K. This time lag is due to a property of the electromagnetic field, whereby *changes* in it (such as the increase of magnetic field strength for instance) are not transmitted instantly throughout the field, but *propagate* from the change center or source (which here is the wire AB) outward at a constant speed, characteristic of the surrounding medium, and entirely independent of the slowness, abruptness, or other characteristics of the change considered.

This speed, as results from both experimental and theoretical investigation, is equal to the speed of propagation of light, 300,000 kilometers per second (182,000 miles). This propagation process may be compared to the propagation of a ripple or wave over the smooth surface of a liquid. But taking place in all directions of space, the wave is not circular but spherical, like a sound wave in air.

The same remarks apply, of course, to the case when the current intensity in the wire AB is decreased from some value I'

to some other value I: the circular lines of magnetic force H' and K' shrink into the lines H and K respectively, at a speed depending solely upon the rate of change of the current in the wire, and with the line H' beginning and ending its motion a little before the line K'.

Simultaneously with this motion of the magnetic lines of force there appears at every point an electrostatic force, as explained in a previous section, perpendicular to the circular lines of magnetic force (hence parallel to AB), proportional to the speed of motion (expansion or contraction) and directed in one way when the lines expand, in the other when they shrink. These two induction fields (magnetic and electric) vary inversely as the square of the distance from the wire AB.

Now, as long as the electric current in the wire AB retains a constant intensity I or I' the electrons in it travel at a constant average speed. The constant electromotive force which is applied at the wire ends, it was shown, then counterbalances exactly the retarding effect of the wire resistance. Both this and the constant applied electromotive force may be disregarded, the electrons being then considered as moving at constant speed in a wire having no ohmic resistance and having no electromotive force applied to it.

A *change* in current intensity, however, meaning a change in the electron speed, requires the application of some accelerating or decelerating force to the electrons (which electromotive force, in the actual case, will add to or subtract from the constant, resistance neutralizing electromotive force applied to the wire). Since this force displaces its points of application (the moving electrons) along its own direction, it accomplishes positive work, and the energy thus spent represents the energy changes of the electromagnetic field, which, as stated, propagate with the speed of light in the form of a wave. In other words, the energy spent is removed from the wire (*radiated*) in the form of a spherical wave having the circuit element AB at its center, and distributed throughout space in the electromagnetic field.[1]

[1] If a is the acceleration of the individual electrons, the speed s at some instant t is equal to

$$s = at,$$

counting the time from the instant the accelerating force was applied to the electrons at rest.

Consider then a fixed point P (Fig. 12) and let w represent the energy flow or energy current passing through an elementary unit area of the sphere surface containing point P. As mentioned before, this has two "components"[1] h and e which are respectively the magnetic and electrostatic components, and which are related to the energy vector by the vector equation

$$w = h \cdot e.$$

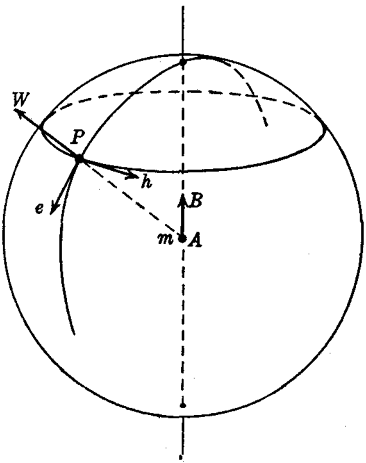


FIG. 12.

It will be noted that the energy vector is inversely proportional to the square of the distance r of the point P considered: the radiated energy, crossing in its outward motion successive concentric spheres of increasing radius, the surface area of which increases as the square of the radius (that is, as r^2), and the energy per unit area of sphere surface decreases in inverse proportion to r^2.

The distance d covered by each electron during a short time interval Δt being substantially equal to

$$d = s \cdot \Delta t = at \cdot \Delta t$$

the energy radiated by each electron during this time interval Δt (work done by the accelerating force in driving the electrons over the distance d) is hence

$$w = a \cdot d = a \cdot s \cdot \Delta t = a^2 \cdot t \cdot \Delta t$$

and is thus proportional to the square of the acceleration.

[1] These are not geometrical components, combining in accordance with the parallelogram law, but vector components following the rules of vector algebra.

A consequence is that the components e and h of the radiated energy decrease each in inverse ratio to the *first power* of the diameter r. This, together with their propagation speed, radically differentiates the *radiation fields* from the induction fields previously considered, which were seen to decrease inversely as the *square* of the distance, and which therefore become negligibly small at any great distance from the circuit as compared with the radiation fields.

Furthermore, the radiation fields being directly proportional to the *rate of change* of the current, considerable radiation fields may be set up by periodically and frequently reversing the electric current in the circuit—in other words by using high-frequency alternating currents. The radiation fields will then alternately reverse with the same frequency as the current in the radiating circuit. When sweeping over any distant circuit they will in turn set up in it an alternating current of like frequency and of an amplitude directly proportional to that of the current in the radiating circuit.

Underlying Principles of Radio Communication.—Radio communication is effected through a transfer of electrical energy from the transmitting station to the distant receiving station, achieved through the medium of radiation fields.

These, consisting as shown above, of interlinked electric and magnetic fields, are set up at the transmitting station by high-frequency alternating currents generated by appropriate apparatus, which energize so-called antenna circuits, designed to radiate large amounts of energy, either equally in all directions, or else more particularly in certain directions, depending upon the purpose in view. The amounts of energy radiated are varied or modulated in accordance with the signals to be transmitted.

The modulated radiation fields, sweeping over the receiving station, energize its circuits which are, for this purpose, given suitable dimensions, shape, and position (receiving antenna) and the resulting high-frequency alternating currents of modulated amplitude actuate translating devices and serve to reproduce the original modulations.

Before describing the methods, apparatus, and devices used for these various functions, a brief survey of some alternating-current phenomena is made in the following chapter, which are of particular importance for the study of radio circuits.

CHAPTER II

PROPERTIES OF OSCILLATORY CIRCUITS

FORCED OSCILLATIONS

Capacitive Reactance.—Consider a condenser of electrostatic capacity C having, as in Fig. 3, its plates A and B connected respectively to the positive and negative terminals of a source of continuous potential, a direct-current generator G, for instance, so that a potential difference v (equal to the generator voltage), is established between the plates. As explained in the preceding chapter, this condition is the result of the transfer, around the circuit AGB, of a certain number of (negative) electrons from the plate A to the plate B, the condenser charge being then

$$m = Cv.$$

The condenser having thus been charged, no further motion of electricity takes place in the system, and a stable condition is reached. If now the generator voltage is gradually increased at a constant rate, that is to say by a certain fixed number of volts during every second, the condenser charge correspondingly increases, the voltage v and charge m being at every instant tied by the above relation. In other words, during every second the same number of electrons are transferred from the plate A to the plate B, through the circuit AGB, creating a constant flow or current of electricity around this circuit.[1] The *intensity* of this current increases and decreases with the *rate of change* of the potential difference between the condenser plates. The *direction* of current flow reverses with the *direction of variation* of the potential difference. Thus, i being the current intensity, as measured by the charge (or number of electrons) passing through a cross-section of the circuit during every second, and dv/dt and dm/dt

[1] Although the free electrons which constitute the current around the condenser circuit do not cross the dielectric separating the condenser plates- it is often said that a varying potential difference between the plates produces a current flow *through* the condenser.

being the rates of variation of the condenser potential difference and of the charge respectively,

$$i = \frac{dm}{dt} = C\frac{dv}{dt}.$$

If, instead of continually varying in a same direction, the potential difference alternately increases and decreases, the condenser current flows in alternately opposite directions. The average rate of change of a thus pulsating or alternating potential difference being directly proportional to the frequency and amplitude of the potential variation, the current intensity is also proportional to these two quantities.

In particular, if a sinusoidal electromotive force

$$e = E_0 \sin 2\pi ft = E_0 \sin \omega t$$

of frequency f (or of cyclic frequency $\omega = 2\pi f$) and maximum value E_0 is impressed upon the condenser, the corresponding condenser-current intensity is

$$i = C\frac{de}{dt} = \omega CE_0 \cos \omega t,$$

and calling I and E the effective values of condenser current and voltage,

$$I = \omega CE,$$

from which

$$\frac{E}{I} = \frac{1}{\omega C} = X_c.$$

The expression $X_c = \dfrac{1}{\omega C}$ is called the *reactance* of the condenser.

These results are represented graphically in Fig. 13, where the instantaneous values of the alternating electromotive force e and condenser current i are plotted against time. The rate of change of the voltage, which is equal at every instant to the slope of the curve "e," is seen to be positive when the voltage passes from its negative to its positive maximum, and greatest when the curve crosses the time axis. This corresponds to a condenser current i flowing in the circuit in a positive direction, and having its maximum value when the voltage becomes zero. The rate of change of the voltage reverses and becomes negative when the voltage passes from its positive to its negative maximum, and is greatest when the "e" curve crosses the time axis. This corresponds to a current flowing in the circuit in the opposite (negative) direction, having its maximum intensity when the voltage is zero.

The condenser current thus passes through its positive and negative maxima respectively one-quarter period *before* the condenser voltage. It is said to "lead" the voltage by 90 deg., or what amounts to the same, the voltage is said to "lag" behind the current by 90 deg.

These same results are represented vectorially in Fig. 14, the sinusoidal electromotive force and current being represented by

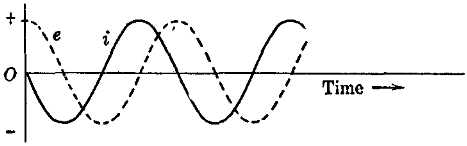


FIG. 13.

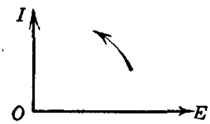


FIG. 14.

vectors OE and OI, respectively, having a common origin O and forming an angle of 90 deg., and rotating about this point O at constant speed, making during every second a number of turns equal to the frequency f of the electromotive force and current. The current vector OI, leading the voltage, is directed in such a way as to assume, in its rotation, the successive positions which the voltage vector will assume one-quarter period *after* it.

Considering now the condenser reactance $X_c = 1/\omega C$, suppose the condenser to be connected to an alternator generating an alternating electromotive force of gradually increasing frequency. The reactance, varying inversely as the frequency, is represented, in function of the latter, by a hyperbola as in Fig. 15. This simply means that for an alternating voltage of unvarying maximum amplitude, the condenser current is small at low frequencies and increases directly as the frequency due to the fact, stated above, that the rate of change of the voltage increases as the frequency.

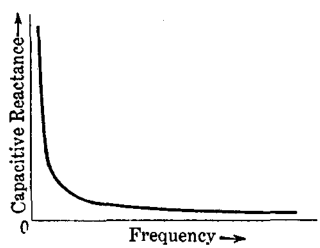


FIG. 15.

Inductive Reactance.—It was shown in the preceding chapter that if a current of intensity i flowing through a coil of inductance L is varied, an electromotive force e is induced in the coil equal to

$$e = -L\frac{di}{dt}.$$

If the variable current i is a sinusoidal alternating current of maximum intensity I_0

$$i = I_0 \sin \omega t$$

then the self-induced electromotive force across the coil is

$$e = -L\frac{di}{dt} = -\omega L I_0 \cos \omega t$$

and, calling I and E the effective values of current and voltage,

$$E = \omega L I,$$

from which

$$\frac{E}{I} = \omega L = X_L.$$

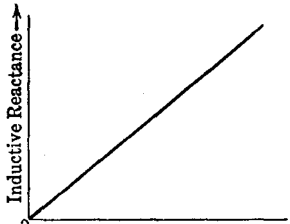


FIG. 16.

The quantity $X_L = \omega L$ is called the *reactance* of the coil. It increases directly with the frequency of the alternating current flowing through the coil, and expresses the fact that the higher the variation frequency of the current, the greater the electromotive force induced across the coil. This dependence of the coil reactance on the frequency is represented graphically in Fig. 16.

Representing now, in function of time, the sinusoidal alternating current i flowing through the coil (Fig. 17) the self-induced electromotive force e is proportional to the instantaneous rate of change of the current and of opposite polarity, and is thus seen to

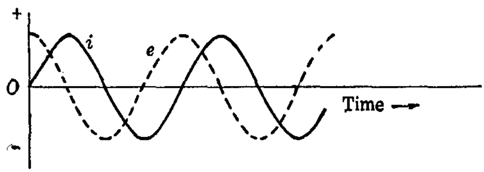


FIG. 17.

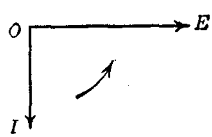


FIG. 18.

"lead" the current by 90 deg. The equivalent vector representation is given in Fig. 18.

Impedance.—Suppose now an alternating current

$$i = I_0 \sin \omega t$$

to be flowing in a reactance X (inductive or capacitive) connected in series with a resistance R (Fig. 19). This current, flowing through the resistance, produces a potential difference e_R between its ends M and N which, according to Ohm's law as stated in the preceding chapter, is equal to

$$e_R = iR = RI_0 \sin \omega t.$$

This potential difference thus passes through its positive and negative maxima, and becomes zero, at the same time, respectively, as the current i. In other words, it is in phase with it. The corresponding relation between maximum or effective values is

$$E_R = RI.$$

This same alternating current i passing through the reactance X sets up between its ends N and P a potential difference e_x which is in quadrature with the current, leading it or lagging behind it, depending on whether the reactance is capacitive or inductive, as explained in the preceding section. Thus,

$$e_X = iX$$

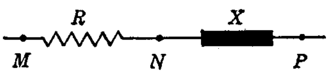


Fig. 19.

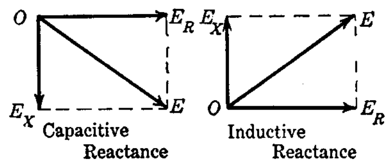


Fig. 20.

and the corresponding relation between maximum or effective values is

$$E_X = XI.$$

The overall potential difference between the ends M and P of the circuit branch MNP is, at every instant, equal to the sum of the instantaneous values e_R and e_X of these two potential differences, thus

$$e = e_R + e_X.$$

But, owing to the fact that these two potential differences are not in phase with each other, the maximum or effective value E of the overall potential difference e is not equal to the sum of the maximum or effective values E_R, E_X of the individual potential differences e_R, e_X, it being necessary to take this phase difference into account. Representing the conditions vectorially (Fig. 20) the total potential difference E is seen to be

$$E = \sqrt{E_R^2 + E_X^2} = I\sqrt{R^2 + X^2} = IZ$$

where $Z = \dfrac{E}{I} = \sqrt{R^2 + X^2}$ is called the *impedance* of the circuit branch MP. For respectively capacitive and inductive reactance X, it is equal to

$$Z = \sqrt{R^2 + \frac{1}{\omega^2}C^2}$$

and

$$Z = \sqrt{R^2 + \omega^2 L^2}.$$

Also, the potential difference E is seen to lead the current I, or lag behind it (depending on the nature of the reactance) by an angle a such that

$$\tan a = \frac{E_X}{E_R} = \frac{X}{R}.$$

Like the reactance, the impedance and phase angle are thus functions of the frequency of the current flowing through the circuit. This is of fundamental importance in radio, for it permits, through combinations of inductive and capacitive reactances in a given circuit system, to constitute impedances which may be almost any desired function of the frequency of the applied electromotive force or of the current flowing through it. The current, voltage, and phase relations being, in such circuits, primarily dependent on the frequency, it then becomes possible to separate currents or electromotive forces of different frequencies selectively. Some simple cases are studied in the following paragraphs.

Series Resonance.—Consider an alternating current of maximum or effective intensity I and cyclic frequency ω flowing through an inductance L connected in series with a capacity C and resistance R (Fig. 21). As explained before, this current produces alternating potential differences E_L, E_C, and E_R between the ends MN, NP, PS of the inductance, capacity, and resistance respectively equal to

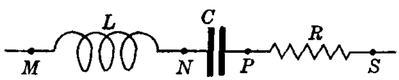


FIG. 21.

$$E_L = \omega L I,$$

$$E_C = \frac{I}{\omega C},$$

$$E_R = RI.$$

The potential difference E_R is in phase with the current I, while the potential differences E_L and E_C, respectively, lead the current and lag behind it by 90 deg. These two potential differences are hence 180 deg. out of phase with each other. In other words, they are in phase opposition, the one reaching its positive maximum when the other reaches its negative maximum, and conversely. The potential difference E_X between the points M and P is therefore equal to the difference between E_L and E_C,

$$E_X = E_L - E_C = \omega LI - \frac{I}{\omega C} = I\left(\omega L - \frac{1}{\omega C}\right).$$

This potential difference E_X leads the current I or lags behind it, depending, respectively, on whether the inductive reactance ωL is greater or smaller than the capacitive reactance $1/\omega C$. Now this, for fixed inductance L and capacity C, depends *solely* upon the frequency ω of the current flowing in the circuit; and if this frequency be varied (say, increased) from a very small value up, the inductive reactance ωL will increase directly as the frequency, while the capacitive reactance $1/\omega C$ decreases, as already shown in Figs. 16 and 15.

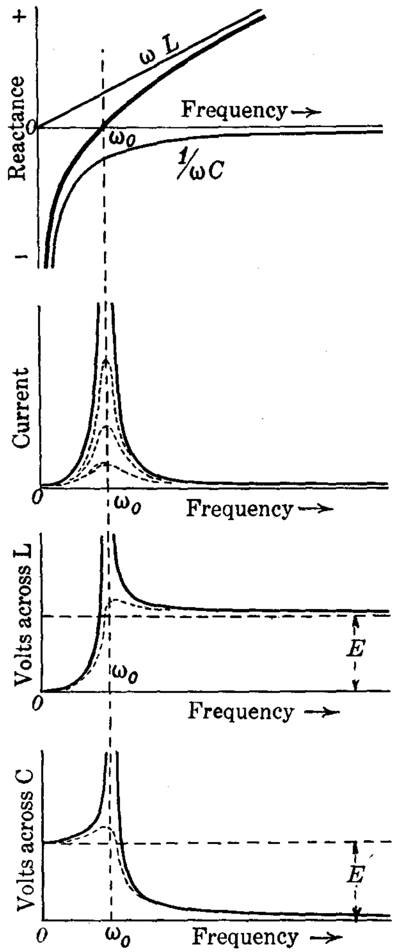


FIG. 22.

The resultant reactance $X = \omega L - 1/\omega C$ being equal to the difference of the two, these two Figs. 16 and 15 are reproduced on a single diagram (upper diagram of Fig. 22), the inductive reactance being, for greater clearness, plotted positively while the capacitive reactance is plotted negatively. The resultant reactance, represented by the heavy line curve in the diagram, is thus, at the smaller frequencies, of the nature of a large capacitive reactance. As the frequency increases, the value of this resultant reactance decreases until, for a given frequency ω_0 the inductive and capacitive component reactances being equal, the total reactance is zero. Thus,

$$X = \omega L - \frac{1}{\omega C} = 0,$$

from which

$$\omega = \frac{1}{\sqrt{LC}} = \omega_0$$

which is called the *resonance frequency* of the circuit. For greater values of the frequency, the total reactance now becomes inductive, and increases with the frequency.

Now, saying that the total reactance X is zero at the frequency ω_0 means, according to the definition of reactance, that for this frequency

$$X = \frac{E_X}{I} = 0,$$

E_X being, as stated above, the potential difference between points M and P of the circuit (Fig. 21) and I the current flowing through it.

If, then, an alternating current of maximum or effective intensity I is flowing in the circuit, the potential difference E_X must be zero, or else, if the circuit is connected to the terminals of an alternator generating an alternating electromotive force of fixed maximum amplitude E, and the resistance R of the circuit is so small that it may be neglected, then the alternating current I flowing in the circuit must be infinite. This latter condition is represented by the full line *resonance curve* in the second diagram of Fig. 22 where the intensity of the alternating current flowing in the circuit is plotted in function of the frequency of the applied alternating electromotive force, the latter being assumed to be of constant maximum amplitude at all frequencies.

Actually, the resistance R of the circuit is never equal to zero, and limits the current intensity, which hence does not become infinite, but simply passes through a maximum at resonance frequency, as shown for various values of resistance by the dotted line curves in the second diagram of Fig. 22.

The circuit impedance being, according to the preceding section,

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

this is, in the present case,

$$Z = \sqrt{\left(\omega L - \frac{1}{\omega C}\right)^2 + R^2}$$

When the frequency is such that the reactance is zero, the impedance reduces to

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

and the current flowing in the circuit is then equal to

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

E being the applied electromotive force. This current is then in phase with the applied electromotive force. When the frequency is different from the resonance frequency ω_0, the reactance is no longer zero, and the current leads the applied electromotive force, or lags behind it, by an angle a such that

$$\tan a = \frac{X}{R} = \frac{\left(\omega L - \frac{1}{\omega C}\right)}{R}$$

according to whether the frequency is, respectively, greater or smaller than the resonance frequency ω_0.

Referring again to diagram 2, Fig. 22, the shape and maximum height of the resonance curve depend in each case on the value of the resistance R. The greater this resistance, the smaller the maximum, and also the flatter or broader the curve, so that the circuit is correspondingly less selective, responding more equally to all frequencies instead of showing a markedly increased current at one particular "resonance frequency."

As a conclusion and summary of the above discussion, when a circuit contains inductance, capacity, and resistance in series and is energized by a source of alternating current, the intensity of the current flowing in the circuit is (if the resistance is slight) negligibly small at all frequencies with the exception of one particular frequency, called resonance frequency and equal to $1/\sqrt{LC}$, when the current passes through a sharp maximum.

If the frequency of the energizing source cannot be altered, resonance may be established by altering either or both the circuit inductance L and capacity C, in order that the expression $1/\sqrt{LC}$ will become equal to the frequency of the energizing source. This process of bringing the circuit into resonance with the source frequency is called *tuning*. If several alternating-current sources of different frequencies energize the circuit simultaneously, the circuit may, through tuning, be successively brought into resonance with each of these sources, the current in the circuit rising sharply every time the circuit is in resonance or in tune. In particular, if the electromotive force impressed on the circuit comprises a plurality of different frequency components, the circuit may be tuned to any one of these components, and the current flowing in the circuit will then be substantially proportional to the amplitude of the corresponding electromotive force frequency component.

Finally, the potential differences $E_L = \omega LI$ and $E_C = I/\omega C$ between the inductance and capacity terminals MN and NP, respectively are functions of both the frequency and current intensity. For an alternating applied electromotive force of constant maximum amplitude and variable frequency, the current varying with the latter as shown in diagram 2 of Fig. 22, the potential differences E_L, E_C will in turn vary as shown in diagrams 3 and 4, of Fig. 22. The full line curves correspond to zero resistance, the dotted line curves refer to cases where resistance is associated with the inductance and capacity. The potential difference between the inductance terminals is thus small when the frequency of the applied electromotive force is less than the resonance frequency, becomes a maximum at resonance, and tends to become equal to the impressed electromotive force E above the resonance frequency.

The potential difference between the condenser terminals varies in reverse manner, being substantially equal to the impressed electromotive force E at frequencies below the resonance frequency, and becoming negligibly small at frequencies greater than this.

This permits the *filtering* out of electromotive forces having frequencies below or above, as the case may be, a particular frequency ω_0. Depending on whether the potential difference between the inductance terminals or condenser terminals is used, the circuit may thus serve as a so-called high-pass or low-pass filter. However, the abruptness of the cut-off is in most cases too small for practical purposes, and more complicated circuit combinations are then resorted to, as pointed out in a later section of the present chapter.

Parallel Resonance.—The case of Fig. 23 differs from the preceding one in that the capacity C and inductance L are connected in parallel across the energizing source, represented here as an alternator generating an electromotive force

$$e = E \sin \omega t,$$

which is impressed simultaneously upon the condenser and inductance coil. Thus, while in the case of the series circuit, the capacity and inductance were carrying a similar current, which would set up across each of them electromotive forces depending upon the frequency of the current, the capacity and inductance have, in the present case of parallel connection, a like impressed electromotive force, and carry currents which depend on their respective impedances, that is, on the frequency.

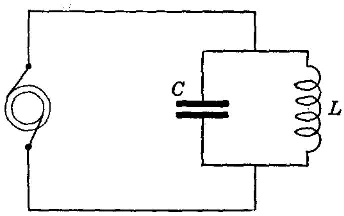


FIG. 23.

Thus, the current i supplied by the alternator divides between the condenser and inductance-coil branches in inverse proportion to their respective impedances. Referring to maximum or effective values, and temporarily neglecting all resistances, the condenser current I_C and inductance coil current I_L are equal to

$$I_C = \frac{E}{X_C} = \omega CE,$$

$$I_L = \frac{E}{X_L} = \frac{E}{\omega L}.$$

Since the condenser current leads the applied electromotive force E by 90 deg. while the inductance-coil current lags behind this same electromotive force by 90 deg., the currents in the two branches flow in opposite directions, and the alternator current is equal to their difference. Thus,

$$I = I_L - I_C = \frac{E}{\omega L} - \omega CE = E\left(\frac{1}{\omega L} - \omega C\right).$$

If then the alternator frequency is gradually increased, the factor $1/\omega L$ will continually decrease while the factor ωC constantly increases. Their difference therefore becomes zero for a frequency given by the relation

$$\frac{1}{\omega L} = \omega C,$$

and thus equal to

$$\omega = \frac{1}{\sqrt{LC}} = \omega_0.$$

At this frequency, the currents in the two branches are equal and opposite, and the current supplied by the alternator is equal to zero. In other words, an alternating current of intensity $I_L = I_C$ and of frequency ω_0 will then flow in the closed circuit formed by the two branches L and C, but *no* current will flow through the alternator. At this frequency there is therefore no interchange of energy between the alternator and the circuit LC, and the alternator may consequently be entirely disconnected and removed from the circuit, in which the current will, nevertheless, continue to flow indefinitely, constituting in this circuit a so-called *free electrical oscillation.* This condition is taken up in detail in later section of this chapter.

Actually, the resistance being not zero, the alternator will, at resonance, supply a current to the circuit, equal to

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

and the alternator current, instead of becoming zero, simply passes through a minimum.

From the above relation

$$I = E\left(\frac{1}{\omega L} - \omega C\right)$$

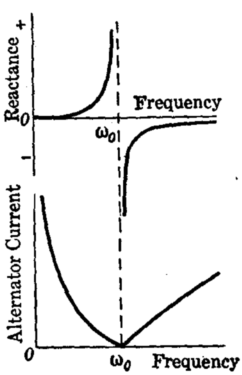


FIG. 24.

the reactance of the circuit connected to the alternator (Fig. 23) is seen to be equal to

$$X = \frac{E}{I} = \frac{1}{\dfrac{1}{\omega L} - \omega C} = \frac{\omega L}{1 - \omega^2 LC}.$$

If the alternator frequency is gradually increased, the reactance, which is inductive for small values of ω, increases in magnitude, becomes infinite when $1 - \omega^2 LC = 0$, that is, when $\omega = 1/\sqrt{LC} = \omega_0$, after which, for frequencies greater than this, it becomes a capacitive reactance of gradually smaller magnitude.

These conditions are represented in the upper diagram of Fig. 24. The lower diagram represents the alternator current in function of frequency, and summarizes the above discussion.

Resonance in Coupled Circuits.—Consider a closed circuit A (Fig. 25) containing inductance and capacity in series (all resistance being neglected for simplicity) and energized by an alternator S connected in the circuit, or coupled to it in any suitable manner, generating in the circuit an alternating electromotive force of constant maximum amplitude E. If the alternator frequency is varied, the current will, as explained before, pass through a maximum at resonance.

Suppose a second closed circuit B having the same inductance and capacity, respectively, as circuit A, to be coupled to this circuit, its inductance coil linking for instance with some of the magnetic lines of force set up by the alternator current flowing in the coil of circuit A. An alternating electromotive force will then be induced in circuit B, which in turn sets up an alternating current in this circuit.

Suppose first that the two circuits A and B are so *loosely coupled* that the alternating current flowing in circuit B does not set up, through its alternating magnetic field, any appreciable electromotive force in circuit A. As the alternator frequency is varied, resonance phenomena will then take place in circuit A as if circuit B were not existant.

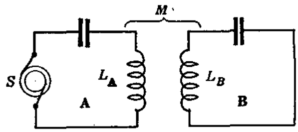


FIG. 25.

On the other hand, the electromotive force induced in circuit B, being of the same frequency as the alternator electromotive force, will produce similar resonance effects in circuit B. That is, the current in this circuit will be a maximum when the alternator frequency is equal to the resonance frequency $1/\sqrt{LC}$ of the circuit.

However, while the alternator electromotive force which energizes circuit A is assumed to be constant at all frequencies, the electromotive force induced in circuit B, being proportional to the inducing current (which is the current in circuit A), will be a maximum when the frequency is the resonance frequency of the circuit. This effect will add itself to that of the alternator frequency variation, so that the resonance curve of circuit B, representing the current intensity in this circuit in function of alternator frequency, will be more peaked than that of circuit A, although the actual intensity of the current in circuit B is, in view of the looseness of the coupling between the two circuits, smaller than that in circuit A.

Tightening the coupling between the circuits seems an obvious way of increasing this induced current. But then the portion of the magnetic field of one circuit linking with the other circuit becomes an appreciable fraction of its total magnetic field, and it is necessary to take into account the electromotive force induced back into circuit A by the current flowing in circuit B. As will be shown presently, this alters the conditions completely.

Saying that the circuits are *closely coupled* is equivalent to considering them as having a part of their respective total inductance in common (if they are inductively coupled; more generally they may be considered as having a portion of their reactance in common). In effect, it makes circuit B a part of the circuit system which is connected to the alternator.

Thus L_A and L_B, being the total inductance in the circuits A and B, respectively, (Fig. 25) and M their mutual inductance, an equivalent system will be that of Fig. 26 in which two directly coupled circuits have each respectively the same total inductance as the circuits of Fig. 25 but have a part M of their total inductance in common. Thus,

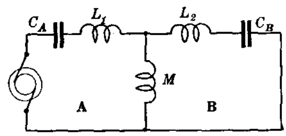


FIG. 26.

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

The alternator generating an alternating electromotive force of adjustable frequency ω, the resonance points of this system may readily be found by computing its reactance. Thus, consider the circuit branch C_AL_1 as connected in series with the parallel arrangement of the two branches C_BL_2 and M. The reactance of C_AL_1 being, as shown in a previous section,

$$\omega L_1 - \frac{1}{\omega C_A},$$

while that of the parallel arrangement of C_BL_2 and M is

$$1 \Bigg/ \left(\frac{1}{\omega M} + \frac{1}{\omega L_2 - \frac{1}{\omega C_B}} \right) = \frac{\omega M\left(\omega L_2 - \frac{1}{\omega C_B}\right)}{\omega\left(L_2 + M\right) - \frac{1}{\omega C_B}},$$

the total reactance of the circuit of Fig. 26 is

$$X = \omega L_1 - \frac{1}{\omega C_A} + \frac{\omega M\left(\omega L_2 - \frac{1}{\omega C_B}\right)}{\omega(L_2 + M) - \frac{1}{\omega C_B}}.$$

The value of frequency for which the alternator current is a maximum is then that which makes this reactance equal to zero (since by definition $I = E/X$), the current being then, as in the simpler cases considered before, limited by the circuit resistance.

Assuming each of the two circuits A and B to have a same resonance frequency ω_0 when considered separately, that is, when not coupled to each other, equal to

$$\omega_0 = \frac{1}{\sqrt{L_A C_A}} = \frac{1}{\sqrt{L_B C_B}},$$

and representing the coupling of the two circuits by their so-called coupling coefficient

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

it is found that the above value of the reactance X of the circuit combination becomes equal to zero for *two* values of the alternator frequency,[1] equal to

$$\omega_1 = \frac{\omega_0}{\sqrt{1+k}}$$

and

$$\omega_2 = \frac{\omega_0}{\sqrt{1-k}}.$$

Thus, when the alternator frequency is varied continuously, the alternator current passes through two maxima, corresponding to the two resonance frequencies ω_1 and ω_2, respectively. It is seen also that for zero coupling (or, in practice for very loose coupling), when $k = 0$, there is only one maximum of current, for the frequency ω_0 which is the resonance of each of the two uncoupled circuits. As the coupling is made tighter, two resonance frequencies ω_1 and ω_2 appear, one below and one above the uncoupled individual frequency ω_0 until, for maximum coupling, when $k = 1$, the two frequencies are

$$\omega_1 = \frac{\omega_0}{\sqrt{2}}$$

and

$$\omega_2 = \infty,$$

[1] In the more general case where the two circuits have *different* resonance frequencies ω_A and ω_B when not coupled together, the coupled circuit system has, as above, two resonance frequencies, equal to

$$\omega = \sqrt{\frac{\omega_A{}^2 + \omega_B{}^2 \pm \sqrt{(\omega_A{}^2 - \omega_B{}^2)^2 + 4k^2\omega_A{}^2\omega_B{}^2}}{2(1-k^2)}}.$$

One of these frequencies is smaller than the smallest of the two frequencies ω_A, ω_B while the other is greater than the greatest frequency ω_A, ω_B.

When making $\omega_A = \omega_B$ this expression simplifies and gives for ω the two values ω_1 and ω_2 given in the text.

which corresponds to the case when the two circuits have all of their inductance in common.

These results are represented in Fig. 27, where the curves 1, 2, 3 and 4 correspond to increasing values of the coupling coefficient. It should be noted that, for reasonably small values of coupling, such as used for curve 2 for instance, the two-current maxima are comparatively close together, and the curve, while rising and falling steeply, has a substantially flat top, contrasting with the single, sharp peak of the single-circuit curve (curve 1). Thus, while a single oscillatory circuit will carry an appreciable current at substantially one frequency only, an arrangement of two such

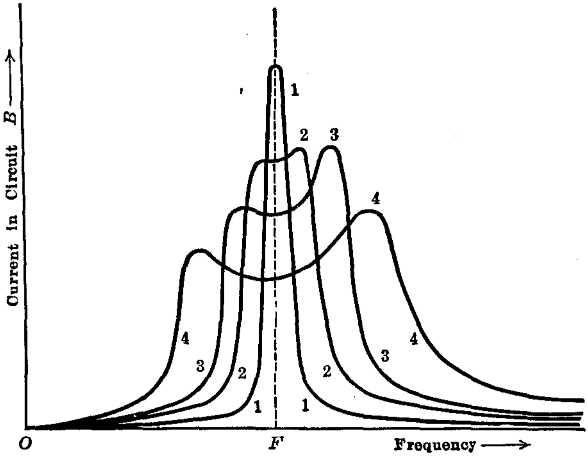


FIG. 27.

circuits, having individually a same resonance frequency and coupled not too tightly, will carry an appreciable alternating current over a band of frequencies, that is, at frequencies comprised approximately between the two resonance frequencies of the circuit-pair arrangement, which may hence serve as a *frequency band filter*.

However, such a two-circuit or two-cell filter may be used for narrow bands only, for when the coupling is made tighter the two resonance frequencies separate and the arrangement will respond, substantially, to these two well-defined frequencies only.

Finally, it is important to note that the currents in the two coupled circuits A and B are substantially in phase with each other at the lower of the two resonance frequencies, while they are 180 deg. out of phase (that is, in phase opposition) at the higher resonance frequency, a condition which finds its application notably in certain vacuum tube circuits, as pointed out later.

FREE OSCILLATIONS

Oscillatory Discharge of a Condenser.—*Physical Explanation.* Consider a condenser C (Fig. 28) 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 become charged. That is, an electric current will for a short time flow in the metallic circuit in the direction $EB'BCAA'E$,

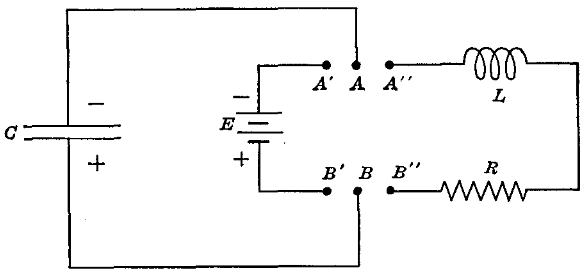


FIG. 28.

accumulating a negative charge on the upper condenser plate and a positive charge on the lower plate, as explained in a previous section. This charging current flows in the circuit until the electrostatic field created between the condenser plates by the charges on them exactly equals and counterbalances the battery electromotive force. The condenser being then fully charged, the amount of energy stored in it as potential energy is, as known,

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

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

Suppose now that the condenser, having thus been charged, is connected to the circuit LR, by disconnecting A and A' from B and B', and connecting A to A'' and B to B''. The potential difference V between the condenser plates being thus applied to the ends of the wire constituting the circuit branch LR, an electric current will start flowing in the circuit CLR in the direction $CBB''RLA''AC$, opposite to that of the charging current.

In other words, the excess of electrons accumulated, during the charging process, on the negative condenser plate, will start moving along the wire *LR* to reach the positive condenser plate and reestablish a neutral (discharged) state of equilibrium.

But these electrons do not, after the closing of the circuit, reach their final speed at once, the magnetic field resulting from their motion and enclosing the circuit tending to oppose speed variations, as explained in Chap. I. Under the continuous acceleration due to the potential difference between the condenser plates, however, the electron speed and therefore the current in the circuit, increase at a rate depending on the inductance of the circuit (since this governs the magnitude of the magnetic field) and on the instantaneous condenser voltage. And since the condenser is discharging, the energy of the system, which at first was entirely stored in the condenser as electrostatic energy, is being gradually transformed into magnetic energy. At the same time, a certain amount of energy is spent in the circuit resistance *R* and is thus converted into heat. When this energy transformation is complete, the condenser is fully discharged and its electrostatic energy entirely disappears.

At this moment the potential difference between the condenser plates is zero, and there being no longer any driving electromotive force in the circuit, the current flow in it will stop. The energy stored in the magnetic field about the circuit tending to prevent changes in the velocity of the electrons, however, the current will not stop instantly, but will continue to flow in the same direction for some time after the condenser is completely discharged. This will, consequently, charge the condenser again, but this time in a direction opposite to the original one, the electrons now accumulating on what was at first the positive plate of the condenser. This creates between the condenser plates a potential difference of a polarity opposite to the original one, and the magnetic energy of the system gradually transforms itself again into electrostatic energy, which is stored in the electrostatic field of the condenser, and into heat, resulting from the expenditure of a part of the energy in the circuit resistance. When the entire magnetic energy has thus been converted, the conditions are therefore similar to the original ones, except for the polarity of the condenser plates. Also, in view of the transformation of a part of the energy into heat, the amount of potential energy now stored in the condenser is less than at the beginning, and the potential difference between the condenser plates is correspondingly smaller.

The charged condenser being still connected to the circuit LR, will again start discharging, but this time in the opposite direction, and the phenomenon will repeat itself a number of times with an electric current flowing in the circuit in alternately opposite directions. The condenser discharge is therefore said to be *oscillatory*, and produces alternate transformations of electrostatic energy into magnetic energy, and of magnetic energy into electrostatic energy, each transformation being accompanied by a loss in the form of heat of a part of the energy of the system. The oscillatory discharge is therefore said to be *damped*, and the oscillation will stop when all the energy originally stored in the condenser has been dissipated as heat in the resistance.

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

It is possible that the resistance may be so high as to prevent the discharge from being oscillatory, all of the energy originally stored in the condenser as electrostatic energy being dissipated as heat. The discharge is then said to be *aperiodic*.

From the above explanation of the oscillatory condenser discharge, it may be shown that the frequency of the resulting alternating current (number of cycles per unit time) is determined by the constants of the circuit (resistance, inductance, and capacity). Neglecting the resistance R, no energy will be lost as heat during the discharge process, and the original amount of electrostatic energy, equal to $\frac{1}{2}CV^2$, will be entirely converted into an equal amount of magnetic energy $\frac{1}{2}LI^2$, designating by V the original condenser voltage, and by I the maximum current flowing in the circuit when the condenser is fully discharged. Thus,

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

Hence, the smaller the inductance L, the greater the discharge current I. And since, by definition, the current is equal to the quantity of electricity passing one section of the circuit during a unit of time, the greater the current, the more rapidly will the charge or quantity of electricity stored in the condenser be set into motion in the circuit. The above expression thus shows that, for a given condenser capacity C, the time required to once discharge the condenser, and hence also the time required to go through one cycle of the oscillation, that is, the *period* of the alternating-current oscillation, will vary directly as the square root of the inductance. The *frequency*, therefore, varies inversely as the square root of the inductance. In a similar way, for a given inductance L, the period varies as the square root of the condenser capacity, and the frequency therefore varies inversely as this.

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 this frequency in terms of the circuit constants is given quantitatively in a later paragraph.

Summarizing the above discussion, an alternating current may be generated in a circuit containing inductance and capacity by first charging the condenser and then letting it discharge through the inductance. The amplitude of this alternating current continually decreases, due to the resistance of the circuit. The frequency of the current is determined by the inductance and capacity of the circuit, and hence may be made substantially as great or as small as desired. Indeed, it has in this manner been possible to generate oscillations of frequencies of many million cycles per second.

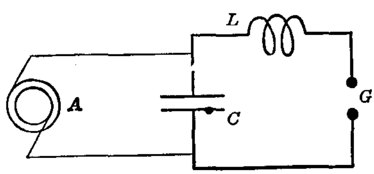


Fig. 29.

However, on account of the damping of the oscillations, the actual duration of a train of oscillations is, in general, not more than a fraction of a second. In order to have a succession of oscillation trains, it is then necessary to recharge the condenser after each oscillation has died out. This may be done by inserting a spark gap in the oscillatory circuit and charging the condenser by means of a source of alternating or pulsating voltage.

This method is illustrated in Fig. 29. The oscillatory circuit comprises the condenser C, inductance coil L, and gap G. The plates of the condenser are connected to the two terminals of an alternator A generating a sine wave electromotive force. If the gap G is so wide that it will 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 electromotive force, and may be represented by a sine curve as in Fig. 30, the potential difference varying between the maximum values $+V$ and $-V$.

If the gap G, which like the condenser C receives the full alternator voltage, is set so that a spark will jump across it when a voltage V', smaller than V, is impressed upon it, then the conditions become very much different. Thus when the

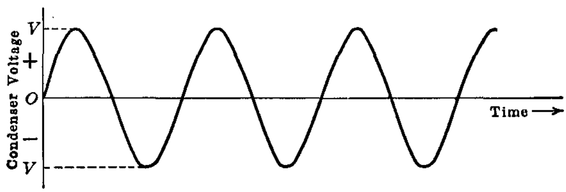


FIG. 30.

alternator voltage rises from zero to the value V' (Fig. 31) 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 is set, the latter breaks down, a spark bridging the gap, ionizing the gas between the electrodes, and filling the space between them with conducting metal vapors. This suddenly closes the oscillator circuit

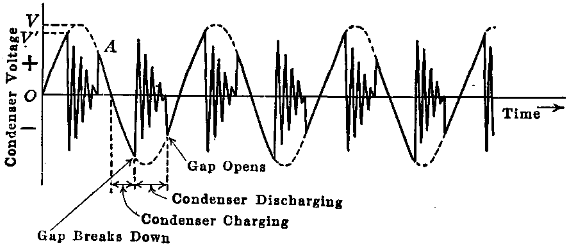


FIG. 31.

LGC, and the condenser C discharges through the inductance L. If the resistance of the wires making up this circuit, and that of the spark across the gap G are not too great, the condenser discharge will be oscillatory, as explained before. This creates in the circuit LGC a damped alternating current of a frequency determined by the values of L and C, and produces across the condenser C an alternating electromotive force of this same frequency, as shown in Fig. 31. When the entire amount of potential energy originally stored in the condenser by the alternator has been dissipated, the oscillation ceases, 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 process repeats itself.

A series of successive oscillation groups may thus be obtained in the oscillatory circuit LC. The frequency of the oscillations in each group, being determined by the constants of the oscillatory circuit, is entirely independent of the alternator frequency. The frequency of the oscillation goups, however, is in the above explanation equal to twice the alternator frequency, there being one oscillation group for each half-cycle. Further, if the voltage A (Fig. 31) impressed on the condenser after the discharge is sufficient to again break down the spark gap, a new discharge will take place before the end of the half-cycle of the alternator. The frequency of the oscillation groups thus depends on the alternator frequency as well as on the gap adjustment. In general, this adjustment is such as to give but one discharge for every half-cycle of the alternator.

The periodical succession of oscillatory discharges, requiring the periodical charging of the condenser to a certain potential, may also be accomplished by using an induction coil (spark coil) instead of the alternator, as is done frequently in low-power radio sets. The electromotive force charging the condenser is then pulsating, instead of being alternating, which does not materially change the above explanation.

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

Mathematical Interpretation of the Oscillatory Condenser Discharge.—As explained in the preceding section, the amount of electrostatic energy variation in the oscillating system of Fig. 28 during a certain interval of time is equal to the sum of the energy lost as heat and the magnetic energy variation during the same time interval. The algebraic sum of these three quantities is, therefore, equal to zero.

In the time interval dt the variation of electrostatic energy is

$$d(\tfrac{1}{2}Cv^2) = Cvdv,$$

while the variation of magnetic energy is

$$d(\tfrac{1}{2}Li^2) = Lidi,$$

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 intensity in the circuit at the instant considered, v the condenser voltage at the same instant, and di and dv the current and voltage variations during the period of time dt. Since, by definition, the current i is equal to the quantity of electricity dq set in motion in the time dt,

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

and since

$$q = Cv,$$

it follows that

$$dq = Cdv$$

and

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

so that finally the above equation becomes

$$\frac{d^2v}{dt^2} + \frac{R}{L}\frac{dv}{dt} + \frac{v}{LC} = 0,$$

or

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

There are, hence, 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

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

or

$$R^2 > 4\frac{L}{C}$$

then v and i are not periodic functions of time, and the discharge is *aperiodic*. The current and voltage in the circuit may then be represented by curves similar to those of Fig. 32.

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

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

the current i and voltage v are periodic and exponential functions of time having a cyclic frequency

$$\omega_n = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}},$$

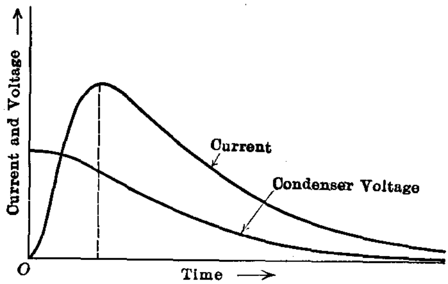


FIG. 32.

and a decrement

$$\delta = \pi\frac{R}{\omega L} = \pi R\sqrt{\frac{C}{L}}.$$

The current and voltage in the circuit may be represented, in function of time, by curves similar to those of Fig. 33.

The decrement δ is merely a representation of the damping action due to the circuit resistance, which continually decreases the amplitude of the alternating-current oscillation. The maximum amplitude of the voltage and current decreasing logarithmically, that is, by a constant percentage during every cycle, the value δ given above for the decrement is the natural logarithm of the ratio of the maximum amplitudes of two successive cycles, this being easier to deal with, in the calculations, than the actual ratio. The decrement is zero when there is no resistance in the circuit ($R = 0$).

The damping action may also be represented by the ratio of the average energy lost in the circuit resistance to the average magnetic or electrostatic energy, the logarithmic decrement δ given above being equal to one-half this energy ratio.

Attention is here drawn to the fact that the natural frequency ω_n of the free oscillation is different from the resonance frequency $\omega_0 = 1/\sqrt{LC}$ of the circuit, defined before as the frequency for which the circuit is in tune or in resonance with an external alternating-current energizing source. This natural frequency ω_n of an oscillatory circuit becomes equal to its resonance frequency ω_0 when its resistance R is zero. In this connection it is pointed out that the free oscillation of a circuit having no resistance, being undamped, is comparable to the undamped forced oscillation sustained, in a circuit having resistance, by an external alternating-

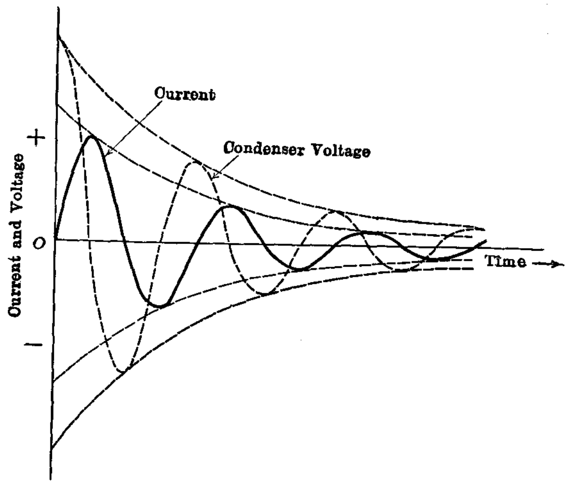


FIG. 33.

current energizing source of frequency equal to the resonance frequency of the circuit. The alternating-current source in this case furnishes, as shown before, a current $I = E/R$ through which it synchronously replenishes the energy lost in the circuit resistance. In other words, the alternator neutralizes the damping action of the resistance.

Most radio circuits are designed with so small a resistance that their natural and resonance frequencies may generally be considered as substantially equal to each other.

Free Oscillations of Loosely Coupled Circuits.—Consider an oscillatory circuit A (Fig. 34) containing fixed inductance and capacity and a spark gap, and in which free oscillations may be set up as explained in the preceding section. Consider also a second oscillatory circuit B *loosely* coupled to circuit A (that is, as explained before, so coupled that the current in circuit A inducing a current in circuit B, this latter current will not induce an appreciable current in circuit A). Assume circuit B to have a fixed inductance and adjustable capacity constituted, for instance, by a variable condenser.

Circuit A oscillating at its natural frequency, determined by its inductance, resistance, and capacity, as explained before, the damped alternating current flowing in it at each spark across the gap, induces in circuit B an alternating electromotive force of decreasing amplitude. Conditions in circuit B are thus the same as though an alternator were inserted in series with it, generating an electromotive force equal at every instant to the electromotive force actually induced by circuit A in circuit B. This was studied previously, except that oscillation in circuit A being damped, the electromotive force acting in circuit B is of constantly decreasing amplitude.

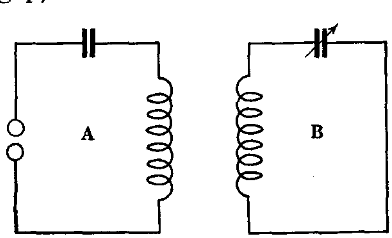


FIG. 34.

Reasoning then for circuit B, as in the case of alternator-excited series resonance, it follows that the alternating current set up in circuit B is a maximum when its variable condenser is adjusted to tune this circuit to resonance with circuit A. And if an ammeter be inserted in circuit B and the condenser in this circuit be varied, the resonance curve of circuit B representing the current intensity in this circuit in function of its natural frequency, will pass through a single maximum when circuit B has substantially the same natural frequency as circuit A.

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 continue after 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, the oscillations in the two circuits will have substantially the same characteristics as the oscillations of circuit A.

Free Oscillations of Closely Coupled Circuits.—If the two circuits A and B are *closely* coupled, the current set up in circuit

B by the alternating current flowing in circuit A induces in the latter an electromotive force which, combining with the electromotive force operating in circuit A, modifies the current flowing in this circuit. As in the case of coupled circuits energized by an alternator, the two circuits constitute a single-circuit system having two oscillation frequencies, even though the two circuits A and B may have, individually, the same natural frequency.

And since at each of these two resonance frequencies the phase and amplitude relations permit a stable operating equilibrium, the system of two closely coupled oscillatory circuits will simultaneously oscillate freely at these two frequencies. Thus, the circuits A and B will each carry two damped alternating currents of more or less different frequencies, which, combining in each circuit according to their respective instantaneous intensities, periodically add and subtract, producing beats similar to those resulting from the combination of two sounds of slightly different pitches. These beats, furthermore, occur in the two circuits in such a manner that the resulting oscillation current in one of the circuits is a maximum when the current in the other is a minimum, and conversely.

These conditions may be clearly interpreted, as shown below, in considering the energy producing the oscillation as being alternately stored in the one and in the other circuit A and B, the surging back and forth of the energy from one circuit to the other taking place at a frequency equal to the difference of the two oscillation frequencies of the circuit system AB.

Consider, then, the two oscillatory circuits A and B as before, closely coupled, and having, individually, the same natural frequency, this condition being chosen for greater simplicity of explanation. An oscillation is set up in circuit A by first charging its condenser until the spark gap breaks down, when the condenser discharges. This oscillating discharge, represented by the upper curve of Fig. 35, is damped due to a continual transfer of energy from circuit A to circuit B, and a partial loss of energy in the circuit resistance. This corresponds to a gain of energy on the part of circuit B, in which the induced alternating current increases in amplitude, as shown by the lower curve of Fig. 35.

At a certain time t_1 all the energy originally stored in circuit A will have been transferred to circuit B or dissipated in the resistance. Circuit B, at the same instant, is in full oscillation, having continually received energy from circuit A up to this time t_1. This supply of energy having now dissappeared, the oscillation in circuit B will start decreasing in its turn, on account of the energy losses in the resistance of this circuit. However, due to the extremely short time required for the entire process, the conducting vapors in the spark gap of circuit A have not had time to

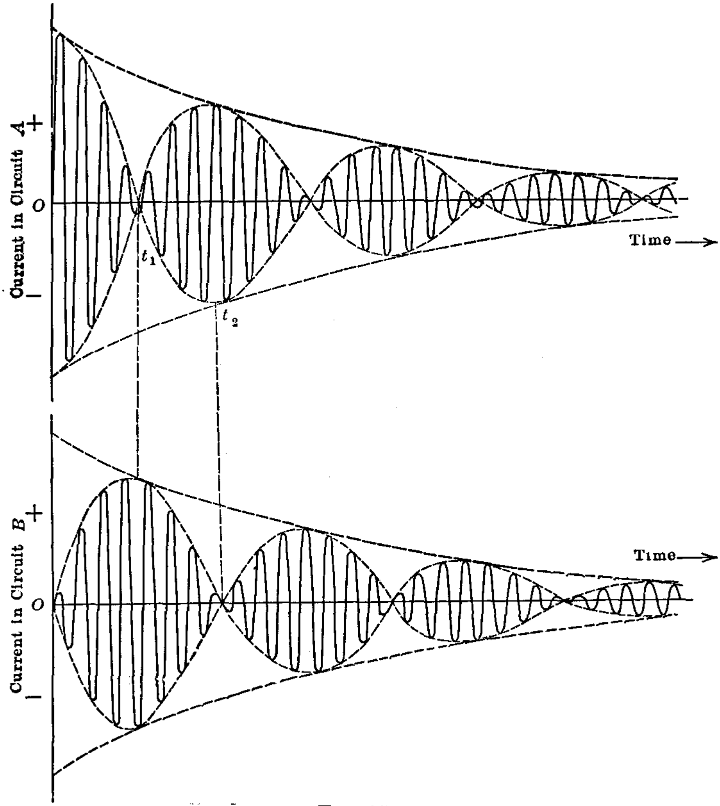


FIG. 35.

escape, cool down, or deionize, so that the gap is still in a conducting state, making A virtually a closed circuit. The alternating current in circuit B therefore reacts upon circuit A and induces an electromotive force which again sets up an alternating current in circuit A. There will thus be a transfer of energy from circuit B to circuit A during the oscillation of circuit B. This increases the decrement of circuit B, in which the oscillation, at a time t_2, will have entirely died out. At this same instant, circuit A is again oscillating, with an amplitude less than that of its original oscillations by an amount equal to the losses in the resistances of the two circuits. 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 system.

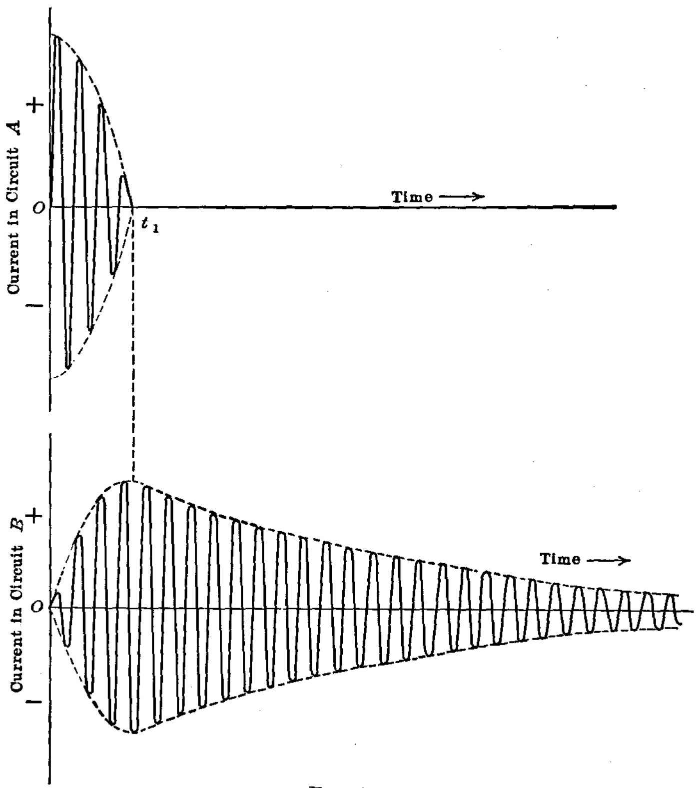


Fig. 36.

It is nevertheless possible, with the circuits closely coupled, to avoid this alternate transfer of energy between the circuits, and thus set up in circuit B a single-frequency alternating current. 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 energy back from circuit B. The resulting conditions are then as represented in Fig. 36. 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 the resistance. At the time t_1 the gap in circuit A is made non-conducting by one of the methods indicated later. Circuit B being then in full oscillation, and setting 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, as if circuit A had been entirely removed.

With this method of setting up or exciting oscillations in the secondary circuit B, the primary circuit generally oscillates for a short period of time only as compared with the oscillation in circuit B. 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. By analogy, this method of exciting the secondary circuit is called "impulse excitation."

With this method, the secondary becomes the main oscillatory circuit. It does not include any spark gap, and is generally a circuit of low resistance and small decrement, in which the oscillations are less damped than in the single oscillatory circuit having a spark gap. The spark gap, which is located in the primary circuit, is of a special "quenched" type, operating as explained presently.

This rapid quenching of the spark gap, which insures the open. ing of the primary circuit and avoids secondary reaction, may be obtained in a variety of manners which all accomplish a rapid deionization and cooling of the air between the gap electrodes. Thus, one type utilizes a blower which continually changes the air within the gap. Another type comprises a multiplicity of small gaps connected in series, the electrodes being provided with metal flanges, which cool down the air between them by heat conduction and radiation. In a third model, known as the rotary gap and used principally in damped wave radio sets in which the primary circuit is periodically charged by an alternator, one of the electrodes is a toothed metal disc or wheel rotating close to a fixed metal stud forming the other electrode. This gap is generally so arranged that one tooth of the rotary electrode will pass in front of the stationary electrode every time the alternator voltage becomes a maximum. The gap then breaks down, but the spark is rapidly extinguished, due to the fact that the rotation of the disc removes the tooth from the stationary electrode.

Wavemeters.—An application of resonance phenomena is embodied in the "wavemeter." This is a device which may be used for measuring the frequency of an oscillating or alternating current, the natural frequency of a circuit, the electrostatic capacity of a condenser or inductance of a coil, the decrement of a circuit or an oscillation, and the radio wave length, which is a

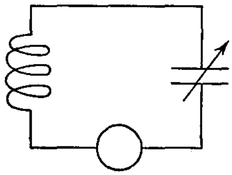


FIG. 37.

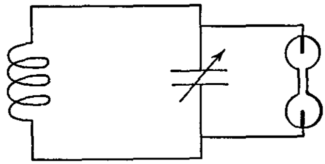


FIG. 38.

quantity defined in the following chapter. It may also be used for generating oscillations of known frequency.

The wavemeter, in its simplest form, is a closed, low-resistance oscillatory circuit, the natural frequency of which may be altered by means of a variable condenser or inductance coil, and which is calibrated so that its natural frequency or its constants are known for each setting of the variable condenser or inductance coil. When used as a measuring instrument the wavemeter circuit also comprises a sensitive current or voltage measuring or indicating device. Thus, a hot-wire ammeter (Fig. 37) may be connected in series with the inductance coil and condenser, or a thermocouple connected to a sensitive galvanometer, or else a neon or Geissler tube may be connected in shunt with the condenser (Fig. 38).

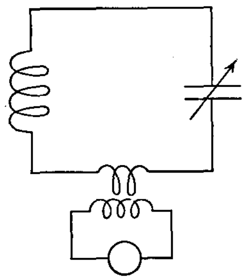


FIG. 39.

Current measuring devices, such as ammeters or thermocouples, are frequently of relatively high resistance. In order to keep down the resistance of the wavemeter oscillatory circuit and thereby increase the sensitivity of the device by reducing the energy losses in it, the current measuring instrument may be loosely coupled to the circuit (Fig. 39) instead of being connected directly in it.

If the wavemeter is used simply as an indicating instrument, the sensitive device may be a small low-resistance incandescent bulb, which glows when the current in the circuit is maximum; or else a telephone receiver may be used in conjunction with a rectifying detector, when working with damped, or modulated, or discontinuous oscillations, as explained in a later chapter.

The use of the wavemeter for measuring the frequency of an oscillation has been explained: it is merely necessary to couple the wavemeter loosely to the circuit under test, and vary the wavemeter constants until maximum current is observed by means of the indicating device. The wavemeter is then in tune with the frequency of the oscillation, and this frequency may then be obtained from the wavemeter setting and calibration. The "*wave length*" of the oscillation, which, as explained in the following chapter, is equal to the ratio of the speed of light (300,000 km. per second) to the frequency of the oscillation, is then obtained at the same time, and many wavemeters are calibrated directly in wave lengths.

The wavemeter may be used for measuring the decrement of an oscillation, and if this oscillation is generated by a circuit oscillating freely, such as is the case when it is excited by the impulse method described 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 of the ammeter being recorded, and also the corresponding electrostatic capacity C of the wavemeter condenser. The condenser setting is then varied until a current equal to $I/2$ is observed, when the electrostatic capacity C' of the wavemeter condenser is recorded. The decrement D of the oscillation is then given by the relation

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

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 specially constructed and calibrated to read decrement directly (decremeters). The description and theory of an instrument of this type is given in Bureau of Standards *Scientific Paper* No. 235.

The above measurements may be done in a similar manner with a wavemeter having a variable inductance instead of variable electrostatic capacity. This, however, may not be 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 of coupling, as well as to the change in the natural frequency of the wavemeter, thus making the result less reliable.

The electrostatic capacity of a condenser may be determined by means of a wavemeter by connecting the condenser to the 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 its free oscillations is measured with the wavemeter, as explained above. The unknown capacity 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 capacity and proceeding along similar lines.

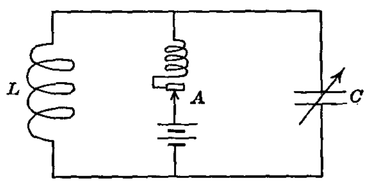


FIG. 40.

Finally, the wavemeter may be used to generate oscillations of standard or known frequency. To this effect the wavemeter condenser is shunted by a buzzer and battery (Fig. 40). 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 connected across the wavemeter condenser through the buzzer winding and thus charges the condenser. But the buzzer circuit being also completed through the inductance L, the vibrator is attracted toward the buzzer coil, and the contact A is opened. The wavemeter condenser C then discharges through the coil L at the natural frequency of the wavemeter circuit. As the buzzer vibrator springs back to its original position, the contact A closes, the condenser recharges, and the process repeats itself.

The three-electrode vacuum tube has greatly improved the above methods, in increasing the sensitivity of the device, in permitting the generation of continuous oscillations of known frequency in a calibrated oscillatory circuit, and also in adapting the heterodyne methods to the accurate measurement of frequency. The principles underlying the operation of such devices are fully developed in later chapters of this book.

CHAPTER III

ANTENNA SYSTEMS AND RADIATION

Closed and Open Oscillators.—The mechanism of energy radiation was explained in Chap. I, where it was shown that electromagnetic waves are sent out by an accelerated electron, and hence also by a circuit carrying an alternating current, and that the radiation fields (electric and magnetic) are directly proportional to the electron acceleration, and, therefore, to the frequency of the alternating current. The freely oscillating discharge of a condenser, permitting to attain extremely high frequencies, should therefore provide means for radiating large amounts of energy. All oscillatory circuits are not suitable for this purpose, however.

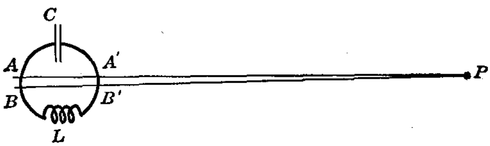


FIG. 41.

Thus, for instance, consider the oscillatory circuit of Fig. 41 comprising a condenser C connected to the ends of an inductance coil L. When a high-frequency, alternating-current oscillation is established in this circuit, in the manner explained in the preceding chapter, interlinked alternating electric and magnetic fields are established around it, which set up alternating electromotive forces and currents in any external circuits which may be present and in a suitable position. Now consider a point P at a comparatively great distance from the oscillatory circuit. When alternating current is flowing in the circuit LC, the current in the section AB of that circuit alternately flows upward and downward, creating an alternating field at point P. But the section $A'B'$ of the circuit carries at every instant a current equal to that in the section AB but in opposite direction, creating at point P a field neutralizing that of the section AB. There will hence be no appreciable radiation from the circuit, except when the point P is so near to the circuit that its distance to the various sections of the circuit can no longer be considered as equal.

Such "closed" oscillators are therefore not generally suitable for long-distance radiation of energy, and so-called "open" oscillatory circuits are used for this purpose, the principle of which is illustrated by Fig. 42. Consider a straight metal wire $A'B'$ cut in its center by a small sphere gap AB. This system constitutes a condenser (from the very definition of a condenser), and electrostatic energy may be stored in it. Thus, connecting A and B to a source of continuous potential charges this condenser by removing some free electrons from one wire and placing a like number on the other. In the case of Fig. 42, the upper wire

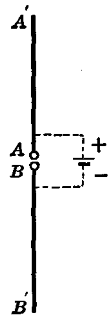


FIG. 42.

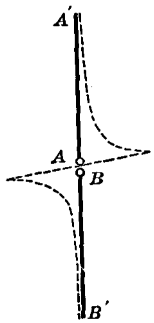


FIG. 43.

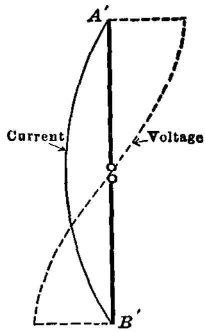


FIG. 44.

AA' will be positive and lose some electrons, while the lower wire BB' gains a like number of electrons and becomes negatively charged. This establishes a potential gradient along the system $A'B'$, as shown in Fig. 43, where the deficiency or excess of electrons is plotted at each point of the wire $A'B'$ perpendicularly to it. An electrostatic field is thus created about the wire, representing a store of potential or electrostatic energy.

If the potential difference between the two sections AA' and BB' is large enough, a spark will bridge the gap AB and the electrons accumulated on the lower section BB' start moving upward under the action of the electrostatic field into the upper section AA', tending to reëstablish the original neutral electrostatic equilibrium. This produces an electric current flow in the wire $A'B'$, with a corresponding continual decrease of electrostatic energy and the setting up of a magnetic field and store of magnetic energy, as explained before. When the entire amount of electrostatic energy has thus been transformed into magnetic energy (with a partial loss due to the resistance of the circuit), the potential difference across AB is zero, but, due to the presence of the magnetic field and exactly as in the case of an ordinary oscillatory circuit, previously considered, the current continues to flow in the direction $B'A'$ until the entire amount of magnetic energy has, in its turn, disappeared. This continued electric current flow accumulates a number of electrons on AA' and removes a like number from BB', charging the system with a polarity opposite to the original one. This first discharge and recharging is followed by a discharge in the opposite direction, and so on, until the entire original store of energy has been dissipated. Thus, the system $A'ABB'$ having inductance and electrostatic capacity, an oscillatory current may be set up in it by the sudden discharge, exactly as in the case of the oscillatory circuits considered in the previous chapter.

The process may be interpreted as consisting in the surging back and forth along the wire of a charge of electricity resulting

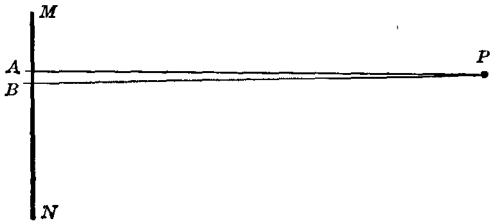


FIG. 45.

from the electrostatic unbalance originally created. That is, a mass or agglomeration of electrons will oscillate back and forth along the wire between its ends A' and B', these electrons accumulating at the end of each half-cycle of the oscillation, at one end or the other of the wire $A'B'$, and moving with maximum speed past the center AB of the wire at the middle of each half-cycle, produce along the wire $A'B'$ a voltage and current distribution as shown in Fig. 44 (maximum voltage at the ends, maximum current intensity at the center of the wire $A'B'$). During each half-cycle the motion of the charge results in the radiation of a certain amount of energy in the form of a wave, in the manner explained in Chap. I.

Contrary to the case of Fig. 41, the oscillations of the straight wire or open oscillator MN of Fig. 45 produce in this a current flowing alternately up and downward, and which is at every instant 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 is then no longer in opposition with the field, due to some other section of the circuit, and energy radiation thus takes place effectively.

Wave Length. Wave Front.—The electric charge surging back and forth along the wire MN, as just explained, may in every respect be likened to the simple electric charge m of Fig. 12. From this it follows that energy is radiated with greatest intensity along a plane perpendicular to MN and cutting MN in its middle, and not at all in the direction of the wire MN. Such radiation results in the setting up, at distant points P (Fig. 46), of alternately and simultaneously reversed magnetic and electric fields H and F, directed perpendicularly to each

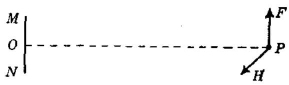


FIG. 46.

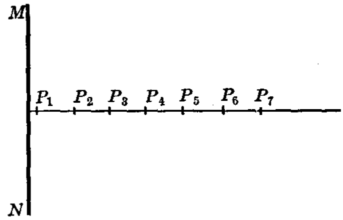


FIG. 47.

other and to the line PO joining point P to the center O of the radiating circuit MN, the electric force F being contained in the plane of OP and MN, while the magnetic force H is normal to this plane.

Now, when an alternating current flows in the conductor MN, the radiation fields around the wire reverse periodically. These reversals, however, do not occur simultaneously at all points of space, but travel from the wire MN outward in all directions with a definite speed equal to the speed of propagation of light. Thus, in Fig. 47, in the first half-cycle when the alternating current rises from a minimum to a maximum, the fields at a point P_1 in the immediate vicinity of the wire, rise similarly and simultaneously. At point P_2 the fields will also rise, but will start doing so a small fraction of time after the corresponding field change at point P_1, the time of lag being that required for the disturbance or wave to travel from P_1 to P_2. Similarly, the fields at points P_3, P_4, P_5, etc., will increase, but will start increasing only when the wave traveling with the speed of light has reached them in turn.

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 decrease 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 now travels outward from MN at the same speed as the preceding wave of increasing field, and reaches all points in succession. *At each point of space* the field reversals thus follow the current reversals in the wire, at a frequency equal to that of the alternating current in the wire. But, due to the finite speed of propagation of electrical disturbances, these field reversals do not take place simultaneously at all points. With an alternating current flowing in the wire MN, there will be a succession of waves of alternately increasing and decreasing fields traveling from MN outward.

It may then 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 P_3, which would be equivalent to a synchronism of the field variations at these 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 obviously equal to the velocity of propagation of the disturbance, divided by the frequency of the alternating current in the circuit. Thus,

$$\text{wave length} = \frac{\text{velocity of light}}{\text{frequency}} = \frac{300{,}000{,}000}{f} = \lambda_m$$

where λ_m is the wave length in meters. It follows 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 50 or even 10 to 5 m. Wave lengths as short as a few millimeters have been obtained in the laboratory by means of specially constructed oscillatory circuits.

Another expression of wave length may be derived from the formula

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

whence

$$\lambda = 2\pi V\sqrt{LC},$$

where λ is the wave length, V the velocity of light, and L and C the inductance and capacity 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.

It also follows from the above explanation that synchronous field variations occur at all points of a sphere having the radiating circuit MN at its center.[1] That is, the fields at all these points are a maximum, minimum, or zero together. The fields at points located on a concentric sphere of slightly greater diameter are also synchronous with each other, but lag slightly in phase behind the fields on the smaller sphere. That is, they pass respectively through their maximum, minimum, or zero values slightly later, when the sphere of maximum, minimum, or zero field intensity has expanded to include these points. This expanding spherical surface, tangent to the lines of force of synchronously varying magnetic and electric radiation fields, is called the *front* of the advancing wave.

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

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

Second, the circuit loses energy by *radiation*, this energy being removed from the circuit and stored in the emitted electromagnetic waves, following a mechanism explained in Chap. I. It was then shown that 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 antenna-current intensity, and is therefore of the form I^2A. By similarity with the ohmic resistance R, which expresses the energy lost in heat, the factor A,

[1] This is true in empty space only, as otherwise the surface is not a sphere when, as is actually the case in long-distance radio transmission around the globe, the propagation speed of the waves is not the same in all directions. This is studied in a later section of this chapter.

representing the energy lost or radiated into space, is called the *radiation resistance* of the antenna or radiating circuit. The greater this radiation resistance, the greater the amount of energy radiated by the circuit.

Third, energy is lost due to the fact that the dielectric surrounding the radiating circuit is not perfect, but partially conducting. Thus, on account of the "open" structure of the circuit and its generally large dimensions, energy is absorbed by the conducting materials in the surrounding objects, such as supporting masts or towers, trees, houses, metal structures, etc. This *dielectric*

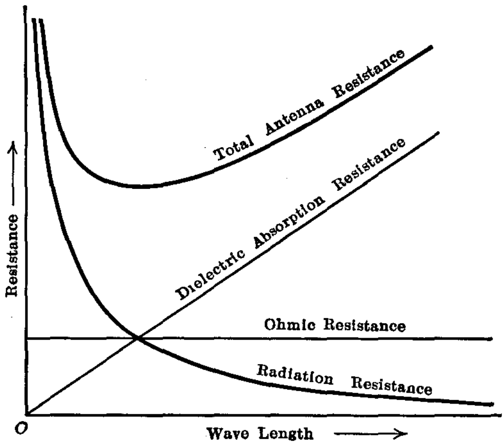


FIG. 48.

absorption loss is approximately proportional to the wave length.

The *effective resistance* of the antenna or radiating circuit is the sum of the three components indicated above. It therefore varies with the operating frequency, as represented in Fig. 48, which also shows that this resistance passes through a minimum value for a given value of wave length, at which the sum of the three components is a minimum.

Antenna Constants and Types.—As stated before, the natural wave length of a freely oscillating circuit is determined by the circuit constants (inductance and capacity). For open oscillators, these constants are more or less uniformly distributed along the entire circuit. Thus, in the oscillator of Fig. 41, the capacity comprises the small capacity of the extreme wire ends A' and B' with respect to each other as well as the larger capacity of the neighboring central wire portions.

Theoretical investigation, as well as wavemeter measurements, show that the free oscillation of a circuit, as shown in Fig. 41, has a wave length substantially equal to twice the overall length $A'B'$ of the wire. The practical conditions of installation of such an antenna system therefore restrict its use to very small wave lengths only, of the order of only a few meters.

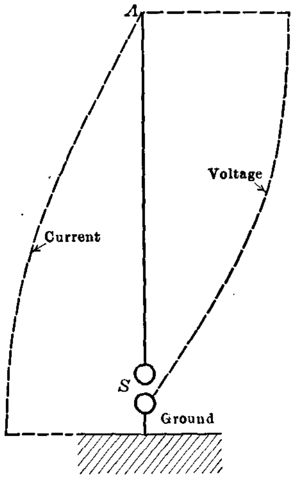


FIG. 49.

For the longer wave lengths, only one of the two conductors AA' and BB' is used, the other being replaced by the ground on which the antenna is erected, or a so-called counterpoise consisting of a network of wires buried in or laid on the ground. A simple antenna circuit then takes the form of a single wire suspended vertically and grounded at its lower end (Fig. 49). This being, as just explained, essentially one-half of the antenna of Fig. 41, it oscillates at one-quarter wave length, meaning that its natural wave length is roughly equal to four times the height of the vertical wire.

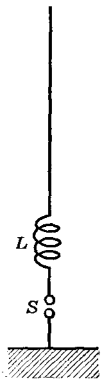


FIG. 50.

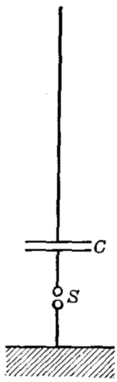


FIG. 51.

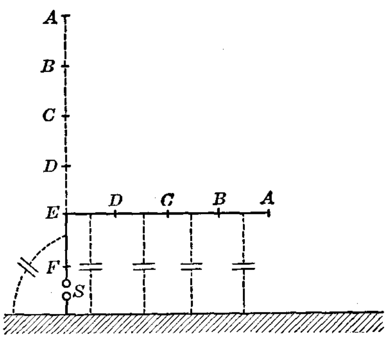


FIG. 52.

In order to change the wave length of a vertical wire antenna, it is then necessary to vary the length of the wire. Or an inductance coil L (Fig. 50), may be inserted in series with the antenna. This increases the total inductance of the circuit, and hence also the wave length, as shown by the wave length formula given on page 62. In order to shorten the wave length, a condenser C (Fig. 51) may be inserted in the antenna, the total capacity of the oscillatory circuit being thereby reduced. This is due to the fact that this condenser is in *series* with the capacity of the antenna system.

Another method consists in changing the distributed constants of the antenna by changing its shape. Thus, in Fig. 52, by suspending horizontally a length AE of the originally vertical antenna wire, a so-called *inverted-L* antenna is obtained. This will have only slightly greater inductance than the vertical antenna of Fig. 49, but the wire above point E being brought closer to the ground, the capacities of the various elementary condensers formed by the wire sections AB, BC, CD, DE and the ground will be increased, the dielectric between these sections and the ground being reduced, which correspondingly increases 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.

Still other arrangements may be used by using a plurality of horizontal branches instead of the single branch EA of Fig. 52. Antennae, such as the well known "T" and "umbrella" types, are then obtained. These have a comparatively small inductance and great capacity. It is generally difficult to calculate accurately the constants of these more complicated types of antennae. Approximate formulæ may be found in Bureau of Standards *Circular* No. 74, and other publications.

Receiving Antennae.—Reasons similar to those which made the use of "open" oscillators 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. 41 above, the waves emitted by a distant oscillatory circuit P would set up a current generally too small to be observed, because the electromotive force 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 oscillatory circuit (Fig. 42) this objection is avoided, and reception takes place effectively. Also, in view of the generally great distance separating the transmitting and receiving circuits and the consequent weakness of the fields at the receiving station, the receiving antenna should be linked with as large a portion of these fields as possible, which naturally leads to the use of an open oscillator.

Directional Characteristics.—Consider a vertical wire antenna, shown in plan view by the point A (Fig. 53) and assume that a number of observers equipped with receiving circuits, all identical, are scattered about the antenna. Assume also that each of these receiving circuits has some device permitting one to measure the current set up in it when the antenna A is oscillating. Now, if the observers move toward or away from the antenna A until they all obtain a same current reading in their receiving apparatus, they will find themselves on a circle having A as its center, showing 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, as shown in Fig. 54, where A and B are respectively the grounded and free ends of the antenna. This shows that this type of antenna radiates energy with greater intensity in the direction of its grounded end.

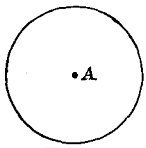


Fig. 53.

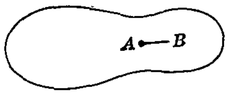


Fig. 54.

More generally, the energy emitted by an antenna may or may not be radiated with equal strength in all directions. Similarly, a receiving antenna may or may not receive signals equally well from all directions, depending upon the shape and arrangement of the antenna. These directional properties are important when communication is to be established between two definite stations, or on the contrary between a plurality of stations scattered in different directions. In order to study these properties it is necessary, however, to distinguish between long and medium waves on the one hand, and short waves on the other, the mode of propagation being different in some respects.

Long and Medium Waves.—It was pointed out before that the energy emitted by a radiating circuit travels away from it in all directions in the form of a spherical wave having the radiating circuit at its center. This is true provided such propagation takes place in a perfect dielectric and with the same speed in all directions.

In radio communication, however, the waves propagate in the earth's atmosphere. This consists essentially of a layer of air (which is a good dielectric) surrounding the globe, bounded on the one side by the earth's semi-conducting surface and on the other by rarefied gases, which are strongly ionized and rendered conducting by the solar and other cosmic radiation.

The waves set up by the grounded oscillator A (Fig. 55), thus enclosed in the layer of insulating air, hence propagate away from the oscillator in all directions. The earth's surface, however, being a poor dielectric and a poor conductor, some energy is absorbed at the foot of the wave, resulting in a dephasing of the magnetic and electric field components and an elliptic polarization of the wave, not studied here. Also, the *propagation speed is altered* (lowered) so that the wave advances more slowly near the ground than higher up in the air. The wave front thus becomes tilted forward at the foot of the waves.[1]

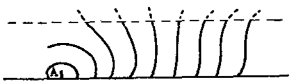


FIG. 55.

On the other hand, in the upper regions of the atmosphere where the air is rarefied and ionized, the propagation speed is *greater* than in the denser portions below. The wave advancing more rapidly, the wave front here also becomes bent forward so that, generally speaking, the entire wave front is inclined toward the ground. And since the direction of propagation of the radiation fields, which are at every point tangent to the wave front surface, is perpendicular to the wave front, this tilting forward of the latter explains why the waves emitted by the antenna follow the curvature of the earth, permitting communication to be established between distant points on the globe.

It also explains the directional properties of geometrically asymmetrical antennae, such as the inverted-L antenna GAB of Fig. 56, for example.[2] Suppose signals to be coming from a distant station located at the left of the figure, the incoming waves traveling from left to right. The electric field set up at the receiving station, instead of being vertical, is then inclined in the direction OF, as just explained. It may, therefore, be considered as the combination of a vertical electric field F and a horizontal component F, respectively, setting up electromotive forces in the

[1] SMITH ROSE, DR. R. L., and R. H. BARFIELD, "Some Measurements on Wireless Wave Fronts," *Experimental Wireless and Wireless Engineer*, pp. 737–749, September, 1925.

KRUSE, R. S., "Horizontal Reception," *Q. S. T.*, pp. 9–17, February, 1926.

[2] Zenneck, *Annalen der Physik*, 1907, vol. 23, p. 846.

vertical and horizontal antenna portions GA and AB as shown by the arrows, which operate in the same direction in the antenna circuit. If, on the other hand, signals come from the opposite direction, that is, from the ungrounded side of the antenna, the waves travel from right to left of the figure, the field is tilted as shown by OF (Fig. 57). The electromotive forces set up in the two branches GA and AB then oppose each other, totally or partially neutralizing each other and hence reducing the intensity of the recorded signals.

Referring again to Fig. 56, and recalling previous discussion: When an alternating current is flowing in the distant transmitting antenna, the electric field OF at the receiving station alternately reverses, acting successively in the directions OF and FO. In this manner an alternating electromotive force is set up along the antenna circuit GAB in which, as explained before, a mass or

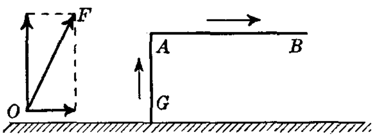


FIG. 56.

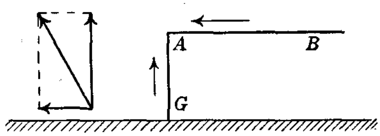


FIG. 57.

charge of electricity is set into oscillation, traveling back and forth along the antenna wire and accumulating at each half-cycle end in the ground or at the free antenna end B. This effect, of course, is a maximum when the free oscillation frequency of the circuit is equal to the frequency of the incoming oscillations.

Stated in other terms, an electrical disturbance travels back and forth along the antenna, being reflected back every time it reaches its free end B.

These oscillations build up gradually during the successive cycles of the exciting field, that is, as the successive incoming waves sweep past the antenna, each yielding some of its energy in giving a new impulse to the oscillating charge in the circuit. And these oscillations continue for some little time after the end of the signal, the circuit continuing to oscillate freely until all of the energy gradually accumulated in it is fully dissipated.

Such a system therefore permits *unidirectional reception* due to the differential action, in the circuit, of the horizontal and vertical field components, and the *gradual building up of a large signal current,* through the cumulative action of the successive incoming waves acting in synchronism with the natural oscillations of the system, in other words through *resonance.*

These same characteristics of unidirectional reception and large received current may be obtained in a different manner by means of a so-called *wave antenna,*[1] to be described presently, in which the accumulation of energy collected from the wave is not obtained through resonance (brought forth through the cumulative action of successive waves, and reflection of the impulses at either end of the antenna circuit), but through the continuous action of a single wave during a comparatively large portion of its travel.

A wave antenna consists essentially of a long, horizontal wire AB (Fig. 58) pointing toward the station transmitting the signals which are to be received. Contrary to the inverted-L antenna described before, the wave antenna is connected to ground

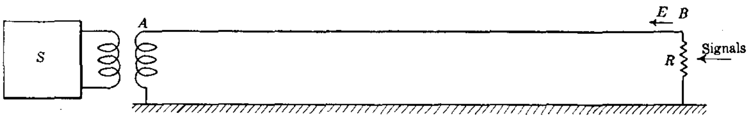


FIG. 58.

through the radio receiving set S at the end *farthest away* from the transmitting station. When a wave (traveling from right to left in the figure) reaches that end B of the antenna nearest to the transmitting station, the horizontal component of its slightly tilted electric field sets up an electromotive force E in the end portion B of the antenna wire. If the wave should then suddenly stop or vanish, this small electromotive force E would propagate or travel from the end B down the wire toward the end A, and the disturbance would thence pass into the ground through the receiving circuit. Actually, the wave continues traveling along the wire with the disturbance, inducing a small electromotive force in each successive portion of the horizontal wire as it is reached. All these elementary electromotive forces, like the initial electromotive force, travel along the wire together, coinciding, adding continually, and building up to many times the initial value, until they finally reach the end A and actuate the receiving set.

It would thus seem that the signal intensity would increase indefinitely with the length of the horizontal antenna wire.

[1] See BEVERAGE, H. H., C. W. RICE, and E. W. KELLOGG, "The Wave Antenna," *Jour. Am. Inst. Elec. Eng.*, March, April, May, June and July, 1923.

This would be the case if the speed of travel of the wave were the same as that of the disturbances along the wire. Actually, this latter speed is *less* than that of the wave (being substantially equal to $1/\sqrt{LC}$, where L and C are, respectively, the inductance and capacity of the wire per unit length), so that after a certain time, the wave traveling faster than the disturbance, the latter is left behind in its motion along the wire, and caught up by the following wave, of opposite field polarity, when the disturbance ceases to grow and begins to decrease in magnitude. The maximum antenna wire length which may be effectively used is, hence, limited to approximately one wave length of the received signals, although this need not be realized very closely (a difference of even 25 per cent will not greatly reduce the signal strength).

A signal coming from the opposite direction (wave traveling from A to B) sets up but a very small disturbance at the receiver end A. This disturbance travels along AB with the wave and builds up in a manner similar to that explained above, and attains its maximum upon reaching the end B. If now this end B were left open, or were grounded through a low resistance connection, this disturbance would be reflected back into the antenna, in the manner explained above in connection with other types of antennae, and would actuate the receiving set at A. The antenna would then not have any unidirectional property. If, on the other hand, the end B is connected to ground through a resistance R, the energy of the disturbance traveling from A to B and thence to ground is transformed into heat and dissipated, and can no longer operate the receiving set. The damping resistance R must, for properly performing this function, be adjusted to the antenna characteristics to a value substantially equal to $\sqrt{L/C}$.

The antenna is thus made unidirectional, receiving with maximum intensity signals which propagate in the direction BA, being substantially unaffected by signals coming in laterally from either side, and, if properly adjusted, completely nonresponsive to signals traveling in the direction AB.

An improvement of this method of reception consists in placing the receiving set S at the same end B as the damping resistance R, thus permitting all adjustments to be made at one end of the antenna line. The receiving set S and resistance R are then both at the end of the antenna nearest to the transmitting station. To obtain this effect, the antenna is made up of two wires AB, $A'B'$, connected as shown in Fig. 59. The received signals, traveling from right to left, set up in the two wires identical electrical disturbances which arrive simultaneously at the ends A and A', where they flow to the ground through a winding W, magnified through the building-up process as explained before. This winding W constitutes the primary of the transformer

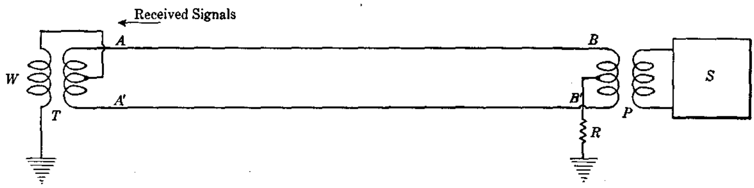


FIG. 59.

T. The current flowing through it induces an electromotive force between the ends A and A' of the antenna wires, setting up a current in the closed circuit $AA'B'BA$ which energizes the receiving set S through the transformer P. On the other hand, a signal coming from left to right produces a surge in the two wires AB and $A'B'$ which, upon reaching the ends B and B', flows to ground through the resistance R, where its energy is dissipated and does not affect the receiving set S, since it flows in opposite directions in the two halves of the primary winding of the transformer P, therefore inducing no electromotive force in its secondary.

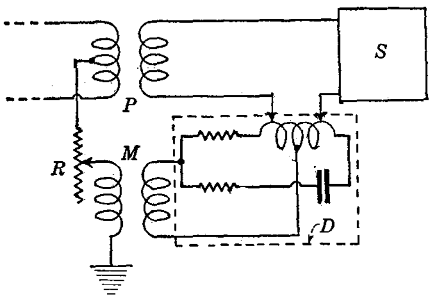


FIG. 60.

In order to reduce further or to eliminate completely currents interfering with the desired signals, the current which flows to ground through the resistance R may be combined with the current in the receiving set S for neutralizing the interfering signals. In order that this neutralizing current may be impressed upon the receiving circuit with the proper phase and amplitude, a phase-adjusting device D (Fig. 60) is connected in the input circuit of the receiving set S, in series with the winding of its energizing transformer P, and coupled more or less closely through the transformer M to the circuit branch carrying to ground the current of the resistance R. Another method consists in connecting an inductance L and capacity C in series with the resistance R (Fig. 61) and adjusted to set up in the antenna circuit a neutralizing wave for compensating, notably, a possible reflection wave originating at the far end of the antenna.

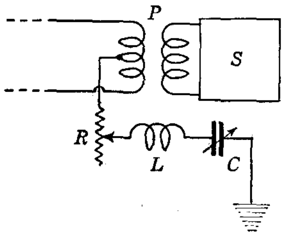


FIG. 61.

Finally, the antenna circuit being untuned and its resonance at signal frequency being rendered impossible through the use of the damping resistance R, it may be used for multiplex reception, that is, the simultaneous reception of a plurality of transmitting stations located in a common general direction. To this effect, a corresponding number of receiving sets S, S' and S'', . . . , are coupled to the antenna circuit and tuned respectively

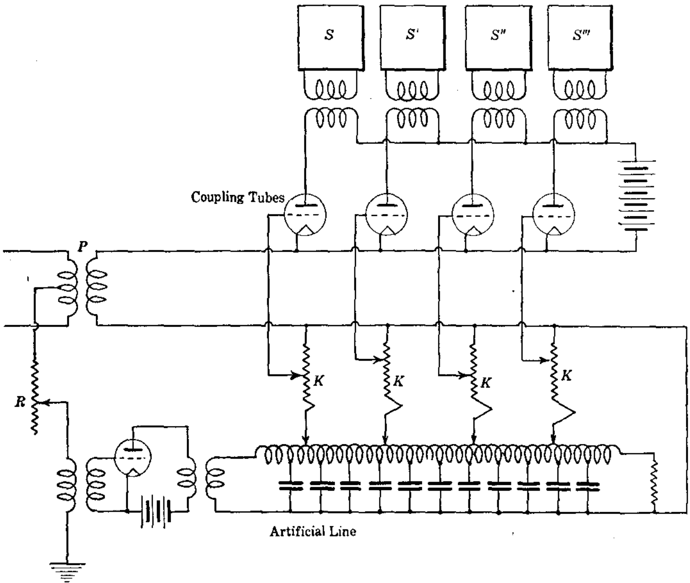


FIG. 62.

to the frequencies of the stations to be received. In order to prevent possible interactions between the various receiving sets, these are electrostatically and magnetically shielded from each other, and furthermore are coupled to the antenna through unidirectional relays, constituted by three-electrode vacuum tubes, as shown in Fig. 62. The damping resistance *R* is adjusted to damp out a mean wave length, and the compensating electromotive forces are obtained, for each set, by tapping off, through an intensity-adjusting potential divider *K*, an artificial line energized by the damping resistance current. As many as nine different receiving sets have thus been actuated from a single antenna.

Short Waves.—When radio communication is effected by means of short waves, of, say, less than 150 m., certain factors which were of lesser moment with the longer waves become of great importance and must be taken into account.[1]

At such high frequencies the waves emitted by a transmitting station *T* (Fig. 63) set up comparatively strong currents in the ground, which consume a correspondingly great percentage of the energy stored in the lower part or foot of the waves and weaken their field intensity. A receiving set located on or near the ground will, hence, receive signals which rapidly become weaker when the distance between transmitting and receiving stations increases, and soon disappear entirely, when some point *R* (Fig. 63) is reached.

That part of the waves which is at a higher altitude above the ground is not absorbed however, and continues traveling in a straight line away from the transmitter. The earth's surface being curved into a sphere, this straight-line propagation brings the waves to an increasing height above the ground until they reach the layer of ionized conducting air which, as stated above, surrounds the earth's atmosphere. At this moment, as explained before for the longer waves, the waves become bent forward, and several possibilities arise as to their further direction of travel.

Those waves, the line of travel of which rises steeply above the ground (lines *TA* in Fig. 63 for instance), simply alter their direction of propagation somewhat, but continue to travel away from the earth, carrying along energy which is definitely lost for the radio transmission.

[1] Taylor, Dr. A. H., and Dr. E. O. Hulburt, *Q. S. T.*, pp. 12–21, October, 1925. Also Baker, W. G. and C. W. Rice, "Refraction of Short Radio Waves in the Upper Atmosphere," *Jour. Am. Insi. Elec. Eng.*, June, 1926.

Other waves, the lines of travel of which leave the ground surface and strike the ionized layer under a smaller angle, like the line *TB*, bend slightly upon reaching the layer of ionized air, which then makes them travel down toward the earth, so that they strike the ground at some point *C*, when the signals from the transmitting station *T* will be heard again.

Finally, waves leaving the station *T* at a very small angle (line *TD*) will, after a reflection at *D*, come down grazing the ground surface at *E* and continuing in a straight line, rising again toward the higher atmosphere, where they may be reflected downward a second time, and so on, the waves thus circling the globe in a succession of leaps or reflections.

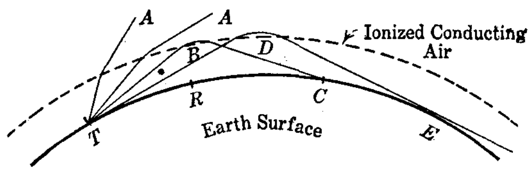


Fig. 63.

Summarizing the above, signals from the transmitting station *T* will be received, with continually weaker intensity, over a distance *TR* due to the so-called *ground wave*, the signals being set up by the lower part or foot of the waves traveling *directly* along the ground from the transmitter to the receiver. A silent zone *RC* follows then, in which no signals are received, after which signals are heard again with good intensity over a distance *CE*, from waves *reflected* downward from the upper atmosphere. A second, third, and even more alternate silent and signal zones may follow after these.

The maximum distance of travel *TR* of the ground or direct wave and the "skip distance" *TC* are, of course, dependent upon the wave length, the former increasing and the latter decreasing with increasing wave length. Above a certain value of wave length the first silent zone *RC* may therefore not exist, the zone *TR* over which the ground wave is being received reaching out beyond point *C*. A receiving station within this overlapping region of the two zones *TR* and *CE*, will hence receive the ground (or direct) wave and also the sky (or reflected) wave. The paths followed by these two waves between the transmitting and receiving stations being of different lengths, the fields set up by the two waves at the receiving station may or may not be in phase with each other, depending upon the actual values of the path lengths. The signal will be of maximum strength if the fields are in phase, and of minimum strength (which may be equal to zero) if they are in phase opposition. There will thus be points where the received signals are of fair intensity and neighboring points where the signals are very faint or even entirely absent.

Finally, all these phenomena, in addition to the signal wave length, depend also upon the height above the ground at which the waves radiated by the transmitting station encounter the layer of ionized conducting air. Now, for a given locality, this height is smallest during the day when the sun's ionizing radiations penetrate the atmosphere to the greatest depth. At night, on the contrary, the air is ionized only by weaker radiations (light reflected by the moon and various other cosmic radiations), so that the height of insulating, un-ionized air is then much greater. The transmission ranges obtained by means of the reflected waves are correspondingly greater during the night. Also, at dawn and at sunset the air rapidly ionizes and de-ionizes, lowering or raising the limit of the ionized layer, and consequently altering the positions, on the earth's surface, of the zones of silence and of greatest signal intensity.

Lastly, the limit between the ionized and un-ionized air is not a smooth surface, being constantly disturbed by air currents and winds which, changing the direction of the reflecting surface, also change that of the reflected waves, sometimes to a considerable extent. The effect is particularly noticeable in those regions where both the direct and reflected waves are received, and where the above-described interference phenomena take place (standing waves resulting in points of silence neighboring with points of great signal strength), and takes the form of rapid and erratic changes in received signal intensity and direction, designated under the general name of "fading."

Loop Antennae.—An open oscillator, as defined at the beginning of this chapter, may be considered as obtained by spreading apart the condenser plates of a closed oscillator. The circuit condenser is then constituted by the antenna, one plate being formed by the ground, the other by the aerial wires. The open oscillator is characterized by the fact that its linear dimensions are of the same order as the wave length of the emitted or received signals, and that electrical energy is readily absorbed by the circuit from incoming electromagnetic waves (receiving antenna) or radiated by the circuit in the form of waves (transmitting antenna).

Similar characteristics may be obtained by means of so-called *loop antennae*, which may therefore be considered as "open" circuits, as just described, constituted in giving to the inductance coil of an oscillator circuit as great linear dimensions as possible. Such loop antennae provide pronounced directional effects. 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 used.

In order to understand the underlying principles, consider first two single-wire vertical antennae AB and $A'B'$ (Fig. 64) of equal length and 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

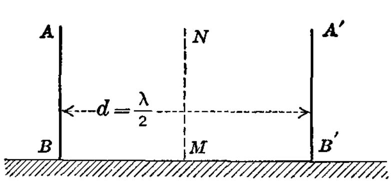


FIG. 64.

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, the field at a distance $\lambda/2$ (or $3/2\lambda$, $5/2\lambda$, etc.) from AB is at every instant 180 deg. out of phase with the field at the wire AB. If, then, the alternating current in the antenna $A'B'$ is 180 deg. out of phase with that in the antenna AB, it will produce in the plane $ABB'A'$ a field which is at every instant in phase with the field produced by the current in AB. Under such conditions, the fields of the antennae will add to each other at all points of the plane containing the two antennae. On the other hand, at all points equidistant from the two antennae, that is, at all points of the plane at right angles to that of the two antennae and located midway between them, the fields of the two antennae 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 antennae and with zero amplitude in the plane at right angles to this and midway between the antennae.

Now, if the two antennae are separated by a distance d smaller than $\lambda/2$, but kept oscillating 180 deg. out of phase, similar effects take place, with the difference however that the resultant maximum field is less than in the previous cases. This may be demonstrated mathematically as follows:[1]

Assume an antenna MN (Fig. 64) to be located midway between the two antennae AB and $A'B'$. This antenna would produce at a point P of the plane $ABB'A'$ a radiation field proportional to the hypothetical current $I_0 \sin 2\pi ft$ flowing in it. The field set up a 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[2]

$$-I_0 \sin\left(2\pi ft - \frac{\pi d}{\lambda}\right).$$

The field at point P resulting from the currents in the antennae AB and $A'B'$ is then proportional to their algebraic sum

$$I_0\left[\sin\left(2\pi ft + \frac{\pi d}{\lambda}\right) - \sin\left(2\pi ft - \frac{\pi d}{\lambda}\right)\right] = 2I_0 \sin\frac{\pi d}{\lambda}\cos 2\pi ft.$$

This expression, and hence the field, is a maximum when $d = \lambda/2$ and gradually decreases as d increases or decreases from this value.

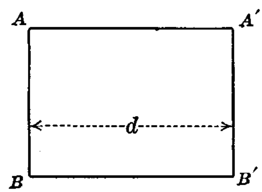


FIG. 65.

Now consider the closed rectangular metallic circuit or loop $ABB'A'$ (Fig. 65) 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 antennae in which the currents are 180 deg. out of phase and it follows, as shown above, that such a circuit has a directional effect giving maximum radiation in the plane of the loop, and zero radiation in a plane at right angles to that of the loop. The radiation will be small, however, if the loop is of small dimensions as compared with the wave length, that is, if the radio d/λ is small. It increases as the wave length

[1] TISSOT, C., *Manuel Elémentaire de Télégraphie sans Fil*, Paris, 1918.

[2] The currents in AB and $A'B'$ are 180 deg. out of phase, as explained before and, hence, are of opposite signs at every instant. This explains the minus sign before the next expression.

decreases, since then d/λ increases and d tends to become more nearly equal to $\lambda/2$.

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, preferably, all of the inductance of an oscillatory circuit, as stated at the beginning of this section. Figure 66 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

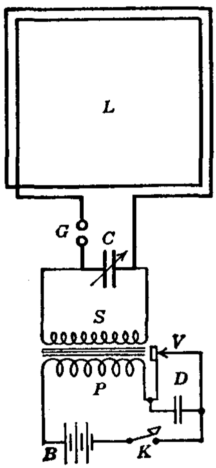


FIG. 66.

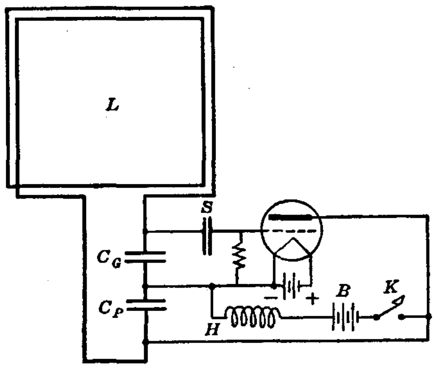


FIG. 67.

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. 67 is a continuous-wave loop transmitting circuit. This is essentially the same circuit as shown in Fig. 160 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 generally made of only one or a few turns of heavy wire—say less than ten. Its linear dimensions seldom exceed a few feet (although much larger loops have been used in permanent stations), and it may be of any convenient shape. Advantages of such loop antennae are their comparatively small dimensions, sharp directional characteristics, and ease of orientation, it being necessary merely to turn the loop to change the direction of transmission of the waves. They are, however, for efficient operation, suited for short waves only, for the reasons previously given.

Loop antennae may be used equally well for receiving signals, for which they are particularly well suited when used in connection with vacuum-tube amplifiers. A simple form of damped or

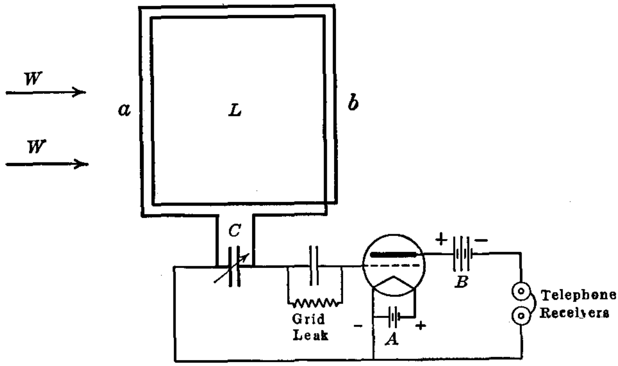


FIG. 68.

modulated wave-loop receiving circuit is shown in Fig. 68. 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. These terminals are connected to a detector and telephone receiver circuit. In the present instance a vacuum-tube detector is used, although, in general, a multi-stage amplifier is required, especially if a loop of small dimensions is used.

The loop antenna has directional receiving characteristics which are similar to its transmitting characteristics, signals being received with maximum intensity when the plane of the receiving loop is parallel to the direction of travel of the waves, and with zero or minimum intensity when the plane is at right angles to that direction. Thus, suppose the loop (Fig. 68) to be in a plane parallel to the direction of motion W of the waves to be received. The side a of the loop will then 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 electromotive force induced across the conductor a is out of phase with, and leads the otherwise equal electromotive force induced in b, thus setting up a current in the loop circuit. The effect, which is dependent upon 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, being a maximum when this distance between a and b equals one-half the wave length, as in the case of the transmitting loop.

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 correspondingly reduced. If the loop is perpendicular to the direction of travel of the waves (hence parallel to the wave front), then all the points of the loop are reached by the wave at the same instant 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 distant transmitting stations. When used for this purpose, the loop is known as a "goniometer" or direction finder, or radio compass.

Constants of Loop Antennae.—The electromotive force e induced in the loop by the alternating magnetic flux Φ set up about the loop by the incoming waves is at each instant expressed by the relation[1]

$$e = \frac{d\Phi}{dt},$$

from which it follows that the effective electromotive force induced in the loop is

$$E = \sqrt{2}\pi f N A H,$$

where N = number of turns of the loop,
A = cross-sectional area of the loop,
H = effective field intensity,
f = frequency of the incoming waves.

This alternating electromotive force sets up a current I in the loop circuit, and if this circuit is tuned to the frequency f, its imped-

[1] See *Electrical World*, vol. 73, No. 10, pp. 464–467, March 8, 1919.

ance reduces to the effective resistance R of the loop, and the current is then

$$I = \frac{E}{R} = \frac{\sqrt{2}\pi f NAH}{R}.$$

In the case of a vacuum-tube detector, connected as shown in Fig. 68, the intensity of the received signals may be indicated by the effective potential E_C developed across the tuning condenser C by this oscillating current I flowing in the loop circuit. If λ is the wave length of the received signals, this condenser voltage is equal to

$$E_C = \frac{mNAL}{\lambda^2 R}$$

where m is a constant.

If, therefore, a given loop antenna is used for receiving signals of different wave lengths, the condenser voltage, and hence the

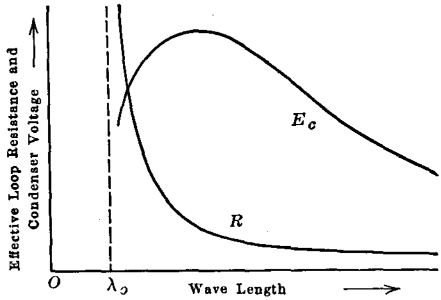


FIG. 69.

signal intensity, tends to increase as the wave length becomes shorter. On the other hand, the effective resistance R of the loop increases with the frequency, or increases with decreasing wave length, thus having an effect on the signal strength tending to offset that of the reduced wave length. The condenser voltage and signal strength, therefore, pass through a maximum value as the wave length of the received signals is gradually reduced. These relations are illustrated by the curves of Fig. 69, which are self-explanatory. In general, the maximum signal intensity occurs at a wave length of two or 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 distributed inductance and capacity. It is therefore good practice to use a loop antenna having a fundamental wave length not exceeding one-half or one-third the signal wave length to be received.

The above formula also shows that the signal strength is greater the greater the area, number of turns, and inductance of the loop. Increasing the area reduces the inductance, however, and increasing the number of turns while increasing the inductance also increases the resistance and distributed capacity by requiring a greater length of wire. These various factors must therefore be considered in the design of a loop for the desired range of wave lengths to be covered. 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 small resistance, inductance, and distributed capacity. On the other hand, for long wave reception the loop should be given a greater number of turns, say at least twenty or thirty.

For a given loop the inductance and distributed capacity increase when the turns are wound closely, and also when the number of turns is increased. The turns of a loop should therefore be so spaced as to give a suitably high inductance without giving too great a distributed capacity, as this would limit the use of the loop to the longer waves. 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 approximately the same for different loop dimensions and turn spacings.[1]

Direction Finding.—As explained before, the intensity of the signals received by means of a loop antenna depends upon 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 electromotive force E and, therefore, the current in the tuned receiving-loop circuit are essentially proportional to the magnetic flux variations in the loop, it follows that this electromotive force 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 is represented in polar coördinates by the two tangent circles of Fig. 70, where the length of the vector OA is proportional to the received current for the position OA of the loop. The current is thus

[1] See paper by Dr. J. H. Dellinger, *Proc. Am. Inst. Elec. Eng.*, October, 1919.

a maximum when the angle α is zero, and equal to zero when α is equal to 90 deg.

Having then a loop receiving circuit, such as shown in Fig. 68, with provisions for rotating the loop antenna about a vertical axis, the loop may be turned, after being tuned to the signal wave length, into such a position that the signals are heard with maximum intensity. Its plane will then coincide with the direction of travel of the waves and point toward the distant transmitting antenna. If the position of the loop is then recorded and a similar operation is repeated from another station of known location, it is then 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.

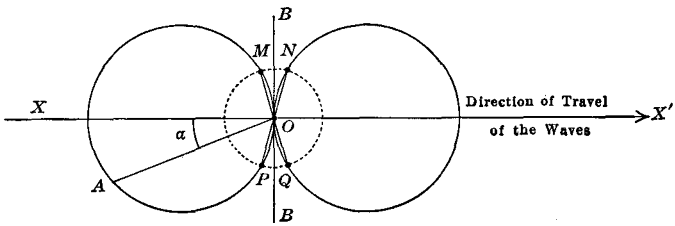


FIG. 70.

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 deg. than when it is approximately zero. That is, the device is 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 nonexistence of a signal, rather than the maximum intensity of the signal. The general practice is, therefore, to turn the loop into a position for which the signals are not heard at all, this permitting a more accurate adjustment. The loop is then normal instead of parallel to the direction of travel of the waves.

This last statement is true, however, only for the ideal case in which the detector and telephone receivers have infinite sensitivity. Actually, a certain very small current is required to operate the receiving apparatus, and if the received current is less than this, no sound will be heard in the telephone receivers. If the vector OM (Fig. 70) represents this minimum current producing a sound in the telephone receivers, no signals will then be heard when the loop occupies any of the positions between OM and ON, or between OP and OQ. In other words, if the loop is rotated from its initial position OX, the signal gradually decreases in intensity and entirely fades out when the loop comes to the angular position OM. No signal is then heard between the positions OM and ON. The sound then reappears when the loop passes the position ON, increases to a maximum for the position OX', decreases to zero for the position OQ, reappears for the position OP, and grows to the original maximum for the position OX. The mean position between these four positions OM, ON, OQ, and OP is then taken as corresponding to the absolute zero position BB of the loop. It follows that the smaller the minimum current required for producing a sound in the telephone receivers, the closer the four loop positions OM, ON, OQ, and OP will come to the actual zero position BB. This minimum current is made smallest by increasing the sensitiveness of the device, through the use of multi-stage vacuum-tube amplifiers.

Unidirectional Goniometer.—The above explanation has shown the loop antenna to have marked directional characteristics, but these are essentially *bilateral*, owing to the fact that when the loop is pointed toward the transmitting station, giving maximum signal intensity, and then turned around 180 deg., the same maximum intensity will be received.

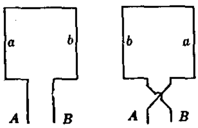


Fig. 71.

It is, however, possible to obtain *unilateral* reception through the simultaneous use of a loop antenna, as just described, and an auxiliary antenna, preferably a simple vertical antenna receiving equally well from all directions.[1] The method is based on the fact that, although the signal strength is not altered when the loop is pointed toward the transmitting station, and then turned around 180 deg. as just explained, the *phase* of the current and electromotive force set up in the loop by the received waves becomes reversed. In other words, turning the loop around is equivalent to leaving it in its original position and reversing its terminal connections, as shown in Fig. 71.

[1] See Bellini, E., "Nouveau Radio Goniomètre avec Levée du Doute," *Onde Electrique*, pp. 233–253, May 1924.

If, then, the current or electromotive force in the loop be made to operate in the receiving set, *and also* the current or electromotive force in the vertical antenna, the action of the loop may be made to add to that of the vertical antenna in one of its positions and subtract from it in the other position, so that these two positions of the loop will give different signal strengths, say maximum and zero, and unidirectional reception will be achieved.

It is necessary, for such operation, that the current and electromotive force set up in the loop by the incoming waves be respectively in phase (or in phase opposition) with the current and electromotive force in the antenna, or else that suitable compensating arrangements be used, in order that the loop and antenna actions add or subtract exactly. Now, the electromotive force set up in the loop by the waves is a maximum when the rate of change of the magnetic field component of the wave is a maximum, that is, when this magnetic component passes through zero. On the other hand, the electromotive force in the vertical antenna is in phase with the electric field component of the wave, hence 90 deg. out of phase with the loop electromotive force. And if both the loop and antenna are tuned to the frequency of the received signals, the currents in them will, respectively, be in phase with the loop and antenna electromotive forces, hence, like them, 90 deg. out of phase with each other.

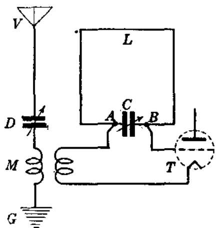


FIG. 72.

In order to bring the two actions into phase with each other the receiving set (of which only the first input amplifier tube T is shown in Fig. 72) is *connected directly* to the terminals A, B of the loop L, which are also the terminals of the loop tuning condenser C, and *coupled inductively*, through a transformer M, to the vertical antenna V, which is tuned by means of the condenser D. In this manner, the electromotive force induced by the vertical-antenna current in the receiving set input circuit is 90 deg. out of phase with this antenna current,[1] and hence in phase (or in phase opposition) with the loop electromotive force. Perfect balance may then be obtained between the two actions by coupling the two windings of the transformer M more

[1] Because it is maximum when the antenna current *variation* is maximum, and therefore when the current passes through zero.

or less closely, so that zero signal is obtained in one of the loop positions, and maximum signal in the other, 180 deg. away from this.

A simplification of this scheme consists in using the loop itself as a vertical antenna, in addition to its function as a loop. The preceding circuit arrangement then takes the form shown in Fig. 73. The loop tuning condenser is here split in two equal parts C, C' mounted on a common shaft and adjustable simultaneously, and connection to the ground G is made from their common point through the transformer M and vertical antenna tuning condenser D as before. The loop current then circulates in the circuit $LACC'BL$ while the vertical antenna current flows in the parallel branches $GDMCAL$ and $GDMC'BL$. As in Fig. 72, the receiving set is connected directly to the loop terminals A and B, and coupled inductively to the vertical antenna LG through the transformer M.

If instead of using a *tuned* vertical antenna, an untuned or *aperiodic* antenna is used, the vertical antenna current is no longer in phase with the antenna electromotive force, but substantially 90 deg. out of phase with it, and is hence in phase or in phase opposition with the loop electromotive force and current, with which it may be combined directly, avoiding the inductive coupling between the antenna and receiving set provided heretofore by the transformer M. The arrangement is represented in Fig. 74, the mid-point of the loop tuning condenser CC' being here connected to ground through a condenser F of such capacity as to considerably detune the vertical antenna LG. The receiving set is connected directly to the loop and vertical antenna, and is thus energized by the electromotive forces developed across one of the loop condensers C' and the vertical-antenna condenser F.

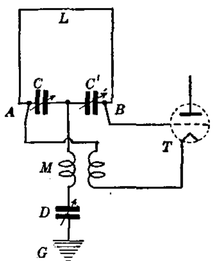


FIG. 73.

In order that the vertical antenna electromotive force may be so adjusted to be equal to the loop electromotive force acting in the receiving circuit, as is necessary for obtaining zero response in one of the two loop positions, the upper end of the vertical antenna is, as shown, connected to ground through an adjustable resistance R. The effect of the vertical antenna on the receiving set may thus be adjusted by setting the resistance R to some suitable value.[1] This most suitable value of course depends on the wave length of the received signals.

A further simplification which experiment has shown to operate practically as well consists in omitting one of the two loop tuning condensers, as in Fig. 75.

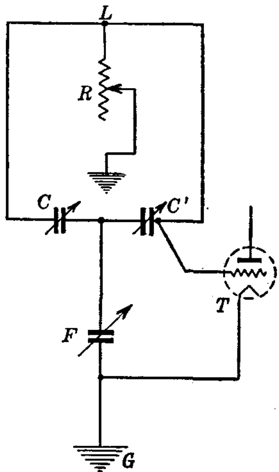


FIG. 74.

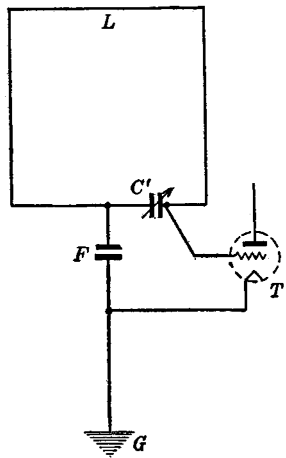


FIG. 75.

Loop Unbalance.—Referring back to the simple loop arrangement of Fig. 68, it is seen that one of the loop terminals is connected to the grid, the other to the filament of the receiving tube. Since the filament is connected to battery A, telephone receivers, and battery B, a comparatively large electrostatic capacity to

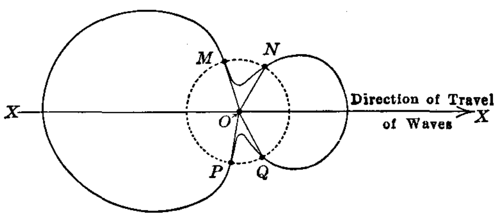


FIG. 76.

ground is established through these between this end of the loop and the ground. This unbalances the circuit and allows the loop to operate in addition, as a vertical antenna, as just described in the preceding arrangements, this capacity to ground operating like the condenser F of the last two figures. This, as just explained,

[1] BELLINI, E., *loc. cit.*

produces a distortion of the ideal curve of Fig. 70, as shown in Fig. 76. 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, which may seriously impair the accuracy of the direction measurements, particularly when the detector and telephone circuit are of lesser sensitivity.

This asymmetry of the curve may be corrected by connecting a condenser of suitable capacity between the grid and the other terminal of the loop (the side connected to the grid of the tube). Or else it may be *increased* and a truly unidirectional effect obtained by connecting the filament end of the loop to ground through a large condenser, or directly to ground, which gives the same results as indicated in the preceding section.

Aircraft Radio Compass.—The single-loop direction finders described above were shown to have maximum sensitivity when adjusted for zero signal. That is, the loop 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, because of engine and wind noises, and the setting of the loop 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 is quite inaccurate, as explained before, on account of the small variation of intensity of the signals for relatively large angular rotations of the loop about its true maximum signal position. This made it desirable to devise a method which would have the accuracy of the zero signal setting with the loop on maximum signal.

These requirements are successfully fulfilled by the circuit of Fig. 77, which illustrates the 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, which is generally used on account of its greater sensitivity, and because a crystal detector may 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 the circuit. The loop L_1 is then shunted by the two parallel connected condensers C_1 and C_2.

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

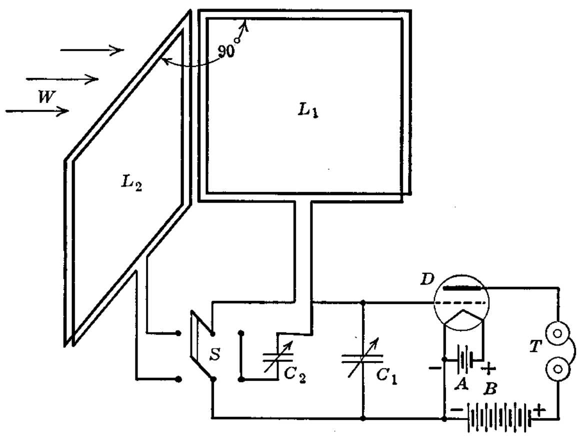


FIG. 77.

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. This results in a decreased electrostatic capacity and increased inductance of the receiving oscillatory circuit. If the condenser C_2 were given a suitable setting before throwing the switch, the product LC of the total inductance and capacity 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 capacity of the condenser C_2. The circuit can thus be made to tune for both positions of the switch S.

With the circuit so adjusted and tuned, its use and operation on an airplane is 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 direction finder on the airplane is 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 device follow. 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. 77) is first closed to the right, thus disconnecting the loop L_2, and making use of the loop antenna L_1 only 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 of maximum audibility. This brings 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 deg. to the loop L_1, will be normal to this direction W, and the waves therefore induce no electromotive force 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 signal intensity. 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 zero signal, and since the loop L_2 is then around its most sensitive position, this small deviation from the actual zero position will manifest itself in a sharp increase or decrease in the intensity of the signals when the switch S is closed from right to left. This method thus permits great accuracy despite the fact that the setting is made for maximum instead of zero signal intensity.

Two methods have been used to construct and instal loop antennae on airplanes. In one, the loops are mounted permanently on the wings and vertical struts of the airplane, and in order to turn the loop antennae, it is necessary to turn the entire craft. 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 bow of 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 course of the airplane is maintained 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 aircraft.

CHAPTER IV

DAMPED-WAVE RADIO TELEGRAPHY

Recalling the previous discussion, radio telegraphic communication is accomplished by generating high-frequency alternating currents in a radiating circuit (open oscillator or antenna), currents being then set up in a distant receiving circuit through the medium of wave-propagated radiation fields. Systems of radio telegraphy making use of the oscillatory discharge of a condenser for producing the high-frequency currents are called *damped-wave* radio telegraph systems, for the reason that the oscillations are not truly constant or continuous alternating currents, but are periodically damped. The condenser discharge circuit generally comprising a spark gap of some sort, the name of "spark set" is also sometimes used for designating a damped-wave transmitting set.

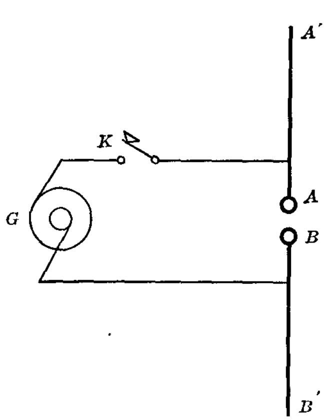


FIG. 78.

Damped-wave Radio Transmitting Circuits.—The operation of the circuit of Fig. 78 as a damped-wave radio transmitting circuit may readily be understood by referring to the preceding chapters. The key K being closed and the spark gap AB suitably adjusted, an oscillating discharge takes place in the open oscillator circuit $A'B'$ every half-cycle of the alternator G, as explained for the closed oscillator of Fig. 29. Thus a succession of groups of damped oscillations takes place in the open oscillator as long as the key K is closed, producing the radiation of a succession of damped wave trains into space. The frequency of the oscillations is 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, is equal to twice the alternator frequency, there being one oscillating discharge per half-cycle.

As explained in the preceding chapter, the actual radiating circuits used in practice generally take the form of an *antenna*, so that the circuit would be as shown in Fig. 79. On account of the rather small electrostatic capacity of such an antenna, it is necessary to charge it, before every oscillating discharge, to a high potential, in order to provide a large store of electrostatic energy for producing the oscillating discharge. If this is not done, the oscillatory current will be of small amplitude, and the range of transmission correspondingly reduced. For this reason, the aerial and ground are connected to the alternator *G* through a step-up transformer *T* developing several thousand volts difference between the aerial and ground.

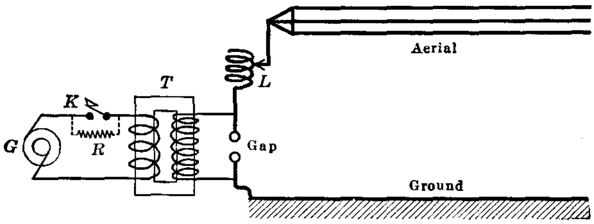


FIG. 79.

The key (operated by hand or automatically) is usually connected in the low-voltage circuit, in series with the transformer primary winding, as shown in the figure. When the key is closed, an alternating electromotive force is impressed through the transformer on the oscillatory antenna circuit, and a train of damped waves is sent out at every half-cycle. By closing the key for shorter or longer periods of time, a short or long series of wave trains is radiated, corresponding to the dots and dashes of the Morse 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 the alternator speed and poor operation. The key may therefore be shunted by a resistance *R* which, when the key is open, reduces the current through the transformer primary, and cuts down the secondary voltage to a value just below the break-down voltage of the spark gap, thus greatly minimizing the alternator load variations.

The key may also be placed in series with the field winding of the alternator. The alternator is then excited only when the key is closed. With this method however, the inductance of the field winding prevents a rapid building up of the alternator field current, and thus reduces the maximum possible sending speed.

As 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 connected in series with it, or else a condenser, and so adjusted as to give the desired wave length.

The above method, 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

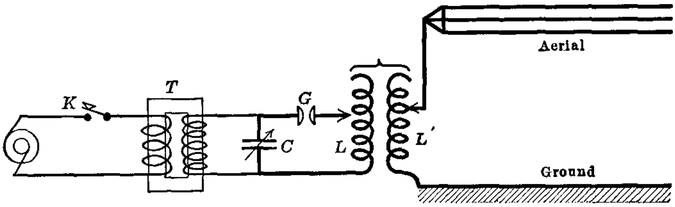


FIG. 80.

words, the oscillations have a high decrement. This results in a flat resonance curve and prevents sharp tuning at the receiving station, as studied further below.

In order to avoid such a high decrement of the oscillations, and prevent a large part of the energy supplied by the alternator from being wasted in the spark resistance instead of radiated, a so-called *indirect excitation* method is employed. This is illustrated in Fig. 80, and reference is here made to what has been said in Chap. II on impulse excitation of coupled circuits. The method consists in coupling the antenna circuit to an intermediate oscillatory circuit containing the spark gap. Oscillations are then excited in the closed oscillatory circuit CLG when the key is closed, in the same manner as described above for the direct excitation of the antenna circuit. 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 Chap. II relating to tuned coupled circuits. As was then shown, oscillations of maximum amplitude are obtained in the secondary circuit (which here is the antenna circuit) if the latter is tuned to the frequency of the closed primary oscillatory circuit, and coupled to it with the critical degree of coupling for maximum power transfer. Furthermore, in order to prevent the energy transferred from the closed oscillatory circuit to the antenna circuit from reacting back upon the closed circuit, which, as known, would cause the antenna to oscillate and radiate at *two* frequencies simultaneously, 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 an adjustment of their natural frequencies. An adjustable coupling may be obtained by varying the relative positions of the coupling coils, rotating one coil with respect to the other, or 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. 80 is illustrated by the curves of Fig. 81. When the key is closed, the alternator electromotive force is impressed upon the transformer *T*, stepped up, and applied across the condenser *C* and gap *G*. This electromotive force is shown by the upper curve (Fig. 81). In every half-cycle, as this impressed electromotive force reaches the value *a*, the gap *G* breaks down, and an oscillating current flows in the circuit *LGC* (Fig. 80). This oscillation is rapidly damped, as shown by the second curve of Fig. 81, principally due to the fact that energy is being transferred to the antenna circuit coupled to it. The oscillating current induced in the antenna is represented by the lower curve of the figure.

As in the case of coupled "closed" circuits, this induced oscillating current grows as the energy is being transferred from the closed 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 the 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 mainly 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, the decrement of the radiated wave is consequently small, permitting sharp tuning at the distant receiving circuit.

Another advantage of using an intermediate oscillatory circuit between the antenna circuit and the power supply is due to the fact 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 electrostatic

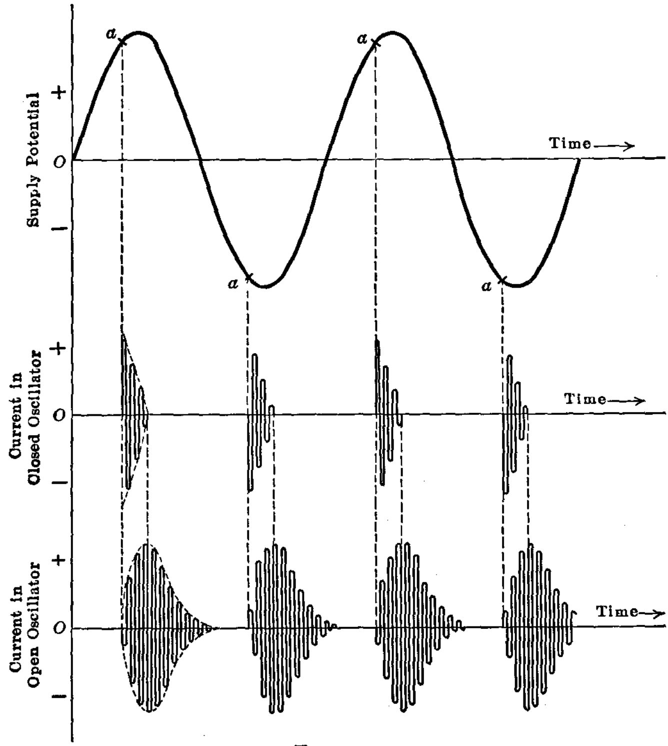


FIG. 81.

capacity of the oscillatory circuit. The capacity of the antenna circuit being small, the energy stored in it is small as compared with that stored in the closed oscillatory circuit, which comprises a condenser of as great a capacity as desired. If, then, the directly excited circuit of Fig. 79 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. 80, using the same antenna and a moderately high potential.

It should be noted that in the circuit of Fig. 80, 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 condenser were interchanged. In this latter case, however, with the gap connected directly across the transformer, the spark produces a short-circuit across the transformer, which may damage the transformer winding and give rise to a "power arc" across the gap, burning up the electrodes.

Of course, the oscillation transformer coupling the antenna and closed oscillator circuit may be an auto-transformer, as shown in Fig. 82, or any other coupling device.

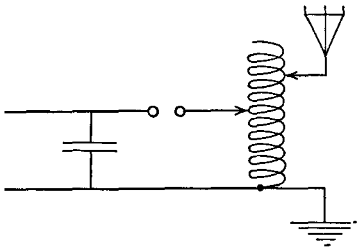


FIG. 82.

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 else coupled to the closed oscillator 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 and placed, say, at a distance of 2 or 3 yd. The closed oscillator circuit is then oscillating at the desired wave length.

The coupling between the antenna and closed oscillatory circuits is then tightened slightly, 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 as to give rise to a double-peaked resonance curve, but should give a single, sharp resonance indication at one well-defined frequency.

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

Damped-wave Radio Receiving Circuits.—The reception of signals sent out by the above methods consists essentially in having the varying radiation fields due to the oscillations of the transmitting circuit set up currents in a suitable receiving circuit, and then observing or recording these currents by means of some appropriate device. In practice, the receiving circuit is tuned to the operating frequency of the transmittng circuit in order to increase the effects of the oscillations in the receiving circuit, and permit, through resonance, the separation of signals emitted by several radio stations on different wave lengths, as described in Chap. II. Also, for the reasons explained in Chap. III, the receiving circuit should preferably have dimensions of the same order of magnitude as the wave length of the signals to be received.

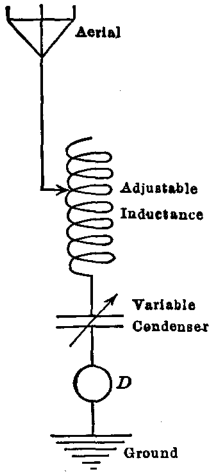


FIG. 83.

A simple receiving circuit consists, then (Fig. 83), of an antenna circuit tuned to the frequency of the transmitted high-frequency oscillations by means of an adjustable inductance and capacity, and some device *D* which will respond to or evidence the electromotive force or current induced in the receiving circuit.

The main problem here is that of providing a device *D* which will respond to the extremely feeble currents flowing in the receiving antenna. An ammeter is out of the question, as it would have too slow an action to follow the dots and dashes sent out at the minimum speed of say twenty-five or thirty words a minute.

A widely adopted method 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. The frequency of the currents induced in the receiving circuit by each wave train sent out by the transmitting circuit being generally high, however, at least 50,000 or 60,000 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 perceive no sound, since the highest audible frequencies are between 16,000 and 20,000 cycles per second. These physical limitations make it necessary to use some apparatus in conjunction with the telephone receiver to suitably modify the currents flowing in the receiving circuit so that they may then actuate the telephone receiver. This device is called a *detector* and its action is explained below.

Referring back to Fig. 81, the lower curve shows the oscillatory current produced in the transmitting antenna when the transmitting key 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, 1,000 groups of oscillations per second. Each one of these sets up oscillatory magnetic and electric fields around the transmitting antenna which, propagated through space with the speed of light in the form of waves or trains of waves, as previously explained, pass over and around the receiving antenna, inducing electromotive forces in it as represented by the upper curve of Fig. 84. The function of the detector is then to *rectify* the current flowing in the receiving circuit under the effect of this induced electromotive force. That is, the detector is a device having unidirectional conductivity, so that current may flow through it in one direction, and not, or only feebly, in the other. This gives rise to a unidirectional current in the receiving circuit instead of a symmetrical alternating current oscillation, as shown in the second curve of Fig. 84. The effect on the telephone-receiver diaphragm during each wave train is then cumulative, as explained presently.

The first impulse *a* (Fig. 84) in passing through the telephone winding produces an attraction of the diaphragm. On account of the high frequency of the oscillation, the second impulse *b* occurs before the diaphragm has had time to spring back in place, and will therefore deflect it further. So will the 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. 84. It therefore vibrates at the frequency of the wave trains instead of the frequency of the oscillations, that is, at twice the frequency of the transmitting alternator. This frequency, which in the case considered, is 1,000 cycles per second, is well within the range of audibility, and produces a sound in the telephones. The wave train frequency, that is, the number of wave trains per second, is for this reason called the *audio frequency* of the trans-

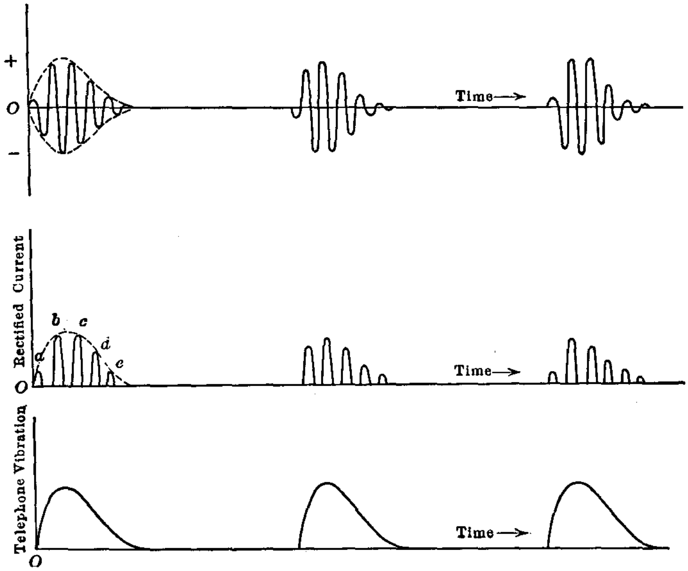


FIG. 84.

mitting 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. 85) of iron pyrite, galena, molybdenum, bornite, or carborundum, in light contact with a fine wire *W*. If the wire touches a suitable spot of the crystal surface, the system has unidirectional conductivity, offering low resistance to current flowing in one direction, say from the wire to the crystal, and high resistance to current flowing in the opposite direction. This is represented graphically by the characteristic curve of Fig. 86, where the electromotive force, positive or negative, applied to the detector terminals is plotted horizontally, and the resultant

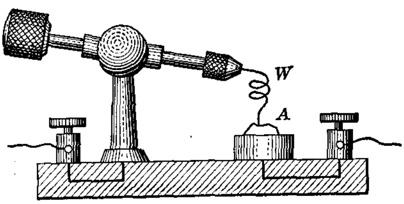


FIG. 85.

current through the detector is plotted vertically. If an oscillation produces approximately equal and opposite voltage variations across the detector, it may be seen from the asymmetrical shape

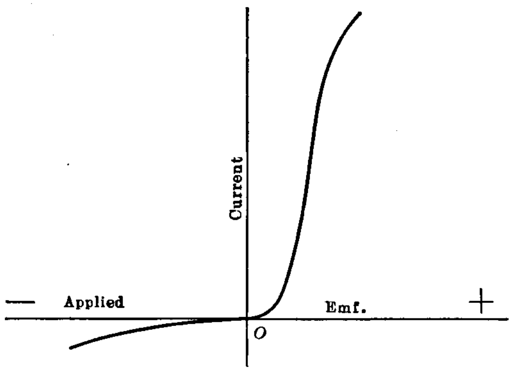


FIG. 86.

of the curve that the current flowing during the negative half of the cycle is negligible as compared to that flowing during the positive half-cycle, and rectification of the current is thus achieved.

Actually the circuit of Fig. 83 is but seldom used in practice, the placing of the detector in the antenna circuit increasing the resistance of the latter, which minimizes or entirely prevents resonance phenomena and renders the set less sensitive. The detector is, therefore, generally placed in a secondary tuned circuit coupled to the antenna circuit, which not only avoids the disadvantages just mentioned, but also permits a greater selectivity.

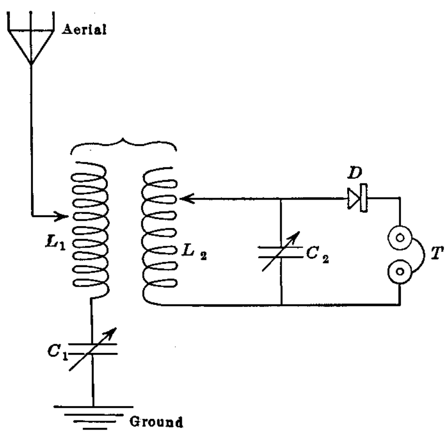


FIG. 87.

Such a coupled receiving circuit is shown in Fig. 87. The antenna circuit may be tuned by means of the variable coil L_1 and condenser C_1, while the coupled or secondary circuit is 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 oscillations induced in the secondary circuit by the current flowing in the antenna give rise to an alternating electromotive force across the coil L_2 which thus finds itself impressed upon the detector and telephone-receiver circuit branch DT in which it produces the flow of a unidirectional current, owing to the rectifying action of the detector as explained before.

It should be stated that, in order that the high-frequency voltage developed across the coil L_2 may operate in the circuit branch DT in the manner just described, it is necessary that the high impedance of the telephone-receiver windings be shunted by a small capacity, or else the high-frequency oscillations will not reach the detector D, being choked out by the telephone-winding reactance. This small "by-pass" condenser is not shown here, the capacity between adjacent turns of the telephone windings being, in most cases, sufficient for the purpose.

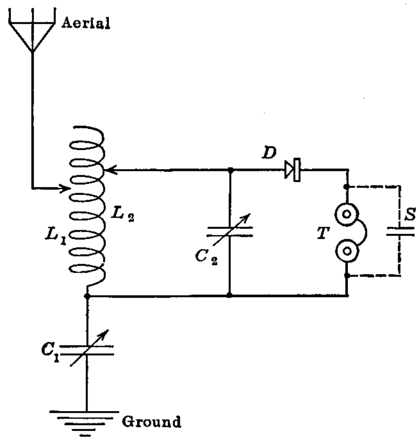


FIG. 88.

The antenna and secondary circuits may of course be coupled in different manners, for instance by means of a common coupling coil tapped off at suitable points, as shown in Fig. 88. The primary and secondary circuits considered independently are seen to be the same as in Fig. 87. A feature of the circuit of Fig. 87 not possessed by that of Fig. 88, however, 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.

Various combinations of capacitive and inductive couplings may also be used, but these are not shown here.

Note the telephone by-pass condenser S (Fig. 88) which has been shown here in order to point out the manner in which it is connected across the telephone receivers.

CHAPTER V

UNDAMPED OR CONTINUOUS-WAVE RADIO TELEGRAPHY

One of the main advantages of damped-wave or spark telegraphy, as described in the preceding chapter, is the simplicity of the apparatus required. However, this is offset by a number of drawbacks, as outlined below, which have gradually led to the use of the more perfect undamped or continuous-wave methods of radio communication.

1. The fact that the oscillations generated in a damped-wave transmitting circuit are of periodically variable amplitude (since they consist in a succession of damped oscillations) and are not a truly continuous high-frequency alternating current, results in the transmitting set operating not only at one single frequency as required, but at a plurality of more or less neighboring frequencies constituting a whole frequency band.[1] The radio station may then be tuned in at the receiving end over a corresponding range of frequencies. In other words, the tuning of the receiving circuit is not as sharp as if one frequency only were radiated by the transmitting antenna.

2. The successive oscillation or wave trains occur at an audio frequency which generally is not an exact submultiple of their radio frequency. If, then, a wave train sets up a slowly damped oscillation in a low-resistance receiving circuit, the successive wave train will not energize this circuit in the same phase, and instead of building up further the oscillation induced by the first wave train, it will during its first few cycles counteract this oscillation and thus prevent in a measure the effective resonance of the tuned receiving circuits. In other words, this dephasing of the successive oscillations of a damped wave set, equivalent to a slight frequency variation between successive wave trains, prevents the sharp tuning of the receiving circuits.

[1] This is fully explained in the study of modulated waves, in the chapter on Radio Telephony.

3. Finally, another advantage of the use of undamped or continuous oscillations may be seen from the following: Suppose an antenna circuit is excited continuously by means of undamped oscillations, or true high-frequency alternating current. The same antenna may be made to radiate the same amount of energy, at the same frequency, by damped oscillation excitation. Only in this case, 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 obviously to permit the use of lower-power apparatus continuously rather than high-power apparatus intermittently.

UNDAMPED-WAVE TRANSMITTING CIRCUITS AND METHODS

The production of undamped or continuous high-frequency alternating currents presents a number of practical difficulties, making it necessary to resort to special constructions of apparatus and also to principles not generally used in the generation of low- or medium-frequency alternating currents. Among the several methods now used for generating the high-frequency alternating currents required for radio communication are the high-frequency alternator, the oscillating arc, and the three-electrode vacuum-tube oscillator. The latter method is studied separately in a later chapter.

High-frequency Alternator

The ordinary alternator construction embodying a wound rotor and stator is not practicable for the *direct* generation of very high-frequency alternating currents. As known, the alternating current generated by such alternators, generally used for low (commercial) frequencies, has a frequency equal to

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

Thus, for producing a current of 100,000 cycles per second with such a machine having, say, a 2-foot rotor diameter and operating at 3,000 r.p.m., the number of poles required would be 4,000, giving a pole pitch of 0.019 inch. Such a construction is manifestly impracticable, in view of the difficulties of placing coils between the successive poles. On the other hand, if the rotor diameter and speed are increased, in order to reduce the number of poles required for a given frequency and increase the space for the winding between them, the rotor will not be capable of withstanding the effect of centrifugal force.

Special alternator designs and modes of operation have therefore been developed, comprising:

1. Cascade alternators with wound rotor and stator, for the *indirect generation* of high-frequency alternating currents through successive frequency multiplication, of which the Goldschmidt alternator described below constitutes a special form.

2. Inductor alternators, for the *direct generation* of high-frequency alternating currents of fundamental or harmonic frequency, such as the Alexanderson, Bethenod, and Latour alternators.

Goldschmidt Alternator.—Consider an ordinary single-phase, two-pole alternator with wound rotor and stator. A current being sent through the field winding, and the machine operated at the rate of f turns per second, the magnetic field through the armature winding will, as known, vary periodically as a result of the motion of the rotor, and will be equal to

$$\varphi = (\text{exciting field flux}) \times \sin \omega t,$$

where

$$\omega = 2\pi f.$$

This periodically variable magnetic flux linking with the armature winding induces in the latter an alternating electromotive force of its own frequency.

When, as is the case in the ordinary alternator, the current through the field winding is a constant direct current, the "exciting field flux" in the above expression is equally constant, and the magnetic field flux φ through the armature has a cyclic frequency equal to ω.

But if, instead of this constant direct current, an alternating current of frequency ω is sent through the alternator field winding, this will set up an alternating exciting field flux

$$\Phi \cos \omega t$$

of same frequency, so that the above expression becomes

$$\varphi = (\Phi \cos \omega t) \sin \omega t,$$

or, through a simple trigonometric transformation,

$$\varphi = \frac{\Phi}{2} \sin 2\omega t.$$

In this case, therefore, the magnetic field flux φ linking with the armature winding, and hence also the electromotive force induced by it in this winding, has a frequency 2ω, which is double the usual, fundamental, frequency of the machine.

Similarly, if an alternating current of frequency 2ω be sent through the field winding, the magnetic field flux φ through the armature will be

$$\varphi = (\Phi \cos 2\omega t) \sin \omega t,$$

or

$$\varphi = \frac{\Phi}{2} \sin 3\omega t - \frac{\Phi}{2} \sin \omega t,$$

which, containing a term of frequency 3ω, induces in the armature an alternating electromotive force of three times the fundamental alternator frequency.

More generally, and reasoning in the same manner, an alternating field current of frequency $n\omega$ will produce in the armature an alternating electromotive force of frequency $(n + 1)\omega$.

Now, in the ordinary alternate-pole machine with wound rotor and stator, as here considered, either winding may be used as the field winding, the other serving then as the armature winding. Suppose then that the stator winding S being used as the field winding, a constant direct current is sent through it. The rotor winding R being connected to some external circuit, an alternating current of frequency ω will be generated in it, as explained before. Now consider the rotor winding R as the field winding of the machine. An alternating current of frequency ω flowing through it, as just stated, an electromotive force of double frequency 2ω will be set up in the stator winding S, now taken as the armature winding of the machine, producing an alternating current of frequency 2ω in the stator-winding circuit.

The stator winding S being now again regarded as the field winding, this alternating current of frequency 2ω flowing through it will, in accordance with the above remarks, set up a current of frequency 3ω in the rotor circuit which now serves as the armature circuit. The process continuing, it is thus seen that one of the two windings of the alternator—the rotor winding in the present instance—carries alternating currents of fundamental and odd harmonic frequencies, while the other winding—the stator winding in this example—carries alternating currents of even harmonic frequencies.[1]

In the ordinary alternator, these higher frequency components decrease rapidly in amplitude on account of the reactance of the alternator windings, which increases with the frequency of the components, and correspondingly reduces them. In the Goldschmidt method of operating the machine, on the contrary, these components are enhanced, the method consisting in connecting

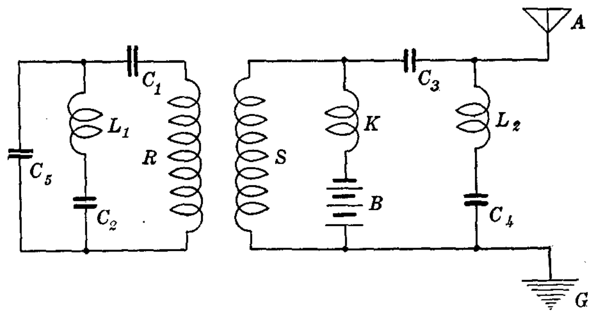


FIG. 89.

the stator and rotor windings, respectively, to external circuits of such inductance and capacity that, combined with the inductance of the alternator winding to which they are connected, they constitute oscillatory circuits resonating at the successive component frequencies.

Thus, the stator S (Fig. 89) being connected to the direct-current generator or battery B, through a choke coil K preventing the high-frequency alternating current from flowing through the battery B, an alternating current of frequency f is generated in the rotor R, which is tuned to this frequency by means of the externally connected condensers and coil $C_1L_1C_2$. As explained, this current of frequency f sets up a current of frequency $2f$ in the stator S, tuned to this frequency by the condensers and coil $C_3L_2C_4$. The stator current of frequency $2f$ produces a current of frequency $3f$ which flows in the resonant circuit RC_1C_5 and which

[1] This property of the ordinary alternator was first described by P. Boucherot, in *La Lumière Electrique*, Paris, Vol. XLVII, pp. 551–561, 1893.

in turn generates a current of frequency $4f$ in the tuned circuit comprising the stator S, condenser C_3, and transmitting antenna AG.

Thus, if the original fundamental frequency f was 10,000 cycles per second, which is about as high as this type of alternator construction permits, the stator frequency will be 40,000 cycles after four successive frequency multiplications or "reflections." However, a limit is set to the number of such successive frequency multiplications which may be made in this manner, by the energy losses in the alternator, which grow very rapidly, reducing at every step the amount of power which may be taken from the generator. This may be understood in considering that, in order to obtain a current of any given multiple frequency, it is necessary to generate in the alternator,[1] currents of *all* the lower intermediate sub-multiple frequencies with amplitudes of the same order, and even greater, than the amplitude of the desired output current. Each one of these intermediate frequency currents produces resistance and iron (hysterisis and eddy currents) energy losses which heat up the generator and increase with the frequency. The efficiency of the machine, therefore, decreases as the frequency increases. Another difficulty is the necessity of accurately tuning the circuits to the several intermediate frequencies. Alternators have, therefore, been developed along other principles, as described presently, permitting the *direct* generation of high-frequency alternating currents without the generation of intermediate frequency currents.

Inductor Alternators.—In the inductor-type alternator both field (direct-current) and alternating-current windings are mounted on the stator, while the rotor is a toothed steel drum or disc without any winding and may thus be rotated at very considerable speeds.

A simple type machine is shown in Fig. 90, which is merely given here to illustrate the principles involved,[2] and which represents a part only of the machine circumference. It comprises a circular winding or coil A, concentric with and perpendicular to the shaft of the machine, and located in a groove in the stator.

[1] The successive frequency components may also be generated in separate machines (separate cascade alternators), but this method is practically as complicated as the present one, and is not described here.

[2] See article by M. Latour, in *Revue Générale de l'Electricité*, IV, pp. 557–564, Apr. 12, 1919.

The latter is made up of packs of steel laminations B separated by non-magnetic spacers C. An equal number of lamination blocks or packs D are mounted on the rotor, separated by spacers E. The coil A being connected to a direct-current generator, in series with a choke coil to prevent the generated alternating current from flowing through the direct-current generator, and the rotor set in motion, the steel blocks B and D will periodically come face to face with each other, practically closing the magnetic circuit around the coil A. Periodic variations of the magnetic flux are thus produced around the coil, the flux being maximum or minimum according, respectively, to whether the magnetic blocks D of the rotor are face to face with the steel projections B of the stator, or with the empty spaces between them. These pulsations of the magnetic flux induce an alternating electromotive force in the coil A, which may be impressed onto a transmitting antenna circuit connected to its terminals.

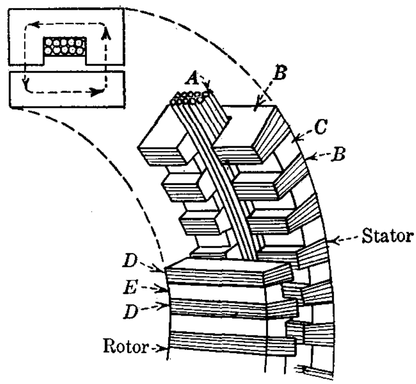


FIG. 90.

The magnetic and electrical conditions in the machine being the same every time the rotor teeth pass in front of the stator teeth, which are in equal numbers, one cycle of the generated alternating current corresponds to a rotation of the rotor equal to the angular pitch of its teeth, so that the output frequency per second is

$$\text{frequency} = (\text{number of rotor teeth}) \times \frac{\text{r.p.m.}}{60}.$$

An objection to this design is that a rather large number of turns is required in the winding A for obtaining an intense magnetic field in the machine with a reasonably small direct current. This in turn increases the impedance and distributed capacity of the winding, correspondingly decreasing the alternating-current output of the machine. It is, therefore, necessary to separate the field (direct-current) and output (alternating-current) windings. Leaving the direct-current winding as just described and shown in Fig. 90, an alternating-current winding may then be provided on each of the two toothed portions of the stator, on either side of the direct-current field coil (this is not shown in Fig. 90, being described in the following paragraph). These two alternating-current windings may then be used separately, in series, or in parallel with each other as desired.

The rotor remains the same as just described, and the magnetic flux pulsations produced by its motion in the stator teeth induce an alternating electromotive force in the alternating-current winding placed around them. There being a same number of teeth along the stator and rotor circumference, the magnetic flux variations occur in phase in all the stator teeth. The alternating current winding must, therefore, be wound in the same direction around all the stator teeth, as shown schematically in Fig. 91. The synchronous magnetic-flux variations in the teeth then induce electromotive forces of same polarity in all coils of the alternating-current winding, which, adding to each other, constitute the output electromotive force of the machine.

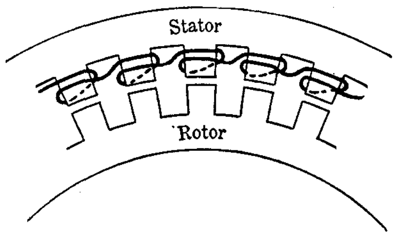


FIG. 91.

This construction, however, does not allow the most complete utilization possible of the space available between successive stator teeth. This at best is very small, so that the intervals between successive teeth are completely filled up with the copper wires and insulation of the alternating-current winding. Referring to Fig. 91, it is seen that there are *two* conductors in each slot or interval which, allowance being made for the insulation on each wire, limits the actual cross-section of wire and increases the total resistance of the winding, reducing the power which may be drawn from the machine.

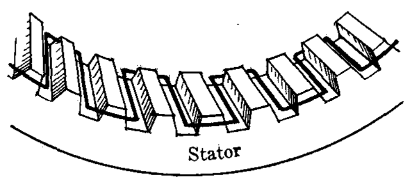


FIG. 92.

For these reasons, a so-called zigzag winding is used, as shown in Fig. 92, requiring only a single conductor to be laid in each slot, and constituting in fact a single-turn alternate-pole winding. Only with this arrangement, the wire winds in *opposite* directions around successive stator teeth, and if, as before, the rotor had the same number of teeth as the stator, the electromotive forces induced in the windings of any two successive teeth would be of opposite polarities and neutralize each other, giving no electromotive force at the machine terminals. In order to avoid this, it is necessary to set up synchronous magnetic flux variations in those stator teeth only around which the wire winds in a same direction, that is, in every other stator tooth. This is accomplished by making the number of rotor teeth equal to one-half the number of stator teeth (counted around *one* stator circumference). The stator teeth and the spaces between them are then half as wide as the rotor teeth and slots, as shown in Fig. 93, in which the same rotor is used as before with the new stator as just described.

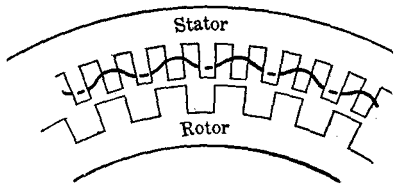


FIG. 93.

In all the inductor alternators described above, the frequency of the generated alternating current is equal to the product of the number of rotor teeth by the number of revolutions per second, as explained before. The output frequency of these machines is limited, however, the speed being limited by the mechanical strength of the rotor, bearings, etc., while the number of teeth is limited by the fact that the number of stator teeth being equal to that of the rotor teeth (homopolar machines) or equal to twice this number (alternate-pole machines), an increase of the number of teeth reduces the space available for the winding and may render the construction impracticable.

Consider again the machine of Fig. 91.[1] Around each stator tooth is wound a coil, and the motion of the rotor induces in each of these coils as electromotive force having a frequency equal to the fundamental frequency of the machine, as defined above. The electromotive forces thus induced in the various stator coils are all in phase with each other, and the coils being connected in series, the output electromotive force of the generator is equal to the sum of these individual electromotive forces.

Instead of thus connecting the first coil to the second, this second coil to the third, and so on, the first coil may be connected to the fourth, the fourth coil then to the seventh, the seventh to

[1] See M. LATOUR, *loc. cit.*, and M. Latour's U. S Patent 1234914.

the tenth, and so on, leaving out two coils out of every three. The output winding thus obtained around one stator circumference then comprises one-third of the total number of coils, and will yield an output electromotive force equal to one-third the original electromotive force, but having the same frequency as before.

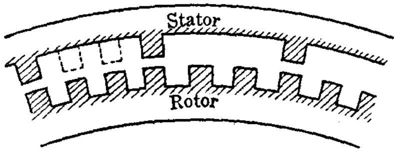


FIG. 94.

The intermediate coils then being not used, may be omitted altogether, as well as the stator teeth on which they were wound, so that the machine assumes the form shown in Fig. 94, which comprises the same rotor as before (Fig. 91) and a stator having only one-third the original number of teeth. The space between successive stator teeth is correspondingly greater than before since the ratio of tooth width to linear tooth pitch instead of being 1:2 as in the original machine, is now equal to one-third this ratio, that is, 1:6. This makes it possible to increase the number of teeth which may be used with good efficiency, and to raise the frequency output of the generator.

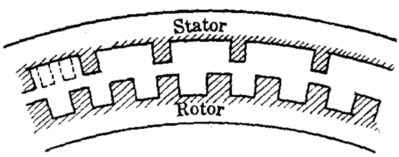


FIG. 95.

The same remarks may be applied to the alternate pole machine of Fig. 93 using a zigzag winding, which is transformed into the machine of Fig. 95 in the same manner as just explained.

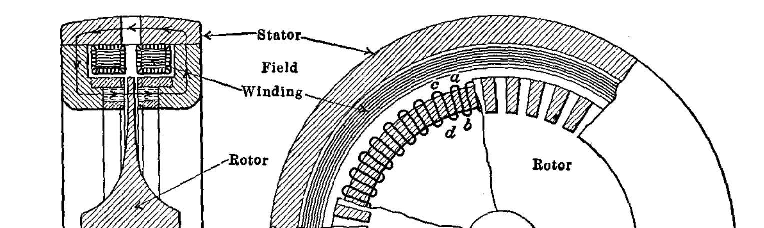


FIG. 96.

As an example of actual alternator construction, Fig. 96 represents schematically a small Alexanderson alternator.[1] The

[1] A more recent design is described by E. F. W. Alexanderson, *Trans. Am. Inst. Elec. Eng.*, pp. 1269–1285, 1919.

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 and reduce windage losses. The stator consists of two grooved steel rims mounted on either side of the rotor, as shown in the figure. 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 and forming a zigzag winding. 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 precautions 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 operating speed, any frequency variation resulting in a considerable reduction of the antenna current, the antenna circuit being then no longer tuned to the frequency of the generated electromotive force. Special quick-acting speed-regulating devices are used for this purpose.

The Oscillating Arc.—While the high-frequency alternator, as described above, transforms mechanical energy into high-frequency alternating-current electrical energy, the oscillating arc affords means for producing continuous waves in a circuit by purely electrical methods, transforming direct current into a high-frequency alternating current. It is but a special case of the use of negative resistance, which is studied in greater detail in a later chapter in connection with the three-electrode vacuum-tube oscillator. The present discussion is intended simply to describe in a general way the functioning of the arc when producing undamped oscillations.

When two carbon electrodes are connected to the terminals of a direct-current generator and an arc established between them, the electric current flowing in the circuit passes through the gap between the electrodes. This gap is filled with highly heated vapors which constitute the actual arc or flame, and which, owing to their high temperature, are strongly ionized and conducting. The degree of ionization and conduction of the arc depends, to a great extent, upon the intensity of the current flowing through the arc, since this current generates the heat at the electrodes. Conversely, the current which flows through the arc depends upon the degree of ionization of the gases. Thus, if these are ionized only to a small extent, the electrical resistance of the arc is great, and a high voltage is required between the electrodes to maintain the current across the arc. On the contrary, if the arc is strongly ionized, even a small potential difference between the electrodes will produce a large current through the arc.

The relation between the current and voltage across the arc may hence be represented by a curve of the shape shown in Fig. 97, showing a large current for a small voltage and a small current for a large voltage. If, then, the current in the circuit be alternately increased and decreased between two values corresponding to points N and P of the curve, the voltage will vary in opposite manner, that is, respectively decrease and increase. In other words, the ratio dV/dI of corresponding voltage and current variations is negative, which is often expressed in saying that the alternating-current resistance of the arc is negative (since the resistance R of a conductor is precisely equal to the ratio dV/dI of corresponding voltage and current variations).

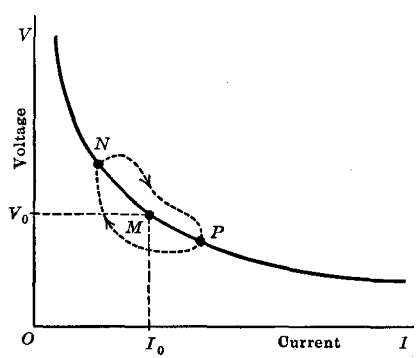


FIG. 97.

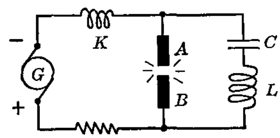


FIG. 98.

Now consider (Fig. 98) a constant voltage generator G connected to the two electrodes A and B of a carbon arc through a choke coil K, and a condenser C connected in series with an inductance coil L across the carbon arc AB. To begin with, assume this shunt oscillatory circuit LC to be disconnected from the arc, with the condenser C not charged and its plates at the same potential. Upon being connected to the generator, the potential difference across the arc will find itself applied to the condenser through the coil L, and the condenser will begin to charge. In other words, a charging current will flow into the condenser. And since the total current supplied by the generator G remains constant, due to the presence of the choke coil K, this flow of current for charging C produces a corresponding decrease in the current flowing through the arc. But referring to the characteristic curve of the arc, a decrease of current through the arc is accompanied by a rise in the potential difference between the electrodes and, therefore, across the circuit branch LC. This further increases the charging rate of the condenser. When the condenser is fully charged, the current flow in the circuit branch LC stops and the current in the arc increases, since the generator current is no longer divided between the two circuit branches. But this increase in the arc current results in a decrease of the potential difference at the arc electrodes, and therefore also across the circuit LC. The condenser thus begins discharging through the arc, increasing the current through the latter and further reducing the electromotive force across the electrodes. When the condenser is fully discharged, conditions are returned to the original state, and the process repeats itself.

These conditions may be followed on the characteristic curve of Fig. 97, which, as explained, represents the corresponding values of arc current and voltage for *steady* conditions, that is, when the arc is not oscillating, as is the case when the circuit LC is disconnected. It is therefore called the *static characteristic* of the arc. Thus, let V_0 and I_0 be the values of voltage and current before the circuit LC is connected, M being the operating point. Then, when the circuit is connected to the arc, as explained above, the arc current decreases during the charging period and the operating point moves from M to 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 thus oscillates between points N and P of the curve.

However, due to the heavy current carried by the arc around the end of the discharge period, when the operating point is at P, the carbon electrodes become quite hot and the space between them is of high conductivity. Due to the great frequency of the oscillations, the arc electrodes do not have time to cool down to their normal temperature when the current decreases, so that the resistance of the arc will be slightly less than normal, and the arc voltage will be consequently smaller. In other words, the operating point in its motion from P to N describes a curve which is *below* the 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, so-called *dynamic characteristic curve,* shown in dotted lines in Fig. 97.

An oscillation following a dynamic characteristic curve, as the one just described, 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 generated oscillations are very nearly sinusoidal.

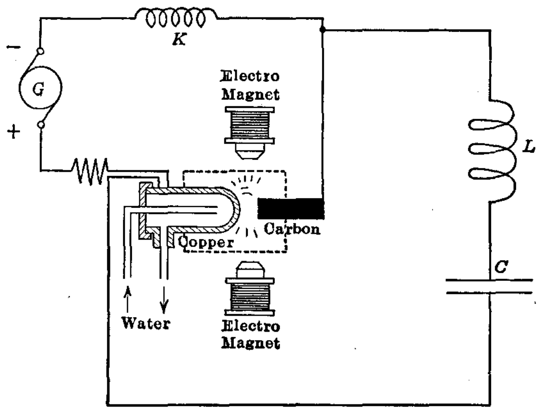


FIG. 99.

In order to obtain oscillations of great amplitude and high frequencies, it is necessary to use an arc having a large negative alternating-current resistance, that is, having a steep characteristic curve. This is obtained only with an arc which is very unstable, such as the Poulsen arc, in which the positive electrode is made of a copper jacket (Fig. 99) 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 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, and the dynamic characteristic of the arc under such conditions is shown in Fig. 100.

Finally, a third kind of oscillation 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 in 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.

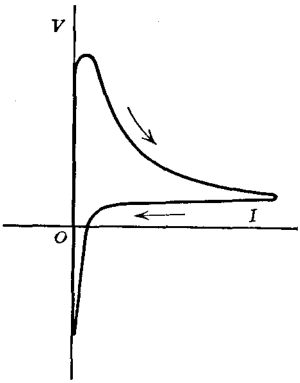


FIG. 100.

CONTINUOUS-WAVE RADIO TELEGRAPH TRANSMITTING CIRCUITS

High-frequency Alternator Circuits.—When a high-frequency alternator is used for energizing the antenna circuit, this circuit may be connected to the machine either directly, or through the medium of transformers. These are particularly useful when coupling several alternators in series or in parallel to the same antenna. The connection may also be made through iron-core frequency-multiplying transformers, a practice which is followed more frequently in Germany.

Signalling by means of alternator-excited circuits is always done in varying the amplitude of the generated current, while keeping its frequency constant. Alternators have generally a high power output. The transmitting key is seldom used for breaking the antenna circuit, unless special multiple-contact high-speed relays be used for this purpose. In some instances, the key is used for making and breaking the direct-current field circuit of the alternator, which of course carries a very much smaller current than the armature (output) circuit.

A more frequent practice is to detune the antenna when opening the key. For this purpose the key is made to short-circuit, when closed, a few turns of the antenna inductance, as shown in

Fig. 101. The antenna circuit being then accurately tuned to the alternator frequency, opening of the key detunes the circuit, which, ceasing to be in resonance, carries then only a negligibly small current.

The detuning of the antenna circuit may also be obtained by using an iron-core inductance in the antenna circuit. The same iron core carries a second winding connected in series with the key and with a direct-current generator. Closing the key then sends a direct current through this second winding, which partially saturates the iron core, changing its magnetic permeability and, hence, altering the value of the inductance and the tuning of the antenna. This principle is used in the Alexanderson magnetic amplifier, which is described more fully in the chapter on radio telephony.

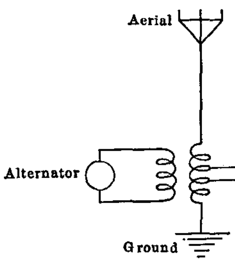


FIG. 101.

Finally, when the alternator is connected to the antenna through ferromagnetic frequency multipliers of even ratio, keying may be effected by making and breaking the direct current used for saturating the iron cores of the multipliers. This is equivalent to disconnecting and connecting the alternator and antenna circuits.

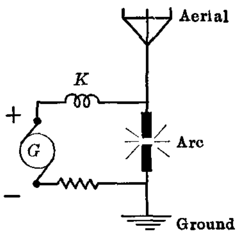


FIG. 102.

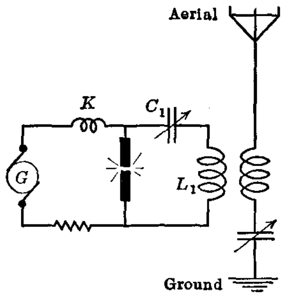


FIG. 103.

Oscillating-arc Circuits.—The oscillating arc may be connected directly in the antenna circuit, as shown in Fig. 102. It may also be made to generate undamped oscillations in an intermediate tuned circuit L_1C_1 coupled to the similarly tuned antenna, as shown in Fig. 103. In this case, care must be taken not to use so close a coupling as to produce a double-humped resonance curve, as this would lead to the same disturbing unstability conditions as described, for the three-electrode vacuum-tube oscillator, on pages 202 to 208.

In the case of an arc set, 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. The detuning method is therefore most commonly used. In this connection, it should be noted that the operation differs according to whether the simple circuit of Fig. 102 or the coupled circuit of Fig. 103 is used.

In case the simple circuit is used, detuning the antenna circuit simply alters the wave length of the generated oscillations without materially changing the amount of power radiated. The circuit thus operates at one frequency when the key is closed and at another frequency when it is open. These two waves are received by the receiving circuit, but this being tuned to the signalling wave, it will be energized more weakly by the spacing wave. Moreover, if the signals are received by the heterodyne method, described further below, the two waves produce two different notes in the telephone receivers and may be distinguished readily, and if the two frequencies are sufficiently different, the spacing wave may even be silenced completely in the receiver, as explained at the end of this chapter.

In case the coupled circuits of Fig. 103 are used, detuning of *either* circuit will produce a sharp decrease of *amplitude* in the antenna current, and the station will, as in the case of the alternator, radiate strongly when the key is closed and weakly (if at all) when it is open.

Signalling Speed.—Both in arc and alternator stations, the various oscillatory circuits and the antenna circuit are made of as low resistance as possible, in order to reduce energy losses. The smaller the ohmic resistance and damping of the circuits, however, the greater their time constant, that is, the longer will be the time required for the oscillations to build up in them when the key is closed and to die down when it is opened. Since commercial transmitting speeds are of the order of several hundred words per minute, this increased time constant of the circuits may actually limit the signalling speed. On the other hand, increasing the resistance (and damping) permits a greater signalling speed but reduces the electrical efficiency of the station.

These conflicting factors must therefore be taken into account, the best overall commercial efficiency being obtained by a proper balance of the electrical and traffic efficiency.[1]

CONTINUOUS-WAVE RECEPTION

The general principles given in the case of damped-wave reception are applicable to undamped or continuous waves, if account is taken of the different characteristics of the latter. The remarks on tuning and coupling of the receiving circuits are of course applicable to undamped waves, and it should be noted that tuning is considerably sharper than with damped waves (see the first paragraphs of this chapter). The resonance curves show a marked and sharply defined peak at resonance frequency, which thus greatly facilitates the elimination of interference from undamped wave sets of other frequencies.

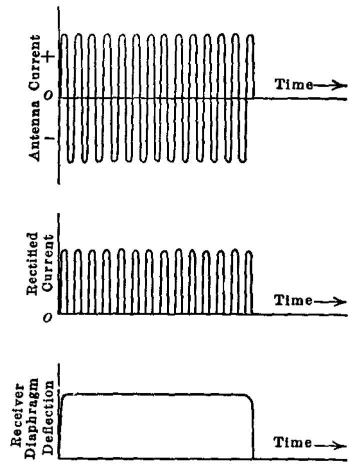


FIG. 104.

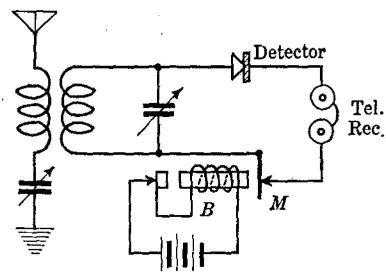


FIG. 105.

The ordinary detector methods cannot be applied in the case of undamped waves, as may be understood 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 oscillation is sent out, inducing in the receiving circuit a continuous high-frequency alternating current, as shown in the upper curve of Fig. 104. If this wave were rectified and made to flow through

[1] This should be contrasted with the case of radio telephony, where the modulation frequency (which here is equivalent to the signalling speed) is beyond the control of the radio engineer, being given by the frequency characteristics of the human voice. Most effective operation is then secured by making the antenna resistance as low as this modulation frequency permits, which is higher than in the case of telegraphy.

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 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, the thus interrupted current,

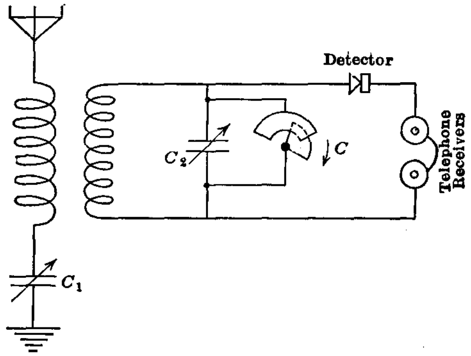


FIG. 106.

when flowing in the telephone circuit, then setting the diaphragm in audible vibration.

Thus, the damped-wave receiving circuit of Fig. 87 may be adapted to continuous-wave reception by connecting an interrupter at M (Fig. 105). This is a vibrating contact operated by a buzzer B adjusted to some audio frequency, say 500 to 1,000 cycles per second. The detected current flowing in the telephone receivers when signals are being received is then interrupted at this frequency and an audible note is heard.

An equivalent method, illustrated in Fig. 106, consists in periodically detuning the receiving circuit at some audio frequency. The receiving circuit is essentially the same as a damped-wave receiving circuit, exept 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 receiver being tuned by means of the condensers C_1 and C_2, the rotation of the condenser C periodically varies the capacity of the secondary oscillatory circuit, thus throwing it periodically into and out of tune. During the detuned intervals it will not respond, or only feebly, to the incoming oscillations. The oscillating current is thus periodically varied in amplitude, or "modulated," and when rectified produces a sound in the telephone receivers.

Still a different method is the tikker method, in which a motor-driven interrupter opens and closes the telephone receiver circuit at a radio frequency but little different from the frequency of the received signals. This method, which does not require any rectifying detector, is not described here, being closely related to the heterodyne method studied below. A vacuum-tube circuit operating on a similar principle as this tikker method is described in a later chapter.

All of the above methods have the common disadvantage that the received energy is periodically not made use of in the telephone receivers, the received current being periodically interrupted. This waste of energy is objectionable, since one of the main advantages of the undamped over the damped-wave method 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 utilizes all the energy received is much to be preferred. For these reasons, and others which will be pointed out later, the *heterodyne reception method* is one of the best and most widely used reception methods for undamped waves.

Heterodyne Reception.—The heterodyne method consists essentially in 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. The frequency of these beats is equal to the difference of the frequencies of the component currents.

Thus, consider the circuit of Fig. 107, 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. 108, 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 sum (or difference) of the corresponding instantaneous values of the component currents. This resultant current is shown by the lower curve of the figure, and is seen to be an alternating current of periodically increasing and decreasing amplitude. The maximum amplitude is equal to the sum of the maximum amplitudes of the component currents, and occurs when the two currents pass together through their positive or negative maxima. And the minimum amplitude is equal to the difference of the amplitudes of the component currents, and occurs when these currents pass together through maxima of opposite polarities.

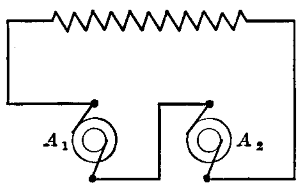


FIG. 107.

The fact, stated above, that the frequency of the beats (that is, the number of beats per second) is equal to the difference of the component frequencies may readily be found mathematically. Thus let the two component currents be represented by the expressions

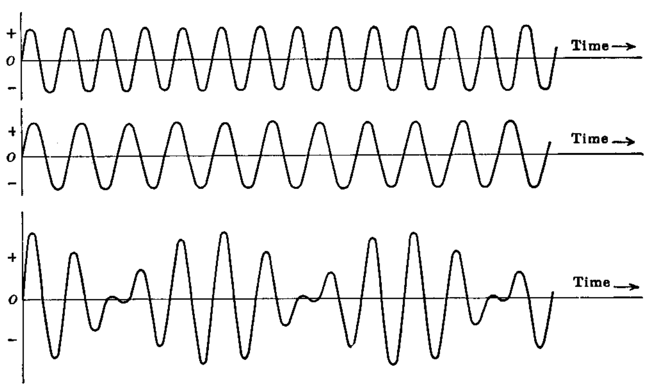


FIG. 108.

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

The resultant current i is then at each instant

$$i = i_1 + i_2,$$

which may easily be put under the form

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

which represents the beat current as described above, and in which the beat frequency $(f_1 - f_2)$ appears in evidence.

Now, applying these principles to the reception of undamped waves, consider the receiving circuit of Fig. 109. The receiving

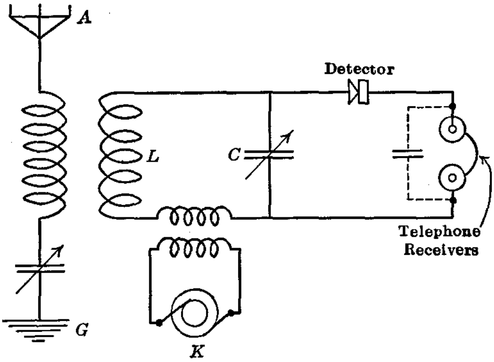


FIG. 109.

antenna circuit AG being tuned to the incoming waves, these induce a current of their own frequency f_1 in the closed oscillatory circuit LC. This current may be represented by the upper curve of Fig. 108. As explained previously, this current when rectified by the detector would not produce a readable sound in the tele-

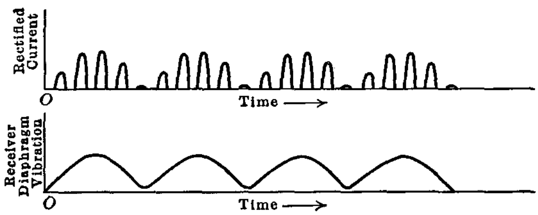


FIG. 110.

phone receivers. A high-frequency alternator K is therefore connected or coupled to the circuit LC and produces in this an alternating current of frequency f_2, adjusted to be slightly different from the frequency f_1, of the incoming signals. This locally generated current may be represented by the second curve of Fig. 108. As just shown, these two currents actually produce a current of variable amplitude (beat current) in the circuit *LC*, as shown by the lower curve (Fig. 108) which, when rectified by the detector and led through the telephone receivers then becomes of the form shown in Fig. 110, and produces vibrations of the receiver diaphragm at a frequency equal to $f_1 - f_2$.

The circuit of Fig. 109 may be modified in many different ways. Thus, the local generator circuit may be coupled to the closed oscillator circuit *LC* through the antenna circuit instead of directly. Another method consists in using a special double-winding telephone receiver, one winding carrying the received current, while the other carries the locally generated high-frequency current.

Advantages of Heterodyne Reception.—A first feature of the heterodyne method of reception is that the note heard in the telephone receivers, being dependent upon the frequency of the current generated at the receiving station, may be adjusted by the receiving operator to best suit the local conditions.[1] Thus, for instance, the beat frequency may be adjusted to be equal to the natural frequency of vibration of the telephone receiver diaphragm. This mechanical resonance greatly reinforces the sound due to the received signals, contrasting these sharply from other interfering undesired signals. This note, in view of its great constancy, is of the nature of a clear, musical whistle.

Another advantage of the possibility of adjusting the pitch of the signals at the receiving station is, that it assists in overcoming interference from other signals to a remarkable degree. Thus, suppose the signals to be received to have 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. The sharp tuning afforded by undamped waves already minimizes the interference. Now, if the locally generated frequency is 101,000 cycles, the notes heard in the receivers will be respectively 1,000 and 500 cycles per second for the desired and the interference signals. These two widely different notes are 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

[1] This is of course impossible with damped waves, since the audio frequency is then determined wholly by the spark frequency at the transmitting station.

interfering signals, which are thus entirely eliminated. This feature becomes of considerably greater importance in the superheterodyne receiver, where the beats are of superaudible (radio) frequency, as will be studied later in greater detail.

Finally, a third important property of the heterodyne method is that it provides an *amplification* of the received signals.[1] This may be explained roughly by the fact that the locally generated energy is added to that received from the transmitting station. 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

$$i = i_1 + i_2.$$

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

$$i^2 = (i_1 + i_2)^2 = i_1^2 + i_2^2 + 2i_1i_2.$$

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 upon 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.

It should not be thought, however, that as great amplification as desired may be obtained by simply increasing the locally generated current intensity. The characteristics of the detector play an important part in this connection. Thus, consider an undamped wave signal as represented by curve 1 (Fig. 111). This same signal, combined with a locally greater high-frequency alternating current of slightly different frequency, is shown in curves 2, 3 and 4, which correspond, respectively, to a local current of smaller, equal, and greater intensity than the received signal current. These curves show at once that the amplitude variation of the combined current increases with the

[1] Heterodyne reception nowadays is effected almost universally by means of vacuum tubes, which of course provide considerable amplification of the signals. But the amplification described here and due to the heterodyne method is independent from the amplifying properties of the tubes, and inherent to the heterodyne method.

the intensity of the locally generated current until the latter is made equal to the signal current intensity (curve 3), when the amplitude of the combined current varies periodically between zero and twice the amplitude of either signal or local current. A further increase of intensity of the local current then does not change this amplitude variation or modulation, but this same modulation occurs about a higher average current intensity, and hence does not come down to zero (curve 4).

Since it is this beat modulation which produces the sound in the telephone receivers, the local current must obviously be made at least equal to the signal current, since this corresponds

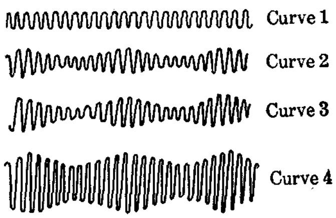


Fig. 111.

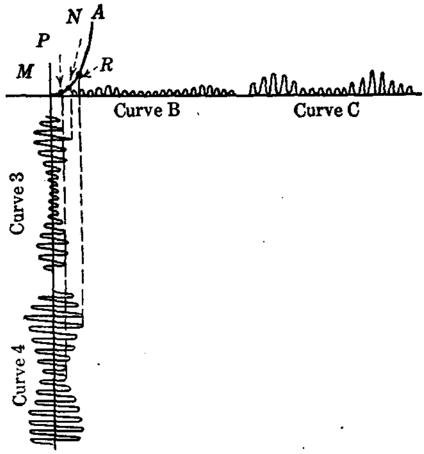


Fig. 112.

to maximum modulation amplitude. The question arises then as to whether the local current should be made greater than the signal current, or simply equal to it, since, as just shown, the modulation amplitude is the same in both cases. The answer depends essentially upon the shape of the characteristic curve of the detector employed. If, as is often the case, this curve has the general shape A shown in Fig. 112, the greatest signal is given by the combined current of curve 4 (Fig. 111), obtained with a local current greater than the signal current. Thus, curves B and C (Fig. 112) respectively represent the "detected" currents yielded by currents 3 and 4 (Fig. 111). In the former case, the slope of the curve portion MN being much smaller than that of the portion PR, which in the latter case serves for the rectification of the modulated part of the high frequency current, the resulting pulsations of the rectified current are correspondingly smaller, and a weaker signal is produced in the telephone receivers.

CHAPTER VI

THE THREE-ELECTRODE VACUUM TUBE

GENERAL PROPERTIES

Electron Emission by Hot Bodies.—As shown in Chap. I, the raising of the temperature of a material increases the agitation of its constituent atoms or molecules. In the case of a metal or other substance containing free electrons, this is accompanied by an ejection or emission of free electrons from the heated body out into the surrounding space, which becomes noticeable at about dull red heat, and increases rapidly with temperature.

If a body is maintained at some temperature T where it thus emits electrons, it will, through the loss of these negative charges, become positively charged, and therefore tend to attract back the emitted electrons. A condition of equilibrium is reached when the attraction of the body on the emitted electrons exactly counterbalances the forces producing the electron emission, this state of equilibrium being, furthermore, characteristic of the substance and temperature. If the temperature is raised, the atomic agitation of the substance and the velocity of the emitted electrons increase correspondingly, the free electrons overcome the attraction of the heated body in greater number, and a new state of equilibrium is established. The number N of electrons emitted at a temperature T (expressed in absolute degrees) is

$$N = AT^c e^{-b/T},$$

where e is the base of the natural system of logarithms, and A, b, and c are constants depending upon the chemical nature, shape, and other characteristics of the body.

This emission of electrons by a heated body makes the surrounding space a good conductor of electricity, which is, by definition, a substance or medium containing free electrons. If, then, a positively charged body is placed near the heated substance emitting the electrons, these will travel from the heated substance to the positively charged body, establishing an electric current through the intervening space.

The Two-electrode Vacuum Tube.—Consider a metal filament F, sealed in an evacuated glass bulb (Fig. 113), containing also metal plate P. When the filament is heated by the current of a battery A, it emits 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, for instance, by connecting a battery B between the plate and filament, there will be established a continuous flow of electrons from the filament to the plate, and therefore an electric current in the circuit $FBPF$.

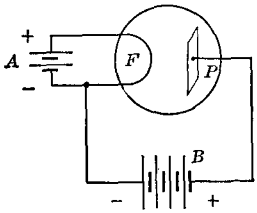


FIG. 113.

Keeping the filament temperature constant, the number N of electrons emitted by the filament per unit time will be correspondingly constant, and the number n of electrons attracted by the plate will depend upon the plate potential with respect to the filament. Thus, starting with a potential difference of zero volts

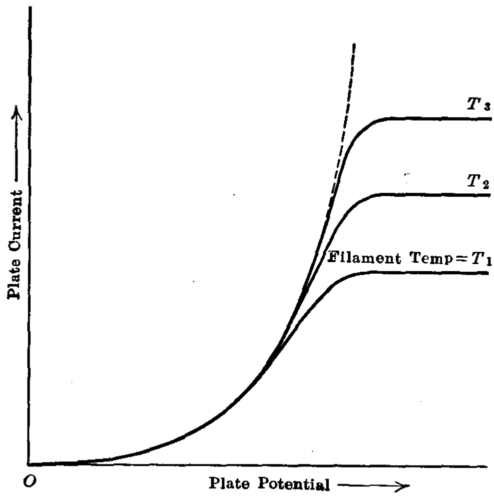


FIG. 114.

between plate and filament, measured between the *negative* filament end and the plate, 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 time are attracted by the plate, so that further increase of plate potential will produce no increase of plate current. This maximum current, beyond which there is no increase for increased plate potential, is called the *saturation current* of the tube for the particular filament temperature (or current) used. This is represented by the curve of Fig. 114.

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

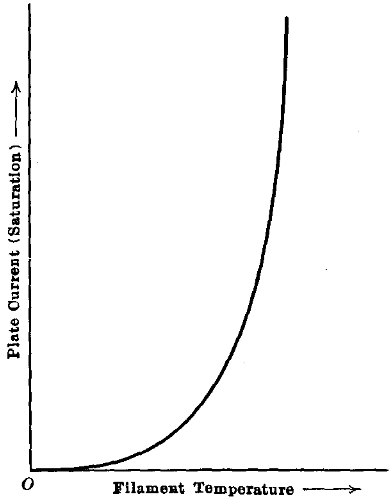


FIG. 115.

The saturation current is thus a measure of the number of electrons emitted by the filament, and may therefore be expressed in function of filament temperature by a relation of a form similar to that given above for the electron emission by a hot body. Thus, the saturation current I for a given filament temperature T in a tube having a perfect vacuum is given by the equation

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

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. 115.

Now, consider the tube with the filament cold (battery A disconnected) and the plate maintained at some constant positive 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, assuming the space within the bulb to be a perfect dielectric. In case gas is present within the bulb, ionization may take place, as described later. If the filament is then gradually heated by gradually increasing the current from the battery A, it will emit electrons at an increasing rate, and the plate current will increase, being at every instant equal to the above expression. However, due to the fact that electrons are continually streaming from the filament to the plate, there are at every instant a number of electrons in the space between the filament and the plate. These all move toward the plate and are absorbed by it, while 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 upon the rate of absorption or attraction by the plate and upon the rate of emission by the filament. The steady increase of the filament temperature increases

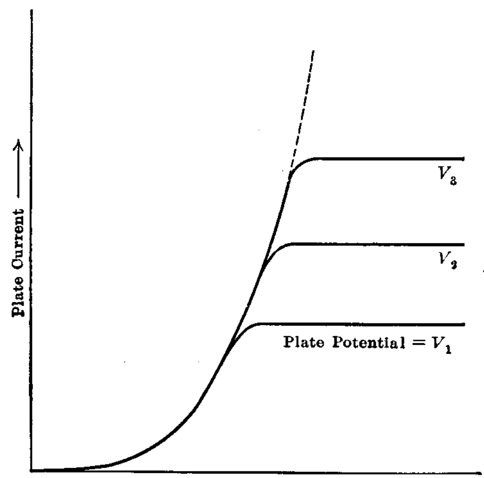


FIG. 116.

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 filament produces a negative *space charge*, the effect of which upon the electrons leaving the filament is opposite to that of the positive charge on the plate.

As the filament current is increased far enough, an equilibrium is reached between the opposing effects of the plate and space charges, and any further increase of filament temperature produces no increase of plate current, any additional electrons in the space of the tube making the negative space charge overbalance the positive charge on the plate and repelling the emitted electrons back toward the filament. In order to obtain a greater plate current, it is then 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. 116.

The maximum plate current which can be obtained for a given value V of plate voltage depends essentially upon the shape, dimensions, and spacing of the two electrodes. For a tube having perfect vacuum, and fitted with a filament 1 cm. long placed along the axis of a cylindrical metal plate of radius r, this current i if given by the relation

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

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 centimetres, and a a constant.

Effect of Gas on the Plate Current.—If the tube contains gas, even in very small amount, the above conclusions no longer apply strictly, the process being modified when the plate potential is such that ionization takes place. Thus, consider the solid line curve of Fig. 117, 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 filament to plate, particularly after the current has reached its saturation value. Now if the tube contains traces 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 of current in the circuit, as shown by the dotted line curve in Fig. 117.

The two-electrode tube described above finds application as an alternating-current rectifier, current flowing through the tube when the plate is positive with respect to the filament, and not when it is negative. This rectifying property, however, is destroyed if the plate is heated to incandescence, as occurs when too high a positive potential is applied to it, increasing the velocity of the electrons and, consequently, the heat developed at the plate when they are stopped by it.

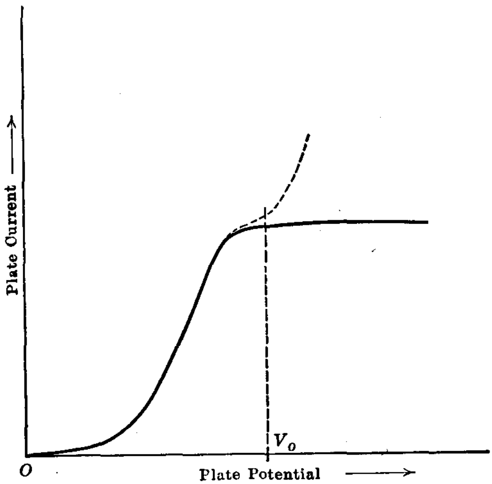


FIG. 117.

THE THREE-ELECTRODE VACUUM TUBE

Other factors than the plate potential and filament temperature may affect the current intensity in the plate circuit of an electron discharge tube, for the state of motion of the particles of negative electricity or electrons which move, within the tube, from filament to plate, is readily influenced by the action of a magnetic or electrostatic field, as explained in Chap. I. A most widely used method of control of the electrons stream utilizes an electrostatic field established within the tube by means of an auxiliary electrode, generally inserted between the filament and plate.

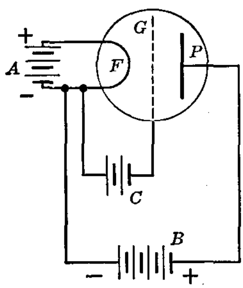


FIG. 118.

This third electrode G (Fig. 118) may be a perforated plate, 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 becomes possible to accelerate or decelerate the electrons moving from filament to plate, to neutralize or increase at will the effect of the space charge, and thus to control the plate-current intensity without changing the plate voltage 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 operating in the plate circuit considerable, yet the energy required to charge the grid to the desired potential is extremely small, due to the small electrostatic capacity of the grid with respect to the filament.

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 filament, the grid not being connected to any circuit.

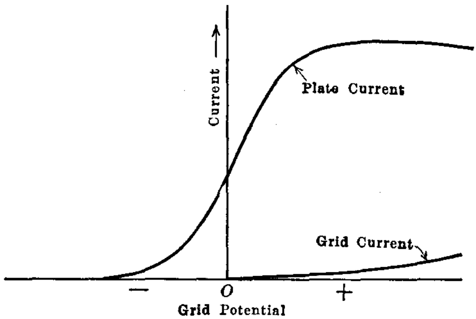


FIG. 119.

The tube then operates like a two-electrode tube, and a stream of electrons flows from the filament F heated by the battery A to the plate P, under the influence of the battery B (Fig. 118). This plate current is limited, as explained for the two-electrode tube, by the space charge due to the electrons present in the space between the plate and filament.

Now, if the grid is given a *negative* potential with respect to the filament, for instance by connecting a battery C with its positive terminal to the filament and its negative terminal to the grid, the negative charge on the grid will add its effect to that of the space charge, repelling the electrons emitted by the filament back toward the latter, and more or less neutralizing the attraction of the electrons by the positively charged plate. The result is a decrease in the plate-current intensity, which decrease will be greater the more negatively the grid is charged. The grid may even be made sufficiently negative to stop entirely the flow of electrons from filament to plate.

If, on the contrary, the grid is charged *positively*, the negative space charge due to the electrons in the tube is neutralized correspondingly, and the plate current increases. For increasing positive grid potentials, the plate-current intensity increases, until the saturation current is reached, corresponding to the existing filament temperature.

This dependence of the plate current on the grid potential is represented by the "plate current" curve of Fig. 119, the filament temperature and plate voltage being kept constant throughout. This curve is called the *static characteristic curve*

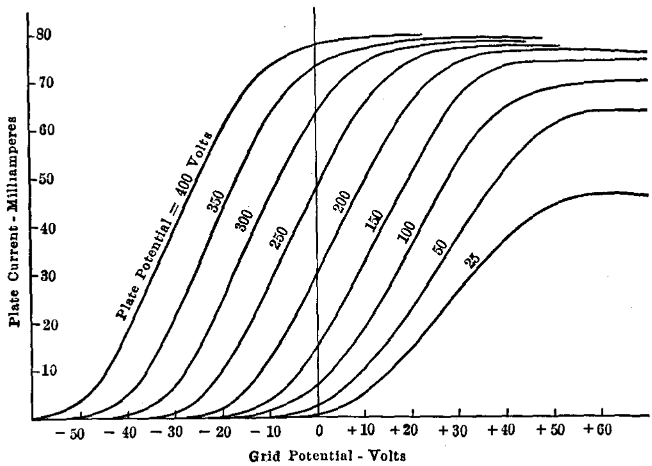


FIG. 120.

of the tube for the particular plate voltage and filament temperature (or current) used. Assuming the latter to be constant, a family of such curves is obtained corresponding respectively to various values of plate voltage, as shown in Fig. 120. The higher the plate potential, the more the curve is shifted to the left, without substantially altering its shape, particularly for the higher values of plate potential.

These same relations may be shown as in Fig. 121 by plotting the plate current in function of the plate voltage for various constant values of grid voltage.

It should be noted that, when the grid is positive, it attracts a few electrons to itself, giving rise to a current in the grid circuit *FCGF* (Fig. 118) which is represented by the "grid current" curve of Fig. 119. In the useful operating range of the tube, this grid current is comparatively very small, ranging from zero to about one-one hundredth or one-tenth of the plate current. It may be entirely avoided through a suitable choice of the operating voltages. Thus, if the plate potential is high, a large negative grid potential is required to stop the flow of plate current. A slightly smaller grid potential permits the plate current to flow, and a still smaller negative grid potential allows the plate current to reach its maximum saturation intensity. Over this whole con-

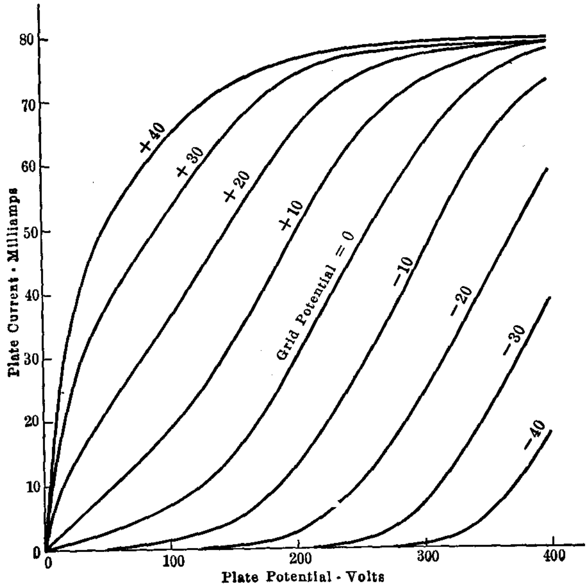


FIG. 121.

trol range the grid potential remains negative, and consequently the grid does not attract any electrons to itself, and no current flows in the grid circuit.

Quantitative Expression of Vacuum-tube Properties.—As just explained, the intensity of the electron stream flowing from filament to plate in a three-electrode vacuum tube may be controlled by varying the potential of the grid with respect to the filament, under conditions for which the grid current is negligibly small or even entirely inexistent. Only the plate current will therefore be considered here, the intensity of which is now to be expressed quantitatively in function of the potentials applied respectively to the grid and plate of the tube, and in function also of certain factors or parameters characterizing the design of the particular tube considered.

Amplification Factor.—Upon escaping from the heated filament, the free electrons emitted by it find themselves under the action of the electrostatic field created about the filament by the charges of the grid and plate, the plate being given a positive potential with respect to the filament by a battery B, and the grid a positive or negative potential by means of a battery C, as described in connection with Fig. 118. This electrostatic field, which drives the electrons onto the plate, is the resultant or combination of the fields set up by the plate and grid charges respectively. It may obviously be established through a variety of possible combinations of grid and plate potentials, and its action on the electrons leaving the filament will be the same irrespective of what particular combination is used.

Thus, if the plate potential is lowered, correspondingly decreasing the plate current, the original field, and also the original plate current intensity may be restored by raising the grid potential—that is, making the grid more positive or less negative. Or else, if the plate voltage E_p is increased, or decreased, by a small amount dE_p, causing, as in the case of a two-electrode tube, a variation dI_p of the plate current, the same plate-current variation may be obtained by keeping the plate voltage constant, and varying, in the same direction, the grid voltage E_g by an amount dE_g, producing the same electrostatic field variation about the filament as the plate potential change dE_p.

The ratio

$$k = \frac{dE_p}{dE_g}$$

of the plate and grid voltage variations which thus produce a same plate-current variation, being equal to the ratio of the corresponding electrostatic field variations, is also equal to the ratio

$$k = \frac{C_g}{C_p},$$

where C_p and C_g are, respectively, the electrostatic capacities of the condensers constituted by the filament and plate, and by the filament and grid. This ratio k thus depends upon the internal structure of the tube only and not upon its operating grid and plate voltages, and is therefore a constant, called the *voltage ampli-*

fication constant or factor of the tube. It expresses the fact that the plate-current variation caused by a grid-voltage variation dE_g is the same as that which would be caused by a plate-voltage variation equal to $k \cdot dE_g$.

It also explains the fact that the grid-voltage, plate-current characteristic curves corresponding to various constant values of plate voltage are substantially parallel to each other, as shown in Fig. 120.

Since a grid potential E_g has the same effect on the plate current as a plate potential kE_g, the current flowing in the three-electrode tube is the same as if the tube were a simple two-electrode tube having a plate potential equal to

$$E = E_p + kE_g,$$

where E_p and E_g are respectively the voltages actually applied to the plate and grid of the tube. The plate current I_p is thus expressed by a relation of the same form as for the two-electrode tube,

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

and, as in the case of the two-electrode tube, no current will flow in the plate circuit if the potential $E_p + kE_g$ is equal to zero or negative,

$$E_p + kE_g = o \text{ or } < o,$$

or

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

The amplification factor k of a three-electrode vacuum tube, as defined above, varies inversely as the spacing between the grid wires, for the closer the mesh of the grid, the greater will be the portion of the electrostatic field of the plate shielded or screened off by the grid. It varies also directly as the ratio of the plate-filament and grid-filament distances, as may be understood from the fact that the closer the grid is placed to the filament, the smaller the potential required to set up a field around the filament equivalent to the field set up about it by the plate. Thus, for obtaining a large amplification factor, it is necessary to use a fine grid mounted at a small distance from the filament, as compared to the distance between the plate and the filament. Voltage amplification factors of 8 to 12 are commonly used in radio tubes, but factors as high as 17 or 20 have been obtained in certain types of tubes.

The amplification factor k is actually not quite constant for a given tube, but varies somewhat with the plate and grid potentials, 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. This shift of the shape and position of the space charge correspondingly alters the relative effects of the plate and grid potentials on the electron flow in the tube.

Also, the filament being heated by a current flowing through it, a potential gradient exists along the filament, and its various points are therefore at different potentials with respect to the grid and plate respectively. It follows that the action of the grid potential on the electron current flow is more gradual than if the entire filament was at a same potential. Thus, if the grid is connected directly to the negative end of the filament, with respect to which it is therefore at zero potential, it will be at a potential of $-V$ with respect to the positive end of the filament, V being the potential difference between the filament terminals. This negative grid potential of $-V$ may, in some cases, be sufficient to stop the electron current from flowing from the filament positive end to the plate. But there being no potential difference beween the grid and the positive end of the filament, an electron current may still flow between it and the plate, which will be stopped only for a greater negative grid potential.

The amplification factor may, however, be considered as substantially constant over the comparatively straight portions of the characteristic curves.

Internal Resistance.—In order to obtain a current flow across an electron discharge tube, it is necessary to maintain the plate at a positive potential with respect to the filament, which may be done by connecting a generator or battery B (Fig. 118) between the filament and plate. Without this battery, the positive charge on the plate would soon become neutralized by the negative charges of the electrons collected by it, and the electron flow would cease.

A current I_p, thus flowing through the tube for a plate voltage E_p, the ratio

$$R = \frac{E_p}{I_p}$$

is the internal plate-to-filament resistance of the tube, according to the definition of resistance given in Chap. I. Indeed this internal tube resistance may be considered as of similar nature as the ohmic resistance of a metallic conductor. As explained in Chap. I, the free electrons moving between the atoms of the metal and constituting the flow of electricity in the conductor, accelerate under the effect of the applied electromotive force until they collide with an atom, when they lose the kinetic energy acquired during the acceleration period, transferring this energy to the atom, which enters into vibration, raising the temperature of the conductor.

In the present case of conduction through a vacuum tube, the "free path" of the electrons is the space between filament and plate. Over this entire path, the electrons accelerate, sometimes to speeds of several thousand miles per second, acquiring considerable kinetic energy. Upon colliding with the plate, they surrender most of their kinetic energy to the atoms constituting the plate, and raise the temperature of the latter,[1] the plate sometimes becoming red hot. The energy thus transformed into heat is, as in the case of metallic conduction, supplied by the battery B, and is equal to

$$W = E_pI_p = RI_p^2 = \frac{E_p^2}{R},$$

as in Chap. I.

The internal resistance being defined as the ratio of plate voltage and plate current, and the plate current being a function of the grid potential, it follows that the internal resistance of the tube is also a function of the grid potential. Keeping the plate voltage constant, it is seen from the characteristic curves of Fig. 120 that the internal resistance, which is infinite for grid voltages for which the plate current is zero, decreases when the grid voltage is made less negative, permitting the flow of an increasingly stronger plate current, until, when the saturation current is reached, the resistance reaches and retains a minimum value.

Differential Resistance.—The internal resistance defined in the preceding section may be called also the "direct-current resistance" of the tube corresponding to given values of plate and grid voltages. In most radio applications of the three-electrode

[1] The balance of their kinetic energy is radiated in the form of a very short electromagnetic wave constituting an X-ray.

vacuum tube, the grid and plate voltages and the plate current are not constant but pulsate about some steady average values. The grid and plate voltages and the plate current may then be considered respectively as the resultant or combination of a constant (direct-current) component and a variable (alternating-current) component. These direct-current and alternating-current components behave independently from each other, and the alternating-current resistance of the tube has no particular relation to its direct-current resistance.

This alternating-current resistance is defined as the ratio

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

of corresponding plate-voltage and plate-current *changes,* these changes, furthermore, being supposed to be small, and the grid voltage remaining constant.

Instead of producing the plate-current variation dI_p by means of a plate voltage variation dE_p, the plate voltage may be kept constant and the plate current altered by means of a grid voltage variation dE_g. This, as known, will be equivalent to a plate-voltage variation $k \cdot dE_g$, so that the alternating-current resistance of the tube may be expressed by the relation

$$r = k\frac{dE_g}{dI_p}.$$

The alternating-current resistance of the tube is also called the differential plate resistance of the tube, being equal to the slope of the plate-current, plate-voltage characteristic curve at the point about which the small current and voltage variations considered take place. It is substantially constant over the straight portions of the curve, but increases around the upper and lower bends of the curve and becomes infinite when the plate current is equal to zero or to the saturation current.

Mutual Conductance.—In addition to the amplification factor

$$k = \frac{dE_p}{dE_g},$$

which is the ratio of the plate- and grid-voltage changes producing a same plate-current variation, and the differential plate resistance

$$r = \frac{dE_p}{dI_p},$$

which is the ratio of corresponding plate voltage and current changes, a third factor is required to express fully in mathematical terms, the interrelation between grid voltage, plate voltage, and plate current, which is the ratio

$$g = \frac{dI_p}{dE_g}$$

of corresponding changes of plate current and grid voltage, the plate voltage being kept constant. This ratio, called the *mutual conductance* of the tube, is equal to the slope of the grid-voltage, plate-current characteristic curve at the point considered. The greater its value, that is, the steeper the curve, the greater the plate-current change produced by a given grid-voltage variation. As may be seen from Fig. 119, the mutual conductance is a maximum at the point of inflection of the curve, corresponding to a plate current approximately equal to one-half the saturation current. It is equal to zero when the plate current is zero or equal to the saturation current, which simply means that small grid-voltage variations then produce no plate-current variation.

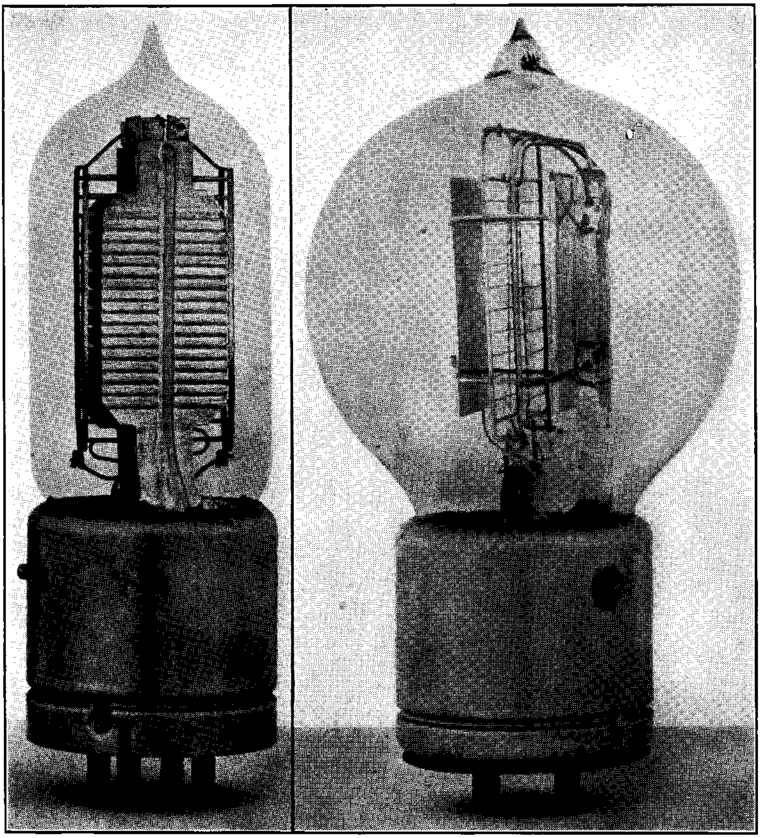
As results from the above discussion, the three factors k, r, and g are connected by the relation

$$r = \frac{k}{g}.$$

Measurement of Vacuum-tube Constants.—The three-electrode vacuum-tube constants k, r, and g defined above may be calculated from the static characteristic curves by computing the expressions defining these factors. They may also be measured directly, a few of the many possible methods being described below.

Amplification Factor.—The amplification factor may be measured by applying known voltages to the plate and grid of the tube and measuring the resulting plate current. Altering the plate voltage by a small amount, the grid voltage is changed until the plate current is restored to its original intensity. The ratio of the absolute values of the plate- and grid-voltage variations is then equal to the amplification factor.

Another so-called dynamic or alternating-current method may also be used. The filament of the tube being heated by a battery A (Fig. 122) the plate circuit is energized by a battery B, while the grid potential is obtained from a battery C. The batteries B and C are connected to the filament through the resistances



(A) (B)

Plate 1.—(*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* 142.)

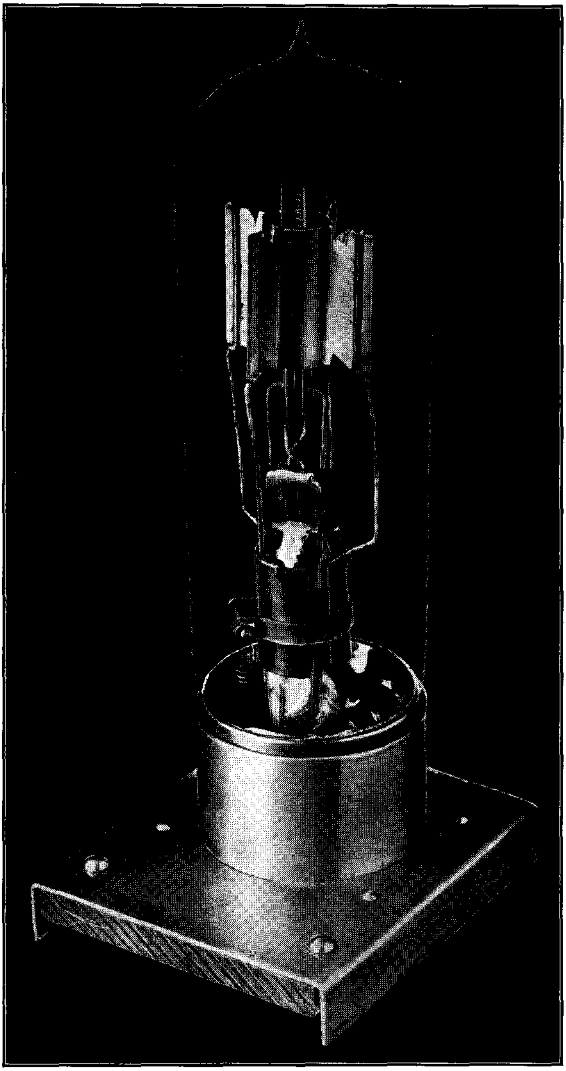


Plate 2.—Three-electrode vacuum tube used as high power oscillator and modulator. The plate has been forced downward to expose the grid and filament axially mounted. Note the cooling flanges on the plate. Signal Corps tube type VT–18.

(*Facing page* 143.)

r_1 and r_2, which may be adjusted by means of a sliding contact connected to the filament. An alternating potential difference being established between the ends of the circuit branch r_1r_2 (in the case of the figure, this is done by connecting these ends to an alternating-current generator through the medium of a transformer), the grid and plate voltages will vary about their original (direct-current) values, and these variations will be 180 deg. out of phase with each other. That is, when the plate voltage rises, the grid voltage decreases, and conversely. The effects of these voltage variations on the plate current are there-

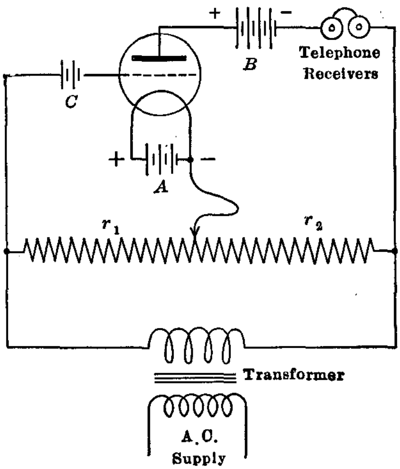


FIG. 122.

fore opposite. They neutralize each other exactly, leaving the plate current strictly constant when the sliding contact is so adjusted that

$$\frac{r_2}{r_1} = k,$$

which condition may be detected by means of a telephone receiver inserted in the plate circuit of the tube (provided the alternator frequency is within the limits of audibility), no sound being heard in the telephone when the resistances are adjusted to the above ratio.

Differential Internal Plate Resistance.—The differential plate resistance of a tube may be determined for any value of plate and grid potentials by computing the slope 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 dI_p/dE_g of the plate-current, grid-voltage characteristic, the resistance being then equal to $k.dE_g/dI_p$.

In the dynamic method, illustrated in Fig. 122, 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 of which are adjustable by means of a sliding contact connected to the bridge circuit, which comprises a telephone receiver. The differential resistance is determined for fixed plate and grid potentials derived from the

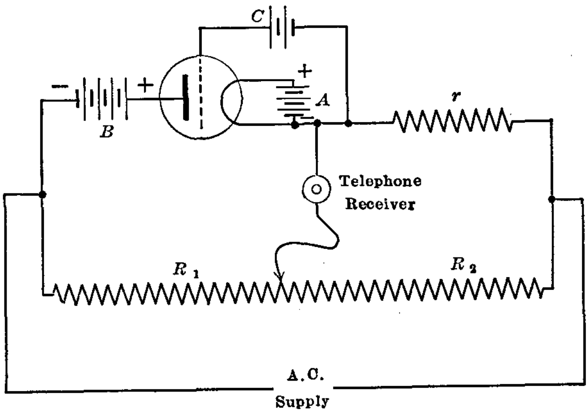


FIG. 123.

batteries B and C. The system being connected to a supply of alternating current of audible frequency, the bridge is balanced to make the telephone receiver silent, when the differential resistance of the tube is equal to

$$\frac{rR_1}{R_2}.$$

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 also be used, but are not described here. Or else, it may be calculated from the amplification factor and differential resistance, applying the relation

$$g = \frac{k}{r}.$$

Various Types of Vacuum-tube Constructions.—The actual construction of three-electrode vacuum tubes may take many different forms, depending upon the power to be handled by the tube, the particular use for which it is intended, etc. Only a few general remarks can be made here on this subject.

A very widely used type of construction is shown in Fig. 124, which is well suited for receiving tubes (which handle a small power only) and also for low-power transmitting tubes. As shown in the figure, a straight filament is mounted along the axis of the helical grid surrounding it, and of the cylindrical plate, which encloses both grid and filament. The three elements are all held on a common glass stem within the bulb, and the terminals (two for the filament, one for the grid, and one for the plate) are brought out of our contact studs mounted in the tube base.

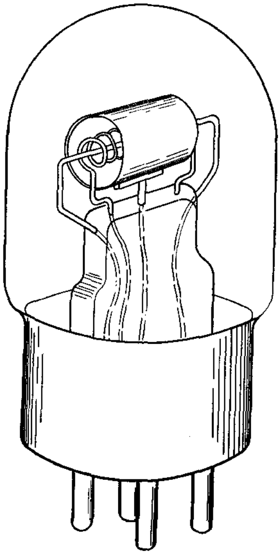


FIG. 124.

Receiving tubes of this type are generally designed for plate potentials of some 20, 40, 60 or 80 volts, and filament voltages of from 2 to 4 volts. The plate is made of sheet nickel or platinum, the grid of nickel or molybdenum wire. The filament, when made of tungsten, generally requires a current of the order of from ½ to 1 amp., and is operated at a bright yellow heat. It is often made of thoriated tungsten, which has the property of abundantly emitting electrons at a dull red heat, and thus requires a much smaller heating current, of the order of from 60 to 100 milliamp. The saturation plate current in this type ranges from about 10 to 50 milliamp. for the receiving tubes and from 100 to 200 milliamp. for the transmitting tubes.

The type of construction exemplified in Fig. 124, while both simple and rugged, has the disadvantage that the plate, grid, and filament terminal wires, being all brought out to the tube base, find themselves quite close to each other. This increases the electrostatic capacity between the tube electrodes, making the tube ill-suited for very short wave reception. It also prevents these tubes from being operated at high plate voltages, as required when large powers are to be handled by the tube, and voltages of 300 to 400 volts are seldom exceeded with this type of construction, on account of the possibility of spark-over between the terminal studs.

These drawbacks are avoided in the construction type of Fig. 125, in which the plate, grid, and filament are arranged in substantially the same manner as in Fig. 124, but with only the two filament terminals brought out to the tube socket. The grid and plate terminals are here brought out to two widely separated points of the glass bulb, constituting the so-called "horns" of

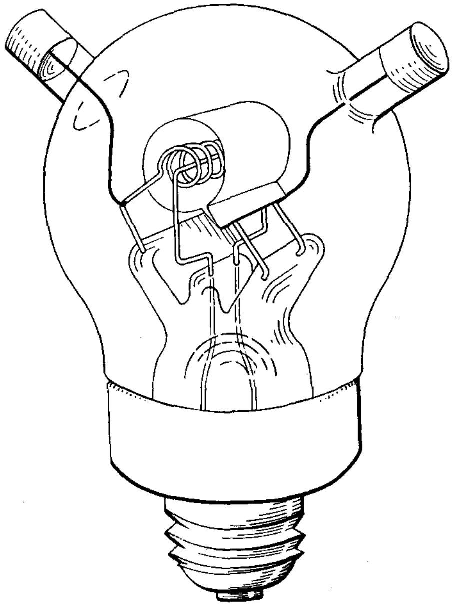


FIG. 125.

the tube. This type of tube, made in small size, is well adapted to the reception of short waves. In larger sizes, it may be used for transmitting purposes with several hundred volts on the plate, sometimes as many as 1,000 volts.

For tubes of very large power ratings, entirely different construction principles are made use of. The reason is that in the preceding types, when the tube electrodes are in a vacuum, the heat developed at the plate by the electron bombardment can be dissipated only by radiation, that is, in comparatively small amounts. This in turn limits the voltage which may be used in the plate circuit and consequently the power handled by the tube.

The principle used here[1] consists essentially in using the plate electrode itself as the closed vessel containing the grid and filament, the plate being then cooled by means of a water circulation in a jacket surrounding it. The filament and grid connections are then brought out through a glass cap sealed over an opening in this "external plate." Of course, the tube being highly evacuated, the seal between the glass cap and metal tube must be perfectly air-tight, and special methods have had to be devised for accomplishing this.

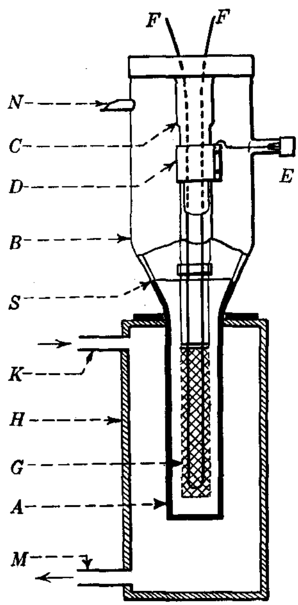


FIG. 126.

The actual construction is schematically represented in Fig. 126. The tube itself is constituted by the metal sleeve *A*, generally made of copper, the opening of which is filled with a large glass cap *B* sealed to the metal sleeve *A* at *S*. Inside the metal tube *A*, which serves as the plate of the tube, is the grid *G* made of metal gauze and forming a bag which surrounds the U-shaped filament. The filament terminals *F* are brought out through a glass stem *C*, which also serves to support the filament mechanically. The grid is held in place by a metal collar *D* fitted around the stem *C*, and to which is attached a connection wire coming out at *E* on the side of the glass cap, constituting the grid terminal of the tube. The nipple *N* serves to pump the air out of the tube when assembled, and is then sealed off in the usual way. When operating the tube, the cylindrical plate *A* is enclosed in a jacket or tank *H* filled with cold water, which is continually renewed through the pipes *K* and *M* in order to cool the tube.

Tubes of this type have been built for power ratings of from 5 to 100 kw. output. A fairly standard size, used in a number of countries both here and abroad, has a power output of 10 kw.

[1] See article by M. W. WILSON in *Elec. Comm.*, August 1922.

and operates with a plate voltage of 10,000 volts. The filament voltage is of the order of from 30 to 40 volts, the filament current being about 3 amp. The filament-to-plate electron current is about 5 to 8 amperes. These figures are given merely to indicate the order of magnitude of the electrical constants of these tubes.

The manufacture of three-electrode vacuum tubes requires that great care be taken to exclude all air and other gases from the tube glass walls and the metal of the electrodes. For this purpose, 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, much in the same manner as when making ordinary incandescent lamps for illumination purposes. Before exhausting the bulb, it is first heated in a special oven to a temperature of about 500°C., and then is pumped out while maintained at that temperature. This is done to drive the gases out of the walls of the tube. For a similar purpose, the filament is heated by means of an electric current sent through it, while 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 tube is finally sealed by melting off the fine glass tube connecting it to the pump.

For tubes enclosed in a glass bulb, like the first two types described above, particularly those employing a thoriated tungsten filament, the inner face of the glass bulb is, further, coated with a fine layer of magnesium, applied by vaporization during the process of manufacture, which prevents any gases remaining in the glass walls of the bulb from escaping later into the empty space of the tube during its use.

CHAPTER VII

THE THREE-ELECTRODE VACUUM TUBE AS AN AMPLIFIER

In the preceding chapter a study was made of the relation between plate current, plate voltage and grid voltage in a three-electrode vacuum tube, and the *characteristic curves* of Figs. 120 and 121 show, respectively, the dependence of plate current upon grid voltage for constant values of plate potential, and of plate current on plate voltage for constant values of grid potential.

Variations of plate current *at constant plate voltage* under the effect of a varying grid potential generally require the battery which energizes the plate circuit to be connected *directly* to the plate and filament of the tube. The plate potential is then always equal to the constant battery voltage.

Conditions are different if, as in Fig. 127, the battery B which energizes the plate circuit is connected to the plate and filament through a resistance R. Thus, E_B being the electromotive force of this battery B, and E_g the potential difference between the grid and filament, and I_p the intensity of the current flowing in the plate circuit of the tube, this current, flowing through the resistance R, produces between its ends a potential difference E_R which, according to Ohm's law, is equal to

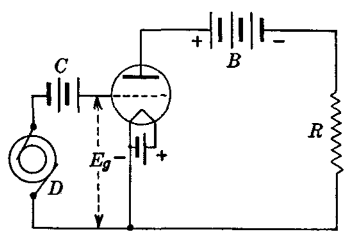


FIG. 127.

$$E_R = I_pR.$$

The plate potential of the tube is then equal to

$$E_p = E_B - E_R = E_B - I_pR.$$

If now the grid potential E_g of the tube is varied by an amount dE_g in such a manner as to produce an increase dI_p of the current in the plate circuit, the potential drop I_pR across the resistance R will correspondingly increase. The battery voltage E remaining constant (the internal resistance of the battery being small as compared to the resistance R and the internal resistance of the tube), it follows that the plate potential E_p will vary (in the present instance *decrease*) by an amount equal to

$$dE_p = -RdI_p.$$

Conversely, if the current I_p is decreased by making the grid more negative, the plate potential will increase.

The operating point of the tube thus no longer follows the static characteristic curve, which corresponds to a constant plate potential, but follows a so-called *dynamic characteristic curve*, such as one of the dotted line curves A and B (Fig. 128) corresponding

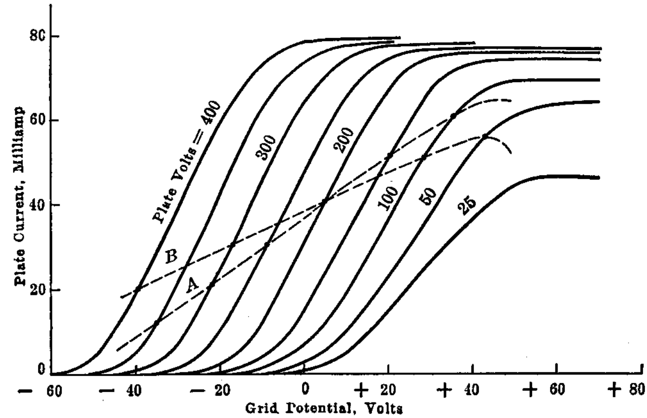


FIG. 128.

respectively to external plate circuit resistances R of 5,000 and 10,000 ohms, which crosses the successive static characteristic curves of the tube.

When a resistance R is present in the external plate circuit of the tube, the plate current variation dI_p produced by a grid voltage variation dE_g thus calls forth a plate voltage variation dE_p. Now, it was shown in the preceding chapter that a grid-potential variation dE_g produces a plate-current variation equal to

$$g \cdot dE_g,$$

where g is the mutual conductance of the tube; and also that a plate-voltage variation dE_p produces a plate-current variation equal to

$$\frac{1}{r}dE_p,$$

where r is the differential (alternating-current) internal resistance of the tube. The total plate-current variation dI resulting from the simultaneous grid- and plate-voltage changes dE_g, and dE_p is then equal to the sum

$$dI_p = g \cdot dE_g + \frac{1}{r}dE_p.$$

And since, as stated above,

$$dE_p = -R \cdot dI_p,$$

it follows that

$$dI_p = g \cdot dE_g - \frac{R}{r}dI_p,$$

and since, as known,

$$g = \frac{k}{r},$$

$$r \cdot dI_p = k \cdot dE_g - R \cdot dI_p,$$

or

$$(r + R)\, dI_p = k \cdot dE_g,$$

or

$$dI_p = \frac{k}{r + R}dE_g.$$

If, then, the grid-potential variation dE_g is a sinusoidal function of time, that is, if an alternating electromotive force

$$e_g = E_{go} \sin \omega t$$

is superimposed upon the constant grid potential supplied by a battery C, an alternating current

$$i_p = \frac{ke_g}{(r + R)}$$

will find itself superimposed upon the steady plate current of the tube. Thus, the alternating current i_p, which flows in the plate circuit of the tube when an alternating potential difference e_g is impressed between the grid and filament, is equal to the alternating current which would flow in a circuit having a resistance equal to $r + R$, at the terminals of which an alternating potential difference ke_g would be established. In other words, the alternating current flowing in the plate circuit is the same as would flow if the external plate circuit of the tube, that is, the resistanec R, were connected to an alternator (instead of the tube) having an internal resistance equal to the differential (or alternating-current) internal plate-to-filament resistance of the tube, and generating an open-circuit alternating electromotive force equal to ke_g, k being the amplification factor of the tube.

Under these conditions, the alternating potential difference arising at the terminals of the external resistance R (and also between the plate and filament of the tube) is equal to

$$Ri_p = \frac{R}{r + R}ke_g.$$

When the external resistance R is zero, this alternating potential difference is also zero. That is, the alternating current flows in the plate circuit without producing any variation of plate voltage. This corresponds to the case considered in the preceding chapter, where the battery is connected directly to the tube. The dynamic characteristic curve then coincides with the static characteristic curve corresponding to the particular plate voltage used.

When the external resistance R is infinite, or in practice when it is very great, the ratio $R/(r + R)$ becomes equal to unity and the alternating potential difference across the resistance (and across the tube) is equal to ke_g, that is to say, it is a k-times amplified reproduction of the alternating electromotive force. In this case, the alternating plate current is zero, the plate current remaining equal to the steady direct current furnished by the battery B, and the dynamic characteristic curve becomes a straight line parallel to the grid-voltage axis.

Voltage Amplification.—As follows from the preceding remarks, the ratio of the alternating electromotive force operating in the plate circuit to the alternating potential difference impressed upon the grid is equal to

$$\frac{Ri_p}{e_g} = k\frac{R}{r + R},$$

and is a maximum, equal to the amplication factor k of the tube when the external plate-circuit resistance is infinite (or, in practice, very great).

This, however, has the drawback of requiring the use of a comparatively high potential battery B for energizing the plate circuit. This may be avoided by replacing the external plate-circuit resistance R by a large reactance, or impedance of small ohmic resistance, permitting the amplification of alternating electromotive forces with a reasonably low potential plate battery B.

Thus, a coil of inductance L and small resistance may be used as shown in Fig. 129. When an alternating electromotive force e_g of cyclic frequency ω is impressed between the grid and filament, by means of an alternator D as shown here, but which may be the electromotive force due to an incoming signal, an alternating current i_p will flow in the plate circuit. This in turn induces across the coil an electromotive force equal to

$$i_p\omega L = \frac{\omega L}{r + \omega L} ke_g,$$

which becomes substantially equal to ke_g when ωL is large, that is, when the inductance of the coil and the frequency are large.[1]

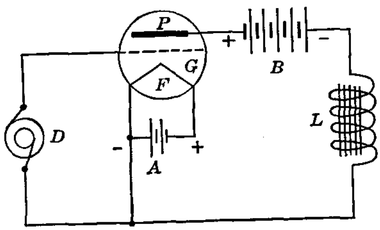


FIG. 129.

Indeed, the reactance connected in the external plate circuit may be made infinite, as required for maximum voltage amplification, by shunting the inductance coil L by a condenser C (Fig. 130) of such electrostatic capacity that the natural frequency

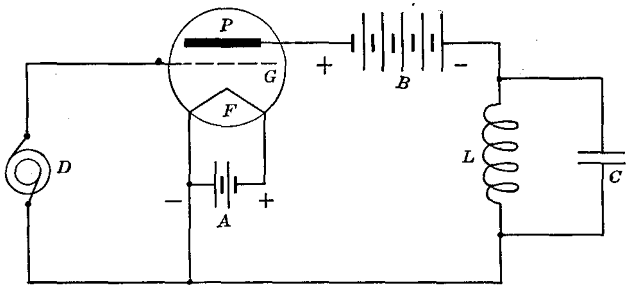


FIG. 130.

$1/\sqrt{LC}$ is equal to the cyclic frequency of the alternating electromotive force applied to the grid circuit of the tube. No alternating current will then flow in the plate circuit of the tube, and an alternating electromotive force ke_g will be developed across the circuit LC (and also between the filament and plate of the tube).

[1] The electromotive force thus developed across the coil L by the alternating plate-current component is, as in the case considered in Chap. II, 90 deg. out of phase with the latter.

It should be noted that if the electrostatic capacity of the condenser C is steadily increased, starting from a zero value, the voltage amplification increases until the natural frequency of the circuit LC is equal to the input frequency, after which it decreases again.

Power Amplification.—When the alternating current developed in the plate circuit of a tube under the effect of an alternating electromotive force impressed between the grid and filament is used for energizing a telephone receiver or some other device, the object in view is not generally to develop a maximum voltage in the plate circuit, but to deliver maximum power to the device connected in or coupled to the plate circuit.

Suppose, for simplicity, this device to be represented by a resistance R connected in the external plate circuit of the tube. The power W consumed in the device as a result of the flow of the alternating plate current through it will be equal to

$$W = I_a^2 R$$

where I_a is the effective value of the alternating current flowing in the plate circuit of the tube. Designating by E_g the effective value of the alternating grid electromotive force, and applying the equations found above, the last expression becomes

$$W = \left(\frac{kE_g}{r + R}\right)^2 R,$$

which is a maximum when $r = R$, that is, when the internal differential plate resistance of the tube is equal to the external plate-circuit resistance (or impedance).

Input Impedance.—As explained in the preceding chapter, the potentials of the plate-circuit battery B and grid-circuit battery C may be so chosen that the grid need never be made positive, with respect to the filament, for controlling the plate-current intensity between zero and its maximum value. The current in the grid circuit is then zero, since the grid does not attract any electrons to itself, and it may at first appear that the grid-potential variations which produce the plate-current variations require substantially no energy from the input source.

However, the grid and filament on the one hand, the grid and plate on the other constituting two electrostatic condensers, an alternating current will flow in the grid circuit whenever an alternating electromotive force is applied between the grid and filament of the tube (irrespective of whether the filament is heated and emits electrons or not). Thus, referring to Fig. 127, the grid-filament capacity is seen to be connected directly across the alternator D which supplies the grid electromotive force, while the grid-plate condenser is connected to the alternator through the external resistance (or impedance) R of the plate circuit of the tube.

The alternating-current source which produces the grid-potential variations will hence set up a current in the plate circuit, by electrostatic induction through the grid-plate capacity of the tube, and supply energy to this circuit, which will be dissipated as heat in the resistance contained in this circuit. Referring to Fig. 127, and designating by C_{gp}, the internal grid-to-plate capacity of the tube, by E_{g_0} and E_{p_0} the values of the alternating components of the grid and plate voltages, the charging current I_{gp} of the grid-to-plate condenser supplied by the alternator D (and which flows from the alternator through the grid-plate condenser, then through the external plate-circuit resistance or impedance R, and back to the alternator) is equal to

$$I_{gp} = \frac{E_{g_0} - E_{p_0}}{1/\omega C_{gp}}.$$

Since $E_{p_0} - E_{g_0}$ is equal to the potential difference between the grid and plate, that is, between the armatures of the grid-plate condenser of the tube, and since, from a preceding section,

$$E_{p_0} = \frac{R}{r + R} kE_{g_0},$$

it follows that

$$I_{gp} = E_{g_0}\left[1 - \frac{kR}{r + R}\right]\omega C_{gp}.$$

In addition to this current, the input source also supplies the charging current of the grid-filament capacity, equal to

$$I_{gf} = E_{g_0}\omega C_{gf}.$$

The input impedance of a tube thus consists of two parallel connected branches, one comprising the reactance of the grid-to-filament capacity, the other comprising the impedance of the circuit branch formed by the grid-to-plate capacity in series with the external plate-circuit impedance. Depending upon the nature and magnitude of the latter, the alternating current from the input source may thus lag behind, lead, or be in phase with the input (grid) electromotive force. The importance of this is particularly great in cascade amplifiers, as explained later, and in vacuum-tube oscillator circuits, notably in oscillating receiving circuits.

CASCADE AMPLIFICATION

In order to obtain greater amplification than that provided by a single three-electrode vacuum tube, the output electromotive force of a first tube may be impressed upon the input (grid) circuit of a second tube, the amplified output electromotive force of which is transferred to a third tube, and so on. This process, in which the amplification is effected in a succession of stages, is called *cascade amplification.*

It is used in radio transmitting circuits for power amplification of the generated high-frequency oscillations and of the modulating impulses. In radio receiving circuits, it serves for the amplification of the small alternating currents set up in the antenna by incoming signals, both before and after detection or rectification of the signals, voltage amplification being used in the successive stages except for the last stage, which is adjusted for power amplification to actuate the telephone receivers or other translating device.

Resistance-capacity Coupled Amplifiers.—The alternating electromotive force to be amplified is, for simplicity, assumed here to be generated by a small alternator A (Fig. 131) connected between the grid G_1 and filament F_1 of a first three-electrode vacum tube 1. To simplify the circuit diagram, the filament-heating batteries have not been shown in the figure.

The plate circuit $P_1B_1R_1F_1P$ of this tube comprises the battery B_1 and high resistance R_1 of the order of 100,000 ohms, the circuit being similar to that shown in Fig. 127. As explained before, the small alternating electromotive force applied between the grid and filament by the alternator A is reproduced, amplified, across the resistance R_1, the ratio of this amplified output electromotive force to the original input electromotive force being substantially equal to the amplification factor k of the tube 1.

This amplified electromotive force is now applied between the grid and filament of a second tube 2, connected up in the same manner as tube 1, and is thus amplified across the resistance R_2 connected in the plate circuit of this second tube. The transfer of the output electromotive force of the first tube 1, developed across the resistance R_1 to the grid circuit of the second tube 2, is effected by electrostatic induction, by connecting the filament F_2 of this second tube to one end of the resistance R_1 and the grid G_2 to the other end of this resistance. A condenser C_1 is inserted between the grid G_2 and the resistance R_1 in order to prevent the large potential drop across this resistance, due to the flow of the direct current component of the plate current of the tube 1 through it, from being impressed between the grid and filament of the second tube.

But the presence of this condenser C_1 in the grid circuit insulates the grid electrode from the rest of the circuit, and when the grid, under the effect of the alternating electromotive force applied to it, alternately becomes positive and negative, it

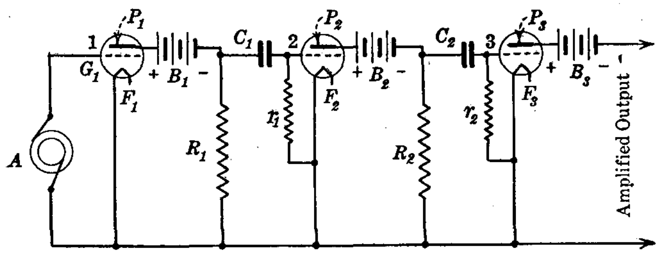


FIG. 131.

attracts during the positive cycle portions some of the electrons emitted by the filament. These in turn impart to the grid a negative charge which, increasing at every successive positive cycle portion, would soon stop the flow of current in the tube and prevent its operation.

It is therefore necessary to connect the grid G_2 to the filament F_2, in order to provide a leakage path for the negative charge collected by the grid. To this effect a resistance r_1 is connected between the grid and filament, of the same order of magnitude as the internal grid-filament resistance of the tube, generally about 1,000,000 to 5,000,000 ohms, depending upon the particular type of tube used. The actual value, while not critical, is of importance however, for if this resistance be chosen too great, it will not permit a sufficient leakage of the grid charge, and will lead to an undue reduction of the plate current in the tube and a consequent reduction of amplification. If it be made too small, it more or less short-circuits the grid to the filament and considerably reduces the electromotive force applied to the grid, correspondingly reducing the amplification.

As a result of the cascade arrangement of the two tubes 1 and 2, an alternating electromotive force is established across R_2, equal to k times the electromotive force across R_1, and hence equal to k^2e_g, where e_g is the electromotive force input to the first tube, assuming that the successive tubes have equal amplification factors.

In a similar manner, this electromotive force k^2e_g is applied to a third, fourth or fifth tube if required, the total amplification ratio being then equal to the product of the amplification ratios of the successive individual amplifier steps or stages.

A drawback of this type of amplifier, as stated before, is the necessity of using high plate-battery potentials. It has, however,

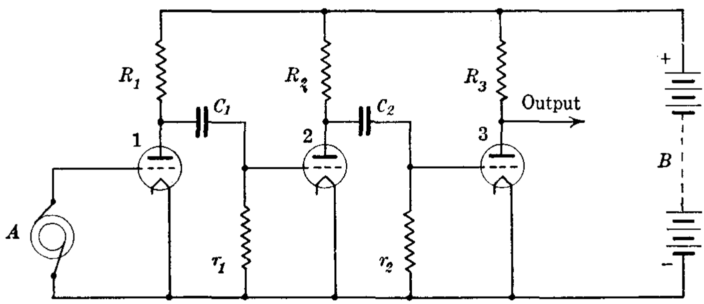


FIG. 132.

the advantage of being very simple to build in small and compact form, particularly as it may be arranged in such a manner that the plate circuits of all the tubes are energized by a single common battery B, while the filaments are heated from another common battery, as shown in Fig. 132, the plate circuits and filaments of the various tubes being simply connected in parallel across their respective batteries.

This common battery system is very widely used in multiple-tube circuits, as it greatly simplifies the entire apparatus.

Impedance-capacity Coupled Amplifiers.—In order further to reduce the size of the batteries required, choke coil coupled amplifiers are often used, as illustrated in Fig. 133. Each tube is connected substantially as the tube of Fig. 131 and operates in the manner described in connection with this figure. The transfer of the output electromotive force of one tube to the input (grid) circuit of the following tube is effected in the same manner as in the case of the resistance amplifier just described, by means of coupling condensers, the grid being, as before, connected to the filament through a high leakage resistance.

Finally, the impedances connected in the plate circuits of the successive tubes may be oscillatory circuits tuned to the frequency of the alternating current to be amplified, as shown in Fig. 134, which permits one to obtain maximum voltage amplification at this frequency, for which the reactances connected in the plate circuits become infinite.

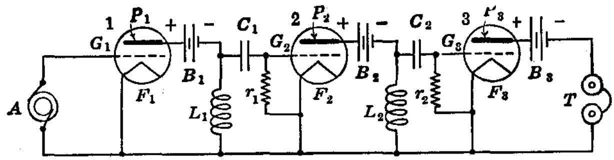


Fig. 133.

This method, called tuned amplification, is particularly useful for radio frequency amplification, in view of the great selectivity which it permits one to obtain, through which signals of only slightly different frequency may be differentiated. Thus, for instance, let F and F' be the frequencies of two equal alternating electromotive forces operating in the grid circuit of the first tube, the frequency F being that of the signals which it is desired to receive. The oscillatory circuits of the amplifier being all

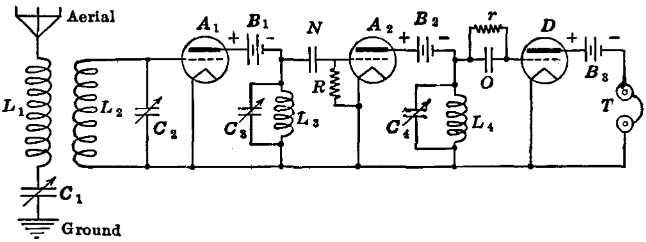


Fig. 134.

tuned to this frequency F, the output electromotive forces of frequencies F and F' developed across the tuned circuit in the plate circuit of the first tube will no longer be equal, due to the resonance of this circuit at the frequency F, so that the electromotive force of frequency F' will, say, be equal to $1/n$ the electromotive force of frequency F. Similarly, in the output circuit of the second tube, the electromotive force of frequency F' will be equal to $1/n^2$ of the electromotive force of frequency F, and so on, this selective action taking place over each amplifier stage, together with the successive amplification of the desired signals.

The relative amplitudes of the two electromotive forces thus become increasingly different as the number of amplifier stages increases, exactly as if the signals were energizing a succession of oscillatory circuits tuned to a same frequency F. Only in this latter case, as explained in a preceding chapter, the circuits must be coupled *loosely* to each other precisely, in order to obtain a sharp frequency selection. This in turn produces a considerable decrease of the absolute signal intensity. On the contrary, in the tuned cascade amplifier the successive oscillatory circuits are not coupled inductively, that is, no *mutual* coupling is established between them. The three-electrode vacuum tube permits the current in one circuit to set up a current in the following circuit, but does not allow the reverse interaction to take place; the tube operates as a relay and provides a *unidirectional* coupling between the two circuits, which clearly differentiates the two methods.[1]

Transformer Coupled Amplifiers.—Instead of using capacitive coupling between the successive stages of cascade amplifiers,

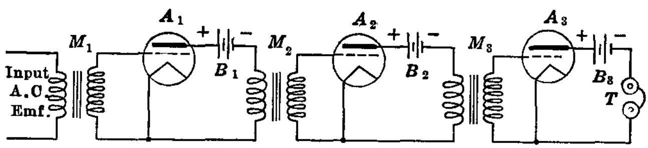


FIG. 135.

as done in the preceding examples, magnetic coupling by means of transformers may be used, as shown in Fig. 135. The alternating electromotive force to be amplified is impressed between the grid and filament of a three-electrode vacuum tube A_1, either directly or through a step-up transformer M_1 of suitable radio and impedance. This alternating electromotive force produces in the plate circuit of the tube A_1 a pulsating current which, flowing through the primary of the coupling transformer

[1] Actually, the input and output circuits of a tube are mutually coupled to each other through the electrostatic capacities of the electrodes within the tube, but the electromotive forces induced through this coupling are negligibly small when the frequency of the current to be amplified is not too great. For very high frequencies, however, the coupling effect becomes of importance, as explained in detail in the following chapter.

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. The second tube A_2 can then be made to amplify further the alternating electromotive force through the medium of a third transformer M_3 and tube A_3, and so on.

This method of amplification, when applied to low-frequency (audio) amplification, utilizes laminated iron-core transformers for coupling the successive stages. In the case of high-frequency (radio) amplification, the transformers generally have no iron core, in order to avoid the energy losses through hysteresis and eddy currents, which would be considerable at high frequencies. Some transformers have however been built for radio frequencies, having cores made of special enamelled soft-steel laminations of a thickness of the order of one-thousandth of an inch.

Direct-current Amplifiers.—In all of the amplifiers described above, the amplified electromotive force developed in the plate circuit of each tube is transferred to the grid circuit of the following tube by induction (electrostatic induction in the case of condenser coupling, magnetic induction with transformer coupling). These amplifiers will therefore amplify only *variations* of the electromotive force impressed onto their input terminals, that is, between the grid and filament of the first tube. In the absence of such variations, the plate current in each of the successive tubes is constant and independent from that in the other tubes. And if the grid potential of the first tube is given a constant value, the current and voltages in the plate circuit of the last tube will not be changed.

However, the plate current of the last (output) tube may be made a function of the *continuous* grid potential of the first (input) tube by using a conductively coupled resistance amplifier, such as illustrated in Fig. 136.

This differs from the resistance-capacity coupled amplifier of Fig. 133, in that the grid coupling condensers C are here omitted and replaced by batteries D of suitable voltage and polarity. The grid voltage of any of the tubes (except the first tube) is then equal to the sum(or difference, depending upon the connection polarities) of the voltage of the battery D and the potential difference between the ends of the resistance R belonging to the plate circuit of the directly preceding tube, due to the flow, through this resistance, of the plate current of the tube to which it is connected.[1]

If, then, the grid of the first tube is connected directly to the filament (zero grid potential), the grid potentials of the other tubes may also be made equal to zero by suitably choosing the potential and polarity of the coupling batteries D. Thus, the plate current of the first tube (as opposed to the *electron* current), flowing in the external plate circuit from filament to plate, produces across the resistance R a potential difference making the grid of the second tube strongly negative. Connecting the battery D with its positive terminal to the grid of the second tube

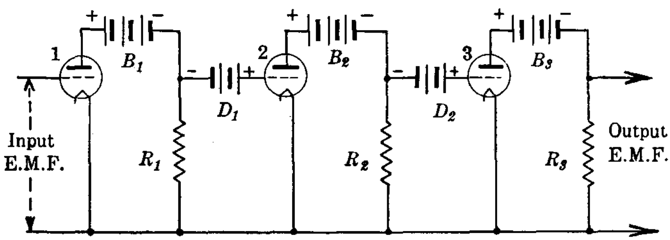


FIG. 136.

and its negative terminal to the resistance R, and choosing its potential equal to the potential difference across the resistance, will make the grid voltage of the second tube equal to zero.

Similar conditions being established for the other tubes, if the grid potential of the first tube is now altered to some other constant value, the plate current of the first tube will be altered correspondingly. This changes the potential difference across the resistance R_1 and also the grid potential of the second tube, bringing about a corresponding change in the plate current of this second tube. This in turn alters the grid potential and plate current of the third tube, and the process continues to the last tube of the amplifier, the plate-current intensity of which will hence assume a different value, within the operating limits of the tube, for every value or grid potential of the first tube. The entire amplifier thus behaves like a single tube of considerable amplification factor. It should be noted, however, that when the plate current in the first tube increases, the plate current decreases in the second

[1] The grid-leak resistances r of Fig. 139 are also omitted here, being not required, since the grids are, respectively, conductively connected to the rest of the amplifier circuit.

tube, increases in the third, and so on, the plate current in the odd-numbered tubes varying in the same direction as that of the first tube, while the current in the even-numbered tubes varies in opposite direction.

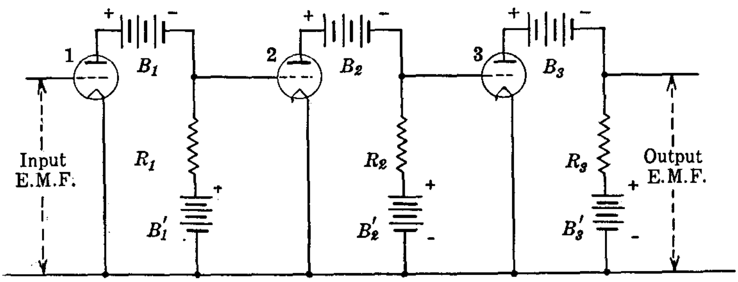


FIG. 137.

The resistance-battery coupled amplifier may be simplified by splitting the plate-circuit energizing batteries of the tubes in such manner that one part of these batteries may at the same time serve as grid-coupling batteries, as shown in Fig. 137, or even by using a single coupling battery, as in Fig. 138, which is derived from the arrangement of Fig. 137.

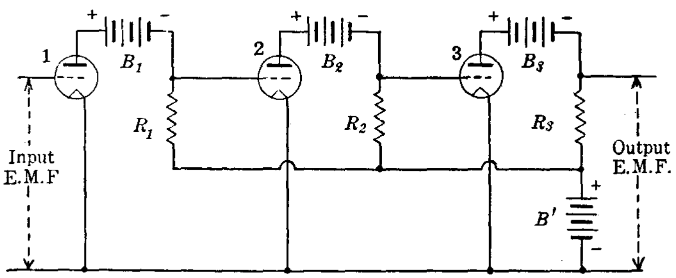


FIG. 138.

Reflex Amplification—The preceding cascade amplification arrangements are widely used in radio reception for amplification of the received signals before and after detection. An illustration is given in Fig. 139, which shows a damped or modulated wave receiving set having three stages of high-frequency and two stages of low-frequency amplification, and utilizing a crystal detector.

The circuit comprises a tuned antenna circuit L_1C_1 coupled to a tuned secondary circuit L_2C_2. This is connected between the grid and filament of a three-electrode vacuum tube A_1 adjusted, like the other tubes of the set, to operate about the straight portion of its characteristic curve, so that symmetrical variations of grid potential give rise to equally symmetrical variations of plate current. The incoming signals are thus successively amplified by the three tubes A_1, A_2 and A_3 coupled by air-core radio-frequency transformers M_1 and M_2. The amplified radio signals are then impressed, by means of a transformer M_3, upon the crystal detector D connected in series with the primary winding of an audio-frequency transformer M_4. This is shunted by a radio-frequency by-pass condenser C_3, so that the radio-frequency cur-

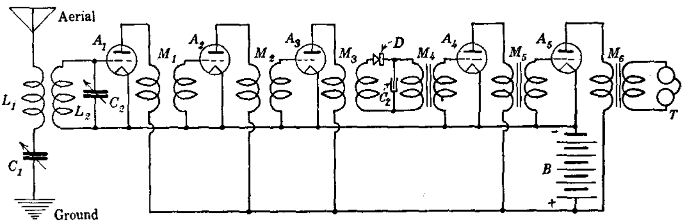


FIG. 139.

rent induced in the detector circuit by the plate current of the tube A_3 will flow through the secondary winding of the transformer M_3, the detector D and condenser C_3, while the audio-frequency current resulting from the operation of the detector will flow through the primary winding of the transformer M_4. This audio-frequency current is then amplified in succession by the tubes A_4 and A_5 and is finally sent through the telephone receivers T.

A multistage amplifier of this kind will amplify the received signal strength considerably—sometimes as much as 10^{14} times—but requires a rather large number of tubes. Methods have therefore been devised, yielding results of the same order of magnitude with a smaller number of tubes, such as the reflex method of amplification, in which the same tubes are used for amplifying the high- as well as the low-frequency currents, the two being, in each circuit, separated by suitable reactance arrangements.

Thus, Fig. 140 represents a reflex amplifier substantially equivalent to the "straight" cascade amplifier of Fig. 139. The radio-frequency alternating current set up in the antenna by the received signals is amplified in the tubes A_1, A_2 and A_3 coupled by transformers M_1 and M_2 exactly as in the case of Fig. 139. The amplified alternating current thus developed in the plate circuit of the tube A_3 is transferred to the circuit of the detector D through the transformer M_3 as before, and the low-frequency rectified or "detected" current is made to flow through the primary winding of the low-frequency transformer M_4. The secondary winding of this transformer, instead of being connected to the grid and filament of a fourth tube, as in Fig. 139, is here connected in the grid circuit of one of the radio-frequency amplifying tubes,

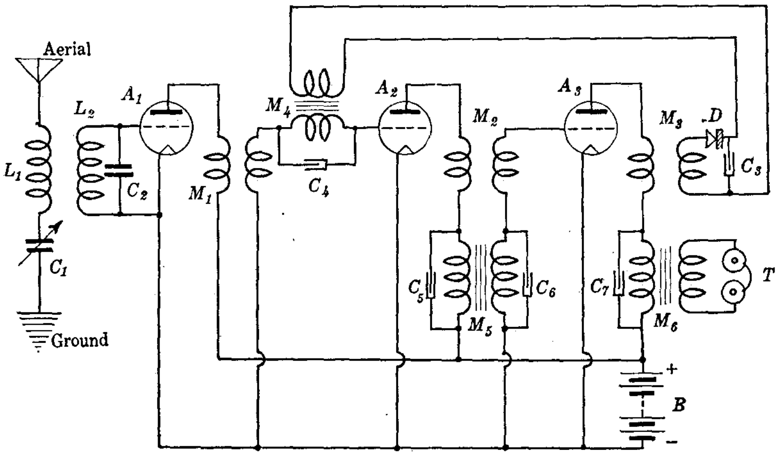


FIG. 140.

A_2 for instance. The resulting low-frequency grid-potential variations of this tube produce corresponding low-frequency plate-current variations which, acting through a transformer M_5, set up amplified low-frequency grid-potential variations of the tube A_3. This in turn actuates the telephone receivers T through a transformer M_6. The tubes A_2 and A_3 are thus made to function as low-frequency amplifiers in addition to their operating as high-frequency amplifiers, and they thus assume the functions of the tubes A_4 and A_5 of Fig. 139. Of course, the windings of the low-frequency transformers M_4, M_5 and M_6, being respectively connected in series with the windings of the corresponding radio-frequency coupling transformers M_1, M_2 and M_3, and opposing a high impedance to the radio-frequency currents, must be shunted by small by-pass condensers, as shown, in order that the radio-frequency currents may flow in the amplifier circuits.

Instead of connecting the low-frequency and radio-frequency transformers in series, they may be connected in parallel, as in Fig. 141, where the transformer pairs M_1M_3 and M_2M_4, are respectively in parallel. Small condensers C must, however, be connected in series with the radio-frequency transformer windings in order to prevent the low-frequency currents from flowing through these windings, which, in view of their low reactance at audio frequency, would otherwise practically short-circuit the audio-frequency transformer windings and render the amplifier inoperative.

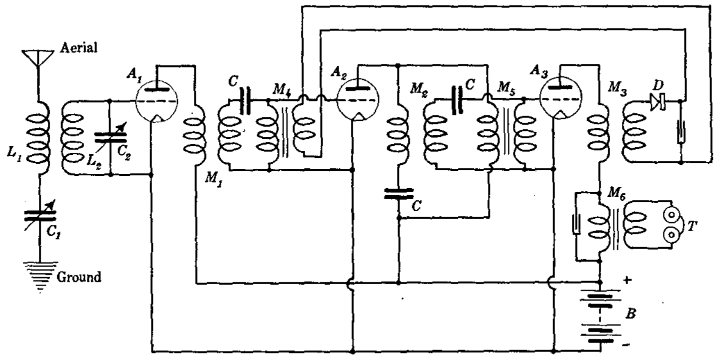


FIG. 141.

Finally, while the radio-frequency currents are amplified by the tube A_1 first, then by the tube A_2 and then by the tube A_3, the low-frequency currents may be amplified by the same tubes in a different order, in the reverse order for instance. Thus, in Fig. 142 the low-frequency impulses are transferred from the detector circuit to the grid circuit of the last tube A_3 through the transformer M_4, then from the plate circuit of this tube A_3 to the grid circ''it of the tube A_2 through the transformer M_5 and lastly from the plate circuit of this tube A_2 to the telephone receivers T through the transformer M_6.

In all of the above examples it may be noted that the first tube is used as radio-frequency amplifier only, and not as low-frequency amplifier. There is no particular difficulty in using it, like the other tubes, for the dual function of high- and low-frequency amplification, but then any low-frequency disturbing currents which may be picked up by the antenna circuit (from power lines, etc.) will find themselves impressed onto the low-frequency amplifier and be considerably magnified, creating serious disturbance.

Also, while transformer coupling was shown in these examples, any other form of coupling may be used, as described before.

The main advantage of the reflex amplifier is that it permits a reduction in the number of tubes and a consequent possible reduction of the size of the batteries required. The reliability of its operation, of course, requires the use of a detector capable of remaining adjusted over long periods of time and sufficiently rugged to withstand possible jars, in the case of portable receiving

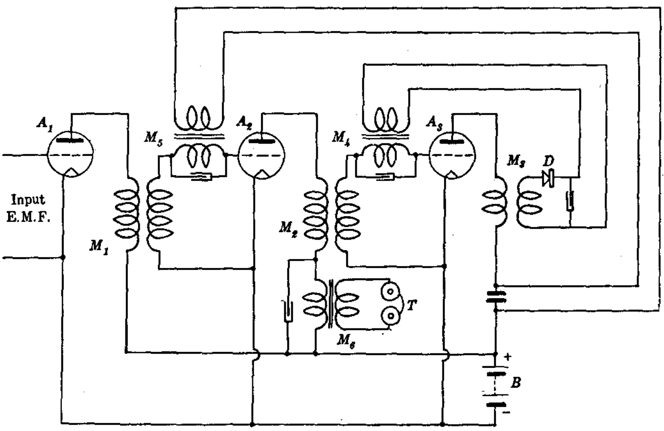


FIG. 142.

sets, the carborundum or perikon detectors being particularly well suited in this case.

A fundamental point also, perhaps better understood after studying the following chapter, is the absolute necessity of separating the radio- and audio-frequency currents, through the use of suitable choke coils or stopping or by-pass condensers, as otherwise the reflex connection would constitute a "feed-back" or retroactive connection through which the amplified output, led back into the input circuit of the amplifier, would set the latter into oscillation and prevent its proper operation. Such separation may, in general, be accomplished by the simple means just mentioned, without the use of more complicated filter circuits, and is sharper the greater the difference between the frequencies of the undetected and detected currents (hence, if the latter are of audio frequency, the greater the frequency of the incoming signals).

Performance of Vacuum-tube Amplifiers.—Vacuum-tube amplifiers, a few types of which have been described above, permit the energy necessary for the operation of the receiving apparatus (telephone receivers, telegraph recorder, etc.) to be supplied at the receiving station, this energy being then simply released or controlled in proper amounts by the incoming signals. This is in direct opposition to the simple reception methods described before in this book (except the heterodyne method), where the energy actuating the receiving apparatus was all furnished by the distant transmitting station. It therefore has permitted the increase of the intercommunication range considerably, for a given power output of the transmitting station. Indeed, it may thus seem that the power output of the transmitting station may be reduced indefinitely, provided a correspondingly powerful radio-frequency amplifier be used at the receiving station. In fact, a lower limit is set to the power of the transmitting station, which must be sufficiently great to set up an electric field about the receiving antenna of greater intensity than that produced about this antenna by the locally generated disturbing fields, such as those due to lightning and other natural causes, which would otherwise drown out the signals.

Such amplifiers are used for the reception of radio telegraphy as well as radio telephony, and although the fundamental principles are essentially the same, the performance of such amplifiers is different depending upon whether unmodulated signals, signals modulated at one single frequency, or finally speech-modulated signals are being received.

In the first case the signals each consist of *one* high-frequency oscillation. Maximum selectivity is then obtained by means of tuned cascade amplification, each stage of radio-frequency amplification being tuned to the frequency of the incoming signals. On the other hand, direct-current amplification must be used after detection, the signals being unmodulated unless the received current is, before detection, combined with a locally generated high-frequency current (heterodyne reception). In this latter case, and also when receiving telegraph signals modulated at some constant lower (audio or radio) frequency, low-frequency tuning may be used after detection, in addition to radio-frequency tuning before detection.

In the case of radio-telephony reception, such extreme selectivity is impossible, as explained in detail in a later chapter, for the amplifier must amplify equally well currents of a whole frequency band. To this effect, resistance-capacity coupled amplifiers are particularly well suited for both radio- and audio-frequency amplification. The other methods are also applicable, but the circuits must be sufficiently damped in order to flatten out their resonance curve, and the transformers, when such are used, of sufficiently generous dimensions not to introduce unduly large harmonics through distortion due to magnetic saturation of the iron cores.

Finally, it should be pointed out here that the maximum number of stages which may be used is limited, in the simple types of cascade amplifiers described above, due to the fact that the comparatively strong magnetic and electrostatic stray fields developed in the last stages of the amplifier, linking with circuit parts of the preceding stages, and also the internal-capacity coupling of the tubes, permit a feeding back of the amplified energy into the input circuits of the amplifier, leading to the generation of undamped oscillations in the amplifier circuits and creating disturbing noises and howling. Special methods must then be used in order to enable the use of a greater number of stages. These will be studied in a section of the following chapter, after the process of oscillation generation has been described.

Push-pull Amplifier.—In order that the alternating electromotive force developed in the plate circuit of an amplifier tube be an exact reproduction of the electromotive force impressed on the grid circuit, that is, in order that amplification be distortionless, it is necessary that the tube be made to operate along the straight portion of its plate-current, grid-voltage characteristic curve. The curve having the general shape of an elongated "S," this condition is best obtained by adjusting the steady (direct-current) plate and grid potentials to such values that, in the absence of any alternating electromotive force in the grid circuit, the operating point of the tube coincides with the central inflexion point A of the curve (Fig. 143). Symmetrical variations of the grid potential about its steady value then produce symmetrical variations of the plate current, which have the same

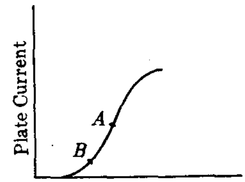


FIG. 143.

shape as the grid-potential variations, at least as long as these are small enough not to extend into the more sharply bent parts of the curve. Even then, the positive and negative maxima of the alternating current developed in the plate circuit are only slightly flattened, and the third harmonic thus introduced in the output of the tube generally remains sufficiently small to be of no material importance.

If the direct current operating point of the tube, instead of being at point A, lies in one of the bent portions of the curve, at B for instance (Fig. 143) symmetrical grid-potential variations produce *asymmetrical* plate-current variations, introducing a second harmonic (double-frequency) component in the output electromotive force of the tube which thus is a more or less distorted reproduction of the input grid electromotive force. This second harmonic is greater the more the position of point B differs from the inflexion point A, and soon becomes of the same order of magnitude as the fundamental frequency term, which is particularly objectionable in the case of radio telephony, strong overtones being thus created which completely alter the original modulations.

In order to avoid this condition and permit distortionless amplification, even though the operating point of the tube is not adjusted to the center point of the curve (which is not always easily done), the so-called push-pull amplification method may be used, as described below. This type of amplifier however, being closely related to the alternating current frequency doubler, the theory of the two devices will be developed simultaneously.

A push-pull amplifier consists essentially of two identical three-electrode vacuum tubes operated at the same continuous grid and plate potentials, respectively, and excited in phase opposition by the electromotive force to be amplified. The fundamental frequency component of the output electromotive force of one tube will then also be in phase opposition (that is, 180 deg. out of phase) with that of the other tube. But the second harmonic (double-frequency term), going through one complete cycle for every half-cycle of the fundamental frequency term, will, in the two tube circuits, be 360 deg. out of phase, which is equivalent to being in phase with each other. If, then, the output windings of the two tubes are connected in series and *boosting* each other, the two fundamental frequency terms, being in phase opposition, neutralize each other, while the double-frequency components, being in phase with each other, add and constitute the output electromotive force of the device, which then operates as a frequency doubler. If, on the other hand, the two output windings are connected in series *bucking* each other, the double-frequency term disappears while the fundamental frequency components, being of opposite polarities, add to each other (since the windings are oppositely connected), and constitute the output electromotive force of the device, which is then an undistorted amplified reproduction of the input electromotive force.

The actual circuit arrangement is shown in Fig. 144. The filaments of the two tubes *S* and *T* being connected to each other,

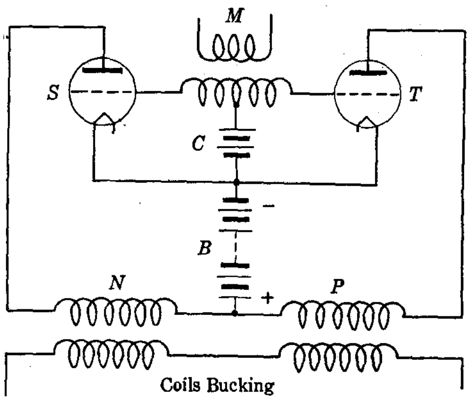


Fig. 144.

the grids of the tubes are respectively connected to the secondary terminals of the input transformer *M*. The middle point of the secondary winding of this transformer *M* is connected to the tube filaments through a battery *C* which serves to adjust simultaneously the grid potentials of the two tubes to a proper value. This battery may actually often be done away with, and is here shown for completeness. The plate circuits are energized in parallel by a common battery *B*, and comprise each an output transformer *N* and *P*, the secondary windings of which are connected in series with each other, either boosting, as in Fig. 145, or bucking, as in Fig. 144, as explained above. When, then, an alternating current is flowing in the primary of the transformer *M*, the alternating electromotive force induced in the secondary makes the grids of the tubes alternately positive and negative, but one of the grids becomes positive when the other becomes negative, and conversely. In other words, the grids are excited in phase opposition. The alternating grid voltages of the two tubes may thus be represented by the curves 1 and 2 (Fig. 146) where GBK is the characteristic curve of the tubes, B being the direct-current operating point. The plate currents, therefore, increasing in the one tube while decreasing in the other, and conversely, are represented respectively by curves 3 and 4, and are seen to be of a much distorted shape as compared with that of the input electromotive force.

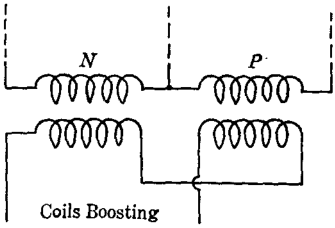


FIG. 145.

Now, if the secondary windings of the output transformers N and P are connected, as in Fig. 145, boosting each other, the output electromotive force of the device developed at the terminals XX is proportional to the rate of change of the *sum* of the two plate currents. This is represented by curve 5, which is at every instant equal to the sum of the corresponding plate-current intensities shown by curves 3 and 4. As seen, this comprises a direct-current component, and an alternating-current component having a frequency equal to twice the frequency of the input electromotive force, the device operates as a frequency doubler. If, on the other hand, the secondary windings *buck* each other, as in Fig. 144, the output electromotive force is proportional to the *difference* of the two plate currents, and, as shown by curve 6 (Fig. 146) is a pure amplified reproduction of the input electromotive force: the device operates as a substantially distortionless amplifier.

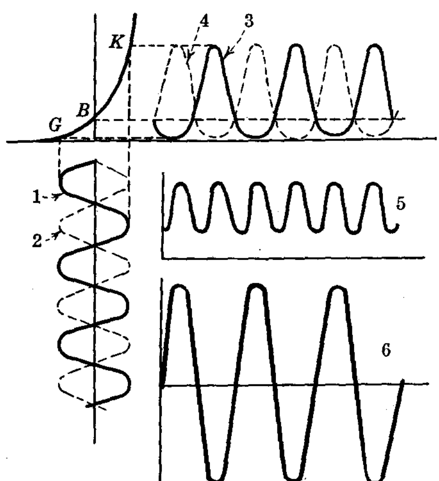


FIG. 146.

Mathematical Theory of the Push-pull Amplifier and Frequency Doubler.[1]—The characteristic curve of the three-electrode vacuum tube may be expressed with accuracy, for values of plate current comprised between zero and the saturation value, by the relation

$$i = ge - me^3 + \frac{S}{2},$$

where i is the plate-current intensity, e the grid potential, g the maximum mutual conductance (slope of the curve at its point of inflexion), S the saturation-current intensity, and m a design constant of the tube, the grid potential e being taken as equal to zero when the plate current i is equal to one-half the saturation value. This last condition simply means that when the grid potential e is zero the operating point of the tube is at point A (Fig. 143) and when it is different from zero, the operating point is on a more or less bent part B of the curve.

A constant potential c being then applied to the grids of the two tubes (Fig. 144), which places the direct current operating point at B, Fig. 143, suppose an alternating current of frequency ω to be flowing in the primary winding of the transformer M (Fig. 144). An alternating electromotive force

$$a = A \sin \omega t$$

is then applied to the grid of the first tube, while an equal and opposite electromotive force

$$-a = -A \sin \omega t$$

is applied to the other.

The grid potentials of the two tubes will then respectively be, at every instant,

$$e_1 = c + a,$$

and

$$e_2 = c - a.$$

The corresponding plate currents i_1 and i_2 of the tubes will hence be

$$\begin{cases} i_1 = ge_1 - me_1^3 + \dfrac{S}{2} \\ i_2 = ge_2 - me_2^3 + \dfrac{S}{2}. \end{cases}$$

[1] The method used here was developed by J. Zenneck, *Jahrb. drahi. Tel. u. Tel.*, Vol. 17, No. 1, January 1921, for the study of ferromagnetic frequency multipliers. Its application to three-electrode vacuum-tube devices is extended in a later chapter to the study of the balanced modulator as used in radio telephony, and of the modulator type heterodyne receiver.

If, then, the secondary windings of the output transformers N and P (Fig. 144) buck each other, the output electromotive force of the device will, as explained above, be proportional to the difference of the two plate currents

$$\begin{aligned} i_1 - i_2 &= g(e_1 - e_2) - m(e_1^3 - e_2^3) \\ &= 2ga - m(6c^2a + 2a^3) \\ &= 2(g - 3c^2m)a + 2ma^3, \end{aligned}$$

or

$$i_1 - i_2 = 2(g - 3c^2m)A \sin \omega t + 2mA^3 \sin^3 \omega t$$
$$= A\left[2(g - 3c^2m) - \frac{3A^2}{2}\right] \sin \omega t + \frac{mA^3}{2} \sin 3\omega t.$$

This expression, which represents the operation of the device as a push-pull amplifier, thus contains one fundamental and one triple-frequency term, the latter of which is generally small as compared to the former. It contains *no* double-frequency term, in spite of the fact that the constant grid potential c brings the operating point of each tube in the bent portion of its characteristic curve.

If, now, the secondary windings of the output transformers N and P (Fig. 145) boost each other, the output electromotive force of the device will, as explained, be proportional to the variable part of the sum of the two plate currents

$$\begin{aligned} i_1 + i_2 &= g(e_1 + e_2) - m(e_1^3 + e_2^3) + S \\ &= 2gc - m(2c^3 + 6ca^2) + S \\ &= 2c(g - mc^2) + S - 6mca^2 \\ &= 2c(g - mc^2) + S - 6mcA^2 \sin^2 \omega t \\ &= [2c(g - mc^2) + S - 3mcA^2] + 3mcA^2 \cos 2\omega t. \end{aligned}$$

This expression represents the operation of the device as a frequency doubler. It comprises a constant term, written between the square brackets, and a double-frequency term which, it will be noted, is directly proportional to the constant grid potential c and hence disappears when c is made equal to zero, that is, when the operating point of the tubes is at the center point of the characteristic curve.

CHAPTER VIII

THE THREE-ELECTRODE VACUUM-TUBE OSCILLATOR

A particularly important application of the three-electrode vacuum tube is its use for sustaining undamped (or continuous) electrical oscillations in an oscillatory circuit. This is achieved by suitably coupling or connecting the input (or grid) and output (or plate) circuits of the tube to each other and to the oscillatory circuit, the process being described in detail below. Direct-current energy supplied to the plate circuit of the tube is thus transformed into alternating-current energy, and the frequency of the alternating current generated, primarily determined by the constants of the oscillatory circuit, is readily adjustable over a wide range, and is remarkably constant, when adjusted. This has permitted the construction of simple continuous-wave radio-transmitting sets as well as extremely flexible heterodyne receiving sets. The vacuum-tube oscillation generator is also widely used in laboratory and industrial work as a source of constant amplitude sinusoidal alternating current, and the frequencies obtainable from this type of generator range from the lowest values—on the order of one cycle per second—to values as high as 300,000,000 cycles per second, or even more.

Physical Explanation.—Consider first the circuit of Fig. 130 studied in the preceding chapter. With the alternator D impressing an alternating electromotive force e between the grid and filament of the tube, the effect in the plate circuit is equivalent to that of an alternating plate electromotive force ke, k being the amplification factor of the tube. If the circuit LC has no resistance and is tuned to the frequency of the alternator D, then the alternating electromotive force ke will set up an alternating current in this circuit, without any expenditure of energy. In this case, the current set up by the battery B in the plate circuit $PBLFP$ remains constant, as was explained in the preceding chapter. 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 electromotive force, and 180 deg. out of phase with the alternating plate electromotive force. The pulsating plate current then follows a dynamic characteristic curve, different from the static characteristic curve of the tube, as explained in the preceding chapter.

When the tube itself is used as an oscillation generator, the grid potential is no longer supplied by an external alternator, but is obtained by coupling the grid and plate circuits. Thus, the circuit of Fig. 147 differs from that of Fig. 130 simply in that the grid, instead of being connected to the filament through an alternator D or other external source of alternating electromotive force, is connected to the filament through an inductance coil L_g inductively coupled to the coil L of the plate oscillatory circuit.

When an oscillation is then in some manner started in the circuit LC, the alternating current flowing in its coil L induces an alternating electromotive force e in the coil L_g, which is then impressed between the grid and filament of the tube. As in the preceding instance, this alternating grid electromotive force e operates like an alternating electromotive force ke impressed between the plate and filament, and therefore across the circuit LC. There are, hence, three possible cases, according to whether this electromotive force is equal to, greater, or smaller than the original electromotive force operating in the circuit LC.

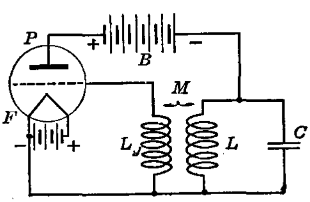


FIG. 147.

In the first case, if the electromotive force ke is *equal* to the alternating electromotive force operating in the circuit LC and in phase with it, the oscillation in that circuit will be sustained.

If the electromotive force ke is *greater* than the electromotive force operating in the circuit LC, then the oscillation in that circuit will grow in amplitude. So will also the electromotive force e induced in the grid coil L_g, and the effect is cumulative, the oscillation in the plate 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. It then stops increasing, as a further increase of the alternating grid potential would no longer produce a corresponding or proportional increase of plate current.[1]

[1] The gradual building up of the oscillations is particularly easy to observe in the case of an audio-frequency oscillatory circuit. It is necessary simply

If, finally, the electromotive force e induced in the grid circuit is such that the equivalent plate electromotive force ke is *smaller* than the electromotive force operating in the circuit LC, then the oscillation will not be sustained, but will decrease in amplitude, and the induced grid electromotive force will decrease with it. The oscillation therefore is damped, but the damping is less than that of a free oscillation of the circuit LC, since some energy is supplied to the circuit, although insufficient to compensate for the energy losses in the circuit.

The oscillatory circuit LC associated in this manner to a three-electrode vacuum tube thus functions like an ordinary oscillatory circuit in which the ohmic resistance (which as shown in Chap. I and II expressed the rate of energy consumption in the circuit) would be, more or less, neutralized by a negative resistance, representing the rate at which the tube, through its operation, supplies energy to the circuit. When this negative resistance is smaller, in absolute value, than the ohmic resistance of the circuit, the free oscillation of the latter is damped, but less than if there were no negative resistance associated to the circuit. When the negative resistance is equal to the positive resistance of the circuit, an oscillation, once started, sustains itself indefinitely. When it is greater than the positive resistance, oscillations are generated spontaneously in the circuit and increase in amplitude until the energy capacity of the system is reached.

Dynamic Characteristic.—As just stated, the three-electrode vacuum tube will, under suitable conditions, set up continuous oscillations in the circuit LC. These conditions are that if an alternating current i is flowing in the circuit LC and inducing an alternating electromotive force

$$e_g = -M\frac{di}{dt}$$

---

to connect a telephone receiver in the plate circuit of the tube, and short-circuit the feed-back grid-coupling coil, connecting the grid directly to the filament and thereby preventing the generation of oscillations. Suddenly opening the short-circuiting switch then puts the tube into oscillating condition. The generated oscillations gradually build up in amplitude, producing a sound of increasing intensity in the telephone receiver, until they reach their final amplitude, when the sound intensity remains constant. Under suitable conditions, the building up process may be made to extend over 1 or 2 sec., or even more, which is readily detected by the ear. When operating at higher (ultra-audible) frequency, the building up process is generally much more rapid, and it is necessary to resort to some form of oscillograph.

between the grid and filament of the tube, through the mutual inductance M of the grid and plate circuits, the resulting alternating plate electromotive force e_p will be equal to and in phase with the electromotive force induced by the current i across the coil L of the oscillatory circuit, that is, equal to

$$e_p = -L\frac{di}{dt}.$$

And since, as in the case of the alternator-excited circuit, the alternating grid and plate electromotive forces e_g and e_p are 180

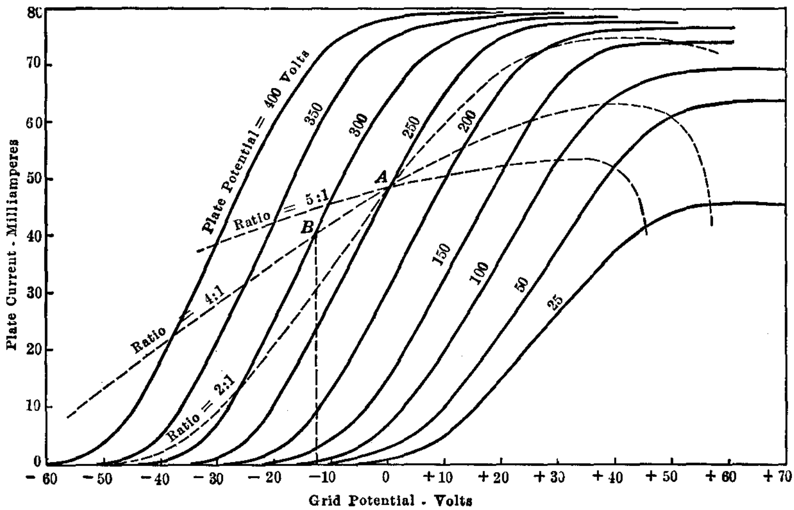


FIG. 148.

deg. out of phase, the ratio of these two electromotive forces is equal to

$$\frac{e_p}{e_g} = \frac{-L}{M} = \text{constant}.$$

This constant is, in fact, the ratio of the transformer constituted by the two coils L and L_g.

These conditions may be depicted graphically in the form of a dynamic characteristic curve giving the plate current as a function of the grid potential. Thus, consider the static characteristic curves of the tube itself, as given in solid lines (Fig. 148) 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 with the grid potential, and in opposite direction to it.

To illustrate, suppose that the ratio of plate to grid potential variations is 4 to 1. Then when 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 plate potential variation being, in this instance, +50 volts while the grid potential variation is −12.5 volts. The reverse variation of grid voltage brings about a reverse motion of the operating point. Dynamic characteristics for three values of the ratio have been shown in dotted lines in Fig. 148, that is, for three different values of coupling between the coils L and L_g. The effect of varying the coupling is thus to vary the steepness of the curve. This is shown in the following paragraphs.

Effect of Coupling on the Oscillation Amplitude.—Suppose the circuit of Fig. 147 to be oscillating. If the coupling

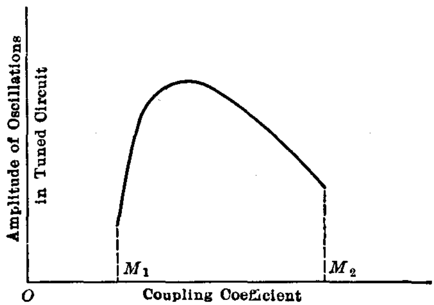


FIG. 149.

coefficient of the two coils L and L_g is gradually *reduced*, a given alternating grid potential induced in the coil L_g will require a gradually *increasing* alternating current to flow in the coil L. In other words, the looser the grid-plate coupling, the stronger the oscillation required for inducing a given grid potential. The oscillation thus increases with decreasing coupling coefficient, until the plate current varies over the entire portion of the characteristic curve comprised between the upper and lower bends. A further decrease of coupling cannot increase the plate current, and the electromotive force induced in the grid coil L_g therefore decreases with the coupling until a value of coupling is reached where the oscillations stop entirely.

On the other hand, if the coupling is gradually *increased*, the generated oscillations decrease in amplitude, and finally stop when the coupling is made too tight. Thus maximum oscillations will be obtained for a well-defined value of the coupling, which is generally only slightly greater than that required for at all generating oscillations.

This dependence of the amplitude of the generated oscillations on the degree of coupling of the circuits is represented graphically in Fig. 149.

Quantitative Theory of Oscillation Generation.[1]—Consider the system of Fig. 150, comprising an oscillatory circuit RLC inserted in the plate circuit of a three-electrode vacuum tube T and coupled inductively to the grid circuit of the tube, as in the preceding example. Let M be the mutual inductance between the grid and oscillatory circuits, and assume for simplicity the resistance R of the oscillatory circuit to be all in the inductance branch of the circuit, the resistance of the condenser branch being negligible. Let i_L, i_C and i_p be the alternating currents[2] flowing respectively in the inductance-coil branch LR, condenser C and plate-to-filament circuit of the tube, these currents being considered as positive when flowing in the directions of the arrows respectively.

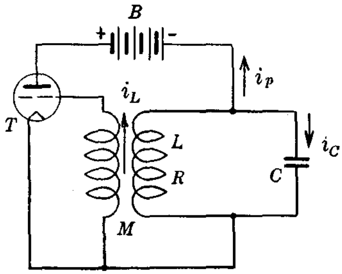


FIG. 150.

Terming r the internal plate-filament differential resistance of the tube and k its amplification factor, it was found in the preceding chapter that the alternating-plate current i_p, plate potential e_p and grid potential e_g are related by the equation

$$ri_p = e_p + ke_g.$$

Since in the present instance

$$\begin{cases} i_p = i_L - i_C \\ e_p = -Ri_L - L\dfrac{di_L}{dt} \\ e_g = -M\dfrac{di_L}{dt} \end{cases}$$

[1] See GUTTON, G., *Revue Générale de l'Electricité*, Vol. VI, p. 14, 1919.

[2] Superimposed upon the direct current furnished by the battery B.

this equation becomes

$$r(i_L - i_C) = -Ri_L - (L + kM)\frac{di_L}{dt}$$

and since, as shown in Chap. II, the condenser current is equal to

$$i_C = C\frac{de_p}{dt},$$

it follows that

$$i_C = \frac{1}{C}\left(-R\frac{di_L}{dt} - L\frac{d^2i_L}{dt^2}\right),$$

so that the equation finally becomes

$$L\frac{d^2i_L}{dt^2} + \left(R + \frac{L + kM}{Cr}\right)\frac{di_L}{dt} + \frac{i_L}{C}\left(1 + \frac{R}{r}\right) = 0,$$

or, the ratio R/r in the last term being very small (since the circuit resistance R is generally of a few ohms while the internal resistance r of the tube amounts to several thousand ohms),

$$L\frac{d^2i_L}{dt^2} + \left(R + \frac{L + kM}{Cr}\right)\frac{di_L}{dt} + \frac{i_L}{C} = 0.$$

Now, it was found in Chap. II that the *free oscillation* of a circuit could be expressed by an equation of the form

$$L\frac{d^2i}{dt^2} + R\frac{di}{dt} + \frac{1}{C} = 0.$$

This differs from the above in that the resistance R in the second term is replaced by the expression $\left(R + \frac{L + kM}{Cr}\right)$ and thus *the circuit LRC associated to the three-electrode vacuum tube in the manner shown in Fig.* 150 *behaves like a freely oscillating circuit* (not connected to any tube) *having same inductance L and capacity C as the actual circuit, but a resistance equal to* $\left(R + \frac{L + kM}{Cr}\right)$, and this equivalent circuit may be considered as an ordinary *freely* oscillating circuit, as shown in Chap. II.

Now, it was pointed out in the first section of this chapter that the operation of the tube as an oscillation generator may be considered as equivalent to adding a negative resistance to the circuit RLC. Actually, as now shown here, it is the term $\frac{L + kM}{Cr}$ which is added to the resistance R of the circuit. In this expression, the quantities L, C, k, r are all positive. The mutual inductance M, however, may be made positive or negative, depending on the connection polarity of the grid-circuit coupling coil to the grid and filament of the tube.

If the mutual inductance M is *positive*, then the resistance of the equivalent circuit, equal to $R + \frac{L + kM}{Cr}$ is *greater* than the resistance R of the actual circuit, and oscillations are damped more rapidly than if the circuit LRC was not connected to the tube. The tube here serves to choke out the circuit oscillations.

If, on the other hand, the mutual inductance M is *negative* the expression $R + \frac{L+kM}{Cr}$ may be made smaller than R; may in fact be made to equal zero or even negative, when

$$M < -\frac{Cr}{k}\left(R + \frac{L}{C}\right)$$

(that is, when M is negative and its *absolute* value is *greater than*

$$\frac{Cr}{k}\left(R + \frac{L}{C}\right)\Bigg).$$

In this latter case, the resultant resistance of the circuit being negative, the oscillations do not decrease but on the contrary increase logarithmically, that is, at a substantially constant rate, until the bends of the dynamic characteristic curve are reached, when a stable oscillation maintains itself continuously. Its frequency is equal to

$$\omega = \sqrt{\frac{1}{CL}\left(1 + \frac{R}{r}\right) - \frac{1}{4L^2}\left(R + \frac{L + kM}{Cr}\right)^2}$$

which reduces to

$$\omega = \sqrt{\frac{1}{CL}\left(1 + \frac{R}{r}\right)}$$

when the stable condition is reached. This is but little different from the natural oscillation frequency of the circuit when not connected to any tube.

The above condition for oscillation generation, that the mutual inductance M be negative and that its *absolute* value fulfil the requirement

$$M \geqq C\frac{r}{R}\left(R + \frac{L}{Cr}\right)$$

may also be written

$$\frac{k}{r}M \geqq CR + \frac{L}{r},$$

or

$$gM \geqq CR + \frac{L}{r},$$

where g is the mutual conductance of the tube, as defined before.

Now consider the characteristic curve of the tube (Fig. 151) and suppose the operating point P of the tube, when not oscillating, to be about the center of the curve, at its point of inflexion, which may be obtained by means of batteries of suitable voltages in the plate and grid circuits. The mutual conductance g of the tube, being equal to the slope of the curve at point P, is thus as great as possible, and if the value of M is large enough to fulfil the above condition, oscillations will start in the circuit. This means that the grid voltage of the tube will oscillate either way about its original value, and point P describes a portion NS of the curve, determined by the amplitude of the oscillations. The average or mean mutual conductance, roughly equal to the slope of the straight line NS, is then smaller than the original value,[1] and decreases when the oscillation amplitude increases until finally the product gM in the above equation becomes so small as to be no longer greater, but simply equal to the right-hand member of the equation. A stable operating condition is then reached.

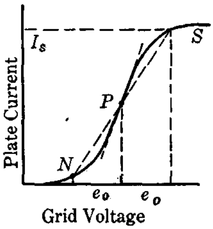


FIG. 151.

Thus, when the direct-current operating point of the tube is about the center of the curve, point of maximum mutual conductance g, oscillations start spontaneously when the value M is made large enough, and of negative polarity. When, under these conditions, the oscillations have grown to such an amplitude that the point P describes the whole curve NS, the plate current in the tube varies between substantially zero and the saturation value I_s. The amplitude[2] of the *alternating*-plate current component is hence equal to

$$I_p = \frac{I_s}{2},$$

[1] HAZELTINE, L. A., *Proc. Inst. Radio Eng.*, April 1918.

[2] HARMS, F., *Jahr. draht. Tel. u. Tel.*, Vol. 15, No. 6, June 1920.

and the corresponding alternating-grid voltage amplitude is

$$E_g = \frac{I_s}{2g_m}$$

where g_m is the mean mutual conductance of the tube (slope of the straight line NS).

And since this alternating grid voltage E_g is induced in the grid circuit by the alternating current (of amplitude I_L) flowing in the plate winding L through the mutual inductance M, it is also equal to

$$E_g = I_L \omega M,$$

from which the current I_L in the oscillatory circuit LC may be calculated:

$$I_L = \frac{E_g}{\omega M} = \frac{I_s}{2\omega g_m M},$$

and since, as explained above, stable oscillations are obtained when

$$g_m M = CR + \frac{L}{r}$$

it follows

$$I_L = \frac{I_s}{2\omega}\left(CR + \frac{L}{r}\right).$$

Effect of Grid Voltage on Oscillation.—If now a different continuous potential is applied to the grid, bringing the direct-current operating point P of the tube around one of the bends of the curve, the mutual conductance g (equal to the slope of the curve at point P) becomes smaller. With the same value of coupling M the product gM may still be large enough to permit oscillation generation. But it may also become too small, and hence require an increase of the grid-plate coupling for permitting oscillations to be generated.

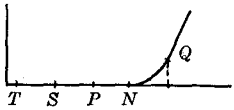


FIG. 152.

Finally if the direct-current operating point P lies in one of the horizontal portions of the curve, the mutual conductance g is equal to zero. The product gM is then also equal to zero, irrespective of how large M is made. Thus, suppose that a negative potential is applied to the grid (Fig. 152), bringing the direct current operating point P of the tube in the lower horizontal portion of the curve. If an oscillation is then started in the circuit LC (Fig. 147) an alternating voltage is induced in the grid circuit, and point

P will oscillate about its original position. If these oscillations take place, say, between points N and S, both located in the horizontal portion of the curve, then the grid-potential variations produce *no* plate-current variations. The tube remains inoperative, and the oscillations in the circuit LC die out as if the circuit was not connected to the tube. If, however, the induced grid voltage is great enough to make point P oscillate between points T and Q, over a part of the ascending curve, then the tube will supply energy to the oscillatory circuit during a part of every cycle, and this energy *may* be sufficient to sustain the oscillations in the circuit, which then grow until a stable condition is reached, as before. But such a large starting alternating grid voltage can only be obtained through either or both a strong initial oscillation in the circuit LC and a close coupling between this circuit and the grid circuit of the tube. The oscillations in this case do not start spontaneously, and when started, set in with considerable intensity.

For even more negative values of the grid potential, the oscillations required for starting the operation are so great that the amount of energy supplied by the tube during the short periods of plate-current flow is not sufficient to sustain them, and the oscillations die out.

Efficiency of the Vacuum-tube Oscillator.—Suppose the grid potential of the tube to be so chosen that the direct-current operating point is about the center of the curve, as in Fig. 151. When no oscillations are generated, the power supplied by the plate circuit energizing battery B is equal to the product I_pE_p of plate current and voltage, and is dissipated in the internal direct-current plate resistance of the tube, being transformed into heat at the plate electrode. When oscillations are generated the plate current and voltage vary alternately above and below their respective steady values. The average plate current and voltage are not changed, however, and the average power furnished by the battery remains the same. But instead of being all transformed into heat at the plate of the tube, a part of it is used to supply the resistance losses in the oscillatory circuit. Under the most favorable operating conditions, the power supplied by the battery is then equally divided between the tube and the oscillatory circuit, and the efficiency of the tube as an alternating-current generator, or more precisely, as a transformer of direct current into alternating current is equal to 50 per cent.

If the direct-current operating point of the tube is shifted to the lower bend of the characteristic curve, however, by using a sufficiently large negative grid voltage, the average plate current will be decreased, and, as explained in the preceding section, the plate current will consist in a succession of short impulses, there being one such impulse for every cycle of the oscillation. The efficiency of the oscillator may then be increased well above 50 per cent, the alternating-current component decreasing, in proportion, less than the direct-current component.

A drawback of this method is that it introduces harmonics in the plate current of the tube, and that the difficulties for starting the oscillations increase with increasing negative grid potential. This latter objection may be remedied in a measure by means of the following method:

Suppose a sinusoidal alternating current is made to flow through the primary winding of a transformer A (Fig. 153). A

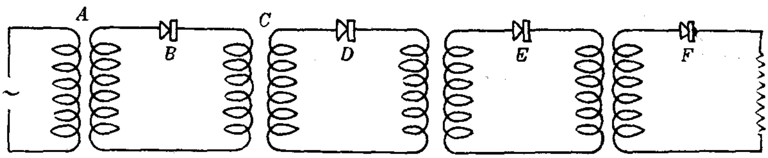


FIG. 153.

sinusoidal alternating electromotive force will be induced in the secondary winding of this transformer, and a current will flow in the circuit connected to this winding. In this circuit, comprising a rectifier B, current will flow when the electromotive force induced in the secondary winding of the transformer A varies from zero to, say, its positive maximum and back to zero, but not during the following, negative, half-cycle. The duration of this current impulse is thus equal to that of one half-cycle of the input current. If this current impulse be made to flow through the primary of a transformer C, an electromotive force will be induced in the secondary of this transformer, which will be positive when the current in the primary rises from zero to its maximum value, and negative when the current decreases back to zero. The circuit comprising a rectifier D, this induced electromotive force will set up a current in the circuit in one direction only, say when it is positive, and the resulting unidirectional impulse thus generated in the circuit has a duration equal to one quarter-cycle of the original input current.

If a third and fourth rectifier circuit E and F be added to the system, the unidirectional impulses set up in these circuits will flow during one-eighth and one-sixteenth cycle of the original input current, respectively, and the phase of these impulses, that is, the point of the cycle of the input current at which they take place, is determined by the connection polarity of the various successive rectifiers.[1]

An application[2] to the three-electrode vacuum tube oscillator is shown in Fig. 154, where the coupling of the grid circuit to the oscillatory circuit LC is made through one rectifier D. The alternating current flowing in the oscillatory circuit LC induces an alternating electromotive force in the rectifier circuit, in which a unidirectional current flows during one-half cycle only. This induces an electromotive force in the grid circuit consisting of a positive and a negative impulse, each of a duration equal to one quarter-cycle of the alternating current flowing in the circuit LC. If the grid is made sufficiently negative, by means of a battery E, normally to stop the plate-current flow in the tube, such plate current will only flow during the positive grid-voltage impulse, and this may be large enough, for suitable transformer ratios, to make the plate current of the tube vary from zero to the maximum, saturation, value. Of course, instead of one intermediate rectifier circuit, a chain of such circuits may be used, as just explained, for further shortening the duration of the plate-current impulses, and increasing the oscillator efficiency. It should be noted however that at the same time the amount of power supplied to the oscillatory circuit during each impulse decreases, hence also the oscillation amplitude decreases.

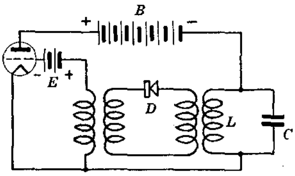


FIG. 154.

Typical Oscillator Circuits.—In the previous discussion the conditions were established which must be secured if a vacuum-tube circuit is to sustain continuous oscillations. Qualitatively speaking, it is simply necessary to couple the plate and grid circuits of the tube to a common oscillatory circuit (or oscillatory circuit arrangement) and in such a manner that the resulting grid-to-plate coupling will be negative. The number of possible circuits answering these requirements is very large, and the actual

[1] This system constitutes, in fact, one branch of a frequency multiplier.
[2] British Patent 213562 to Western Electric Company and H. W. Nicholls.

quantitative relations are special in almost every case. They are, however, often easily obtained in a manner similar to that followed in the previous sections for the particular circuit considered. Despite their differences, these various vacuum-tube circuits may be classified in a few fundamental circuits, which will be briefly described here.

Thus, the oscillatory circuit may be coupled to either the grid or the plate circuit of the tube, or to both these circuits together. Also, the couplings may be electrostatic or magnetic, or both. In all these circuits, an oscillation started in the oscillatory circuit induces an electromotive force in the grid circuit, and the grid-potential variations produce corresponding plate-current

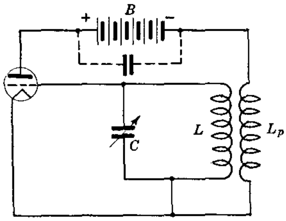


FIG. 155.

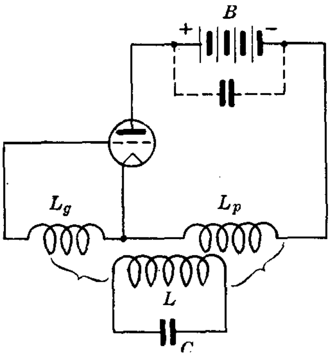


FIG. 156.

variations which in turn induce an electromotive force in the oscillatory circuit, which is in phase with and equal to (or greater than) the electromotive force operating in this circuit.

As a first example, the circuit of Fig. 155 differs from that of Fig. 147 in that the oscillatory circuit LC is inserted in the grid circuit of the tube instead of the plate circuit. The plate circuit here simply comprises a coil L_p coupled inductively to the grid coil L, which at the same time constitutes the inductance of the oscillatory circuit. The mode of operation is readily understandable from the preceding remarks.

In the arrangement of Fig. 156, the grid and plate circuits of the tube are independently coupled to the oscillatory circuit LC. Proper operation is obtained for suitable values of the grid and plate couplings, which may be done, as in the preceding examples, by moving the coupling coils toward or away from the inductance coil of the oscillatory circuit, or rotating them more or less, altering correspondingly the mutual inductance.

Another form of the last arrangement is shown in Fig. 157, in which parts of the oscillatory circuit inductance are inserted directly in the plate and grid circuits respectively. The self-induced electromotive forces set up in these winding parts by the oscillation current operate then in these circuits. They are made of opposite polarities, as required for oscillation generation, by connecting the filament to a point of the circuit inductance located *between* the points connected to the grid and plate. By making the filament connection point a sliding contact, as in Fig. 158, the ratio of plate and grid inductance may readily be altered, and the feed-back coupling may thus be set to its proper value.

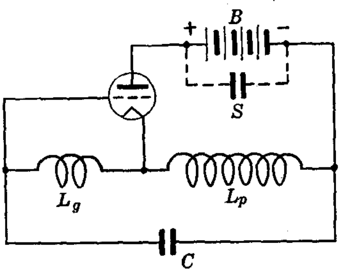


FIG. 157.

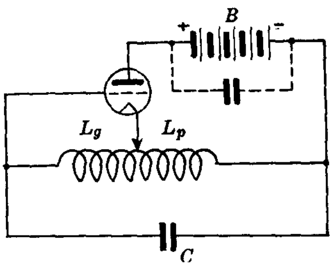


FIG. 158.

In all the above examples, the alternating component of the plate current flows through the battery B which furnishes the direct current to the plate circuit of the tube. In order to reduce high-frequency energy losses in the resistance of this battery, and also the disturbing effects of the electrostatic capacity between battery and ground, it is well to shunt the battery by a by-pass condenser, as shown in dotted lines in the diagrams.

Another way is to connect a choke coil of large inductance in series with the plate-circuit battery, connecting this in parallel with the high-frequency plate circuit. Figure 159 shows an arrangement derived from that of Fig. 155, the oscillatory circuit LC being connected in the grid circuit of the tube, while the plate circuit is coupled magnetically to its inductance L through a coupling coil L_p. The plate circuit is energized by a battery B connected to the filament and plate through a choke coil K. The plate-circuit coupling coil L_p is connected between the filament and plate of the tube, through a condenser H placed in series with it, in order to prevent the flow of direct current through the coil L_p, which would otherwise short-circuit the circuit branch BK. An oscillation being then started in the oscillatory circuit LC, an alternating electromotive force e finds itself impressed between the filament and grid of the tube. The plate circuit being energized by the battery B in series with the choke coil K, an alternating electromotive force substantially equal to ke is developed between the plate and filament, as explained in connection with Fig. 129 (amplifier operation of the tube). This alternating electromotive force being impressed upon the circuit branch L_pH, an alternating current will flow through the latter, which will induce an oscillation sustaining electromotive force in the circuit LC through the magnetic coupling of the coils L and L_p. This induced electromotive force will, of course, be greater the greater the alternating current flowing through the coil L_p, and this in turn depends on the electrostatic capacity of the condenser H. If this be made very small, only a small alternating current flows

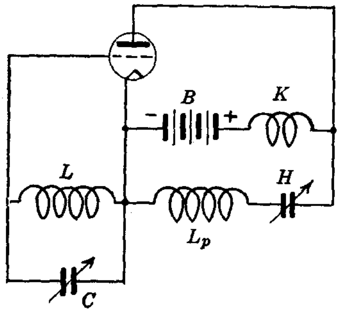


FIG. 159.

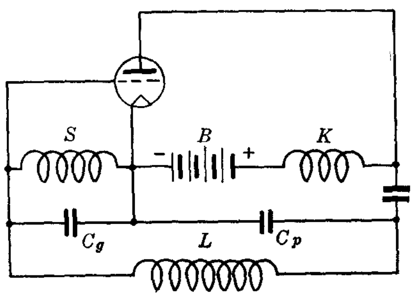


FIG. 160.

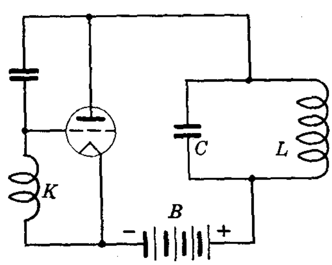


FIG. 161.

through the coil L_p, which may not be sufficient for sustaining the oscillations in the circuit LC. Thus, by suitably adjusting the capacity of the condenser H it is possible to regulate the amount of power fed back from the plate circuit into the oscillatory circuit LC. This constitutes, so to speak, an electrostatic control of the magnetic feed-back coupling of the plate and grid circuits.

A purely electrostatic coupling arrangement is shown in Fig. 160, which is the counterpart of that of Fig. 157. The capacity of the oscillatory circuit is constituted by the condensers C_p and C_g which are inserted, respectively, in the plate and grid circuits of the tube, the reactive potential drops across these condensers constituting the alternating plate and grid operating electromotive forces. Direct current is fed to the plate, as in the preceding example, by a battery B connected in series with a choke coil K, a series energization of the plate circuit being here impossible on account of the plate coupling condenser C_p which insulates the plate from the filament. In the same manner, the grid-coupling condenser C_g insulating the grid from the filament, a conductive grid-to filament must be provided, through a choke coil S or a high resistance in order to provide a leakage path for the negative charge which would otherwise accumulate on the grid, when the latter collects electrons from the filament during the positive half-cycles of the alternating electromotive force.

Electrostatic feed-back coupling is also possible when the oscillatory circuit is connected in only one of the grid and plate circuits, as shown for instance in Fig. 161, where the circuit LC is in the plate circuit, and in Fig. 162, where it is in the grid circuit. Considering this last example, an oscillation started in the circuit LC impresses an alternating electromotive force e between the grid and filament of the tube. The plate circuit being energized by a battery B in series with the choke coil K, an alternating electromotive force ke is developed across the ends of the circuit branch BK, that is, between the plate and filament, as in the case of amplifier operation of the tube. If now a condenser H be connected between the plate and grid, the oscillatory circuit LC will be connected, through this condenser, across the filament and plate, and a fraction of the alternating plate electromotive force ke will be impressed upon the circuit LC where it will, if large enough, sustain the original oscillation.

The operation of the arrangement of Fig. 161, in which the oscillatory circuit LC is inserted in the plate circuit and the choke coil K in the grid circuit, may be explained along similar lines.

The operation is based upon the fact that the reactance of the choke coil K is great at the operating frequency. An infinite reactance, constituted by an oscillatory circuit tuned to the operating frequency, may be used in its place, the arrangement being shown in Fig. 163. An oscillatory circuit is inserted in the plate circuit of the tube and a second oscillatory circuit, tuned to the first circuit, in the grid circuit.

It is noteworthy that the coupling condenser *H* is connected in parallel with the condenser formed by the plate and grid within the tube. The internal grid-to-plate capacity of the tube thus forms a part of the feed-back capacity, and is indeed sufficiently great, in many cases, for permitting the generation of continuous oscillations without the addition of any external coupling condenser. This is particularly the case when the oscillations generated are of high frequency, when the reactance of the internal grid-plate condenser is correspondingly small.

This property, which in the present case is the useful feature of the oscillation generator, is a serious cause of trouble in radio-

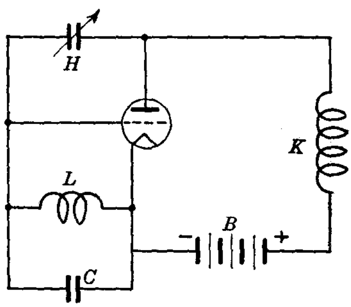


FIG. 162.

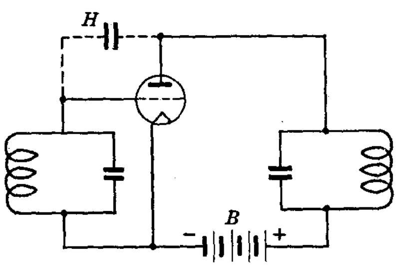


FIG. 163.

frequency cascade amplifiers, which in this manner become capable of spontaneously generating oscillations, preventing the proper operation of the amplifier. This will be discussed presently.

Oscillation Generation and Neutralization in Cascade Amplifiers.—Consider one of the intermediate tubes of a radio-frequency cascade amplifier, as described in the preceding chapter, in Fig. 133 or Fig. 134, for instance.

The plate circuit of this tube comprises a high impedance, across which the pulsating plate current of the tube builds up its output electromotive force. The grid circuit, on the other hand, comprises the high impedance belonging to the external plate circuit of the directly preceding tube. The grid and plate circuits being coupled to each other through the internal grid-to-plate capacity of the tube, the same conditions are seen to obtain as in the last examples of oscillation generators, and it thus often occurs that for this reason the amplifier spontaneously breaks into oscillation, the generated oscillations drowning out the signals to be amplified, and preventing the proper operation of the amplifier.

This generation of oscillations may be prevented in several different manners:

1. The ohmic resistances of the grid and plate external impedances may be increased, by means of a rheostat, for instance, until the positive resistance of the circuits becomes greater than the negative resistance developed by the tube. This, however, decreases the frequency selectivity of the amplifier and artificially increases the energy losses.

2. An equivalent method consists in giving to the grid of the tube a slightly positive potential, which makes it possible for

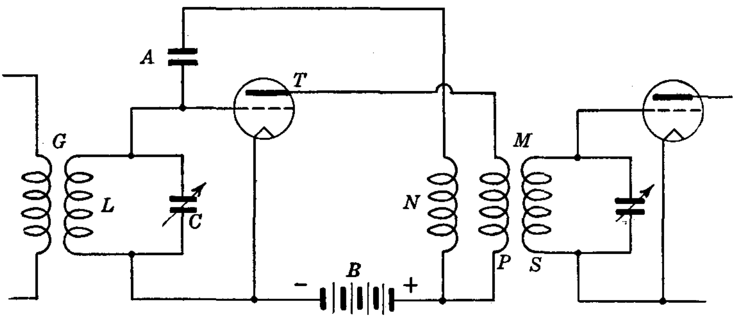


FIG. 164.

the electrons emitted by the filament to enter the grid circuit. Energy losses are thus created in this circuit, which absorb the excess of energy supplied by the tube, and prevent oscillation generation.

3. A better method, not based on the creation of additional energy losses in the circuits, is based on the fact, stated in a preceding section, that a *negative* coupling is required between the plate and grid circuits in order that the tube may generate oscillations, a *positive* coupling producing the opposite effect, which is to dampen out or prevent the oscillations.

A negative coupling being, in the present case, provided through the internal grid-to-plate capacity of the tube, its action may hence be *neutralized* or compensated by establishing a positive coupling of suitable value between the plate and grid circuits[1]. This positive coupling may be electrostatic or magnetic,

[1] L. V. Hazeltine, *Wireless World*, April 1923.

but is preferably constituted in such a way as to have the same form as the negative coupling circuit, the action of which it is intended to neutralize. In this manner the positive and negative coupling reactions vary in the same direction when the operating frequency is altered, and compensate each other over the entire frequency range of the amplifier.

Thus, Fig. 164 represents one of the intermediate tubes T of a multistage cascade amplifier. The grid of this tube is connected to the filament through an oscillatory circuit LC coupled inductively to the plate circuit of the preceding amplifier tube through the interstage coupling transformer G, of which the inductance L constitutes the secondary winding. The plate circuit of the tube T, energized by the battery B, comprises the primary winding P

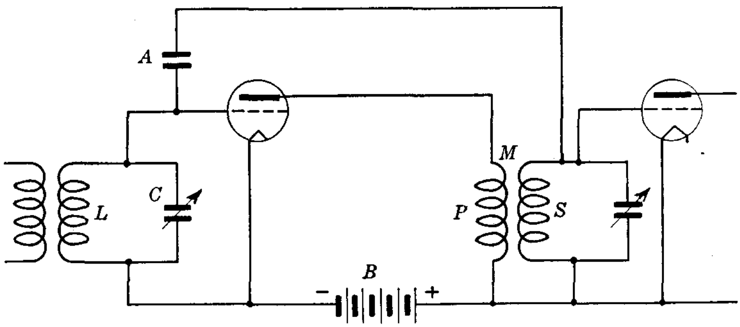


FIG. 165.

of the transformer M through which it is coupled to the following amplifier tube. When then an alternating current flows through the primary of the transformer G, an oscillation is set up in the circuit LC, alternately varying the grid potential of the tube T. This in turn varying the plate current of the tube, an amplified alternating electromotive force is induced in the secondary winding S of the transformer M, operating the following tube of the amplifier.

But at the same time a self-induced electromotive force is set up across the primary winding P of the transformer M through the pulsation of the plate current of the tube T. And one end (the filament end) of this winding P being connected directly to the circuit LC, while the other end of this winding is connected to this same circuit through the condenser constituted by the plate and grid of the tube, this self-induced electromotive force will be impressed back upon the circuit LC where, as shown in the preceding section, it will sustain continuous oscillations.

If now a winding N, coupled to the winding P, is connected to the circuit LC, one end directly, the other through a condenser A, the electromotive force induced in it will, similarly, be impressed on the circuit LC; and if the connection polarity of the winding N is suitably chosen, the electromotive force impressed by it upon the circuit LC will be of opposite polarity to that due to the winding P, which it will hence neutralize exactly for a proper capacity of the condenser A.

Of course, instead of using a separate secondary winding N, the ordinary secondary winding S of the interstage coupling transformer M may be used, as shown in Fig. 165.

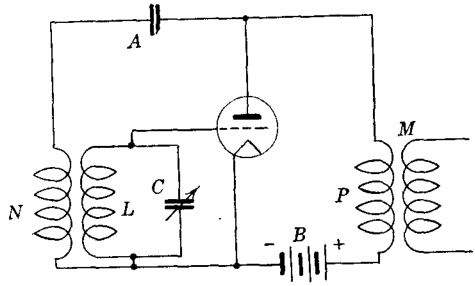


FIG. 166.

Another alternative consists in connecting, between the filament and plate of the tube, a circuit branch AN (Fig. 166) comprising a condenser A and coil N, the latter being coupled inductively to the input oscillatory circuit LC of the tube. The alternating electromotive force developed across the coil P through the operation of the tube sets up an alternating current in this circuit branch AN which induces in circuit LC an alternating electromotive force, and this, for a suitable connection polarity of the coil N and proper capacity of the condenser A, may be made equal and opposite to the electromotive force impressed upon this same circuit LC by the coil P through the internal grid-plate capacity of the tube.

The coil N need not be a separate coil, but may be the primary winding of the interstage coupling transformer used for coupling the output (plate) circuit of the preceding tube to the input (grid) circuit of the tube T of the amplifier.

Or else, the transformer NL may be replaced by an autotransformer, as shown in Fig. 167: the filament of the tube is connected to an intermediate point of the grid-circuit inductance L. One end of this inductance is connected to the grid, while the other is connected to the plate, through a condenser A. Due to the alternating electromotive force developed in the plate circuit, across the ends H and K of the winding P, an alternating current will then flow, on the one hand, in the circuit branch comprising the portion F of the inductance L and the condenser A, and on the other hand in the circuit branch comprising the portion D of the inductance L and the condenser formed by the grid and plate of the tube. The currents flowing in the two portions D and F of the inductance L are thus in opposite directions, hence also the electromotive forces induced by them in the coil L. Complete neutralization is then obtained when, for a suitable capacity of the condenser A, these two are of equal magnitude.

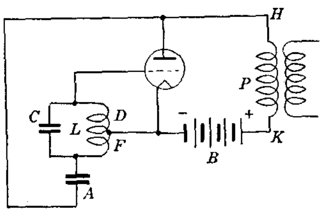


FIG. 167.

In all the above methods, it is possible to regulate the extent of neutralization by means of the capacity of the condenser A. If this be made slightly less than required for complete neutralization, then some energy is fed back from the plate to the grid circuit, not sufficient to sustain oscillations in it, but still capable of reinforcing the signals.

Strictly speaking, the preceding methods prevent oscillation generation due to energy feed-back through the internal capacity of the tube. But oscillations may also be generated in cascade amplifiers due to the fact that the stray electrostatic and magnetic fields developed in the various parts of the apparatus interlink more or less, particularly when the amplifier is of small overall dimensions. It is then necessary to electrostatically and magnetically shield the various condensers, coils, transformers, etc., or to place them in such relative positions that their fields do not link with each other (staggering the inductance coils for instance, or placing them at right angles to each other), and in some cases to reverse the connection polarity of certain of the coils, in order to reverse their fields.

Special Short-wave Oscillator Circuits.—The circuit arrangements described above for the generation of oscillations by means of three-electrode vacuum tubes may be used for almost any value of frequency. When generating currents of extremely high frequency (of the order of 250,000,000 cycles per second), these circuits assume particularly simple shapes, for the tube and the wires connecting it to its batteries are sufficient to provide the required inductance and capacity. Thus[1] the circuit of Fig. 168 is the equivalent of that of Fig. 157. The wires shown in heavy lines constitute the grid and plate inductance, while the capacity is constituted by the internal grid-plate capacity of the tube. A large condenser shunts the plate-circuit energizing battery, which is, like the filament heating battery, connected in series with choke coils, in order to confine the high-frequency currents to the oscillatory circuit. In making the plate and grid

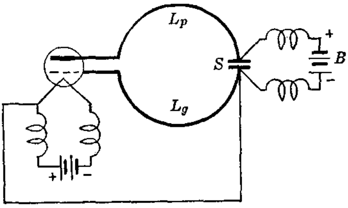


FIG. 168.

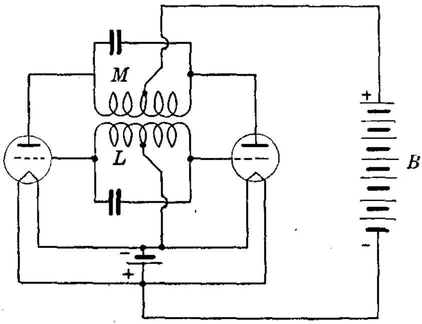


FIG. 169.

wires about 1 ft. long, the generated current has a fundamental frequency of the order to 80,000,000 cycles per second, and contains fairly strong second, third, and fourth harmonics.

Another arrangement,[2] shown in Fig. 169, makes use of two tubes operated symmetrically. The filaments being connected together, the grids of the two tubes are connected to each other by an inductance coil L while the plates are connected by a coil M wound oppositely to coil L. Each coil being shunted by a condenser, the middle point of the grid coil L is connected to the filaments, while the middle point of the plate coil M is connected to the positive terminal of the energizing battery B, the negative terminal of which is connected to the filaments. The tubes thus

[1] GUTTON and TOULY, *Comptes Rendus Academie des Sciences*, Vol. 168, p. 271, Paris, 1919.

[2] MESNY, R., "Les ondes très courtes," *Onde Electrique*, p. 26, January 1924.

operate 180 deg. out of phase with each other, the grid of one tube being positive when that of the other is negative, and conversely.

For very high-frequency generation, the condensers reduce to the internal tube capacities, while the inductances M and L are single-turn windings of opposite directions, as shown in Fig. 170, and of a few inches diameter.[1]

Comparatively strong oscillations may be generated by this method, even at very high frequencies, and the arrangement being symmetrical, the wires connecting the batteries to the circuit carry no high-frequency currents, so that the choke coils and by-pass condensers may be dispensed with entirely, and the effects of battery capacity to ground are also eliminated.

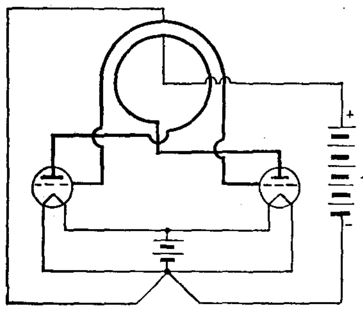


FIG. 170.

Radio Telegraph Transmitting Circuits.—Vacuum-tube oscillators as described above are widely used in undamped (continuous) wave radio telegraph transmission, for setting up high-frequency alternating currents in the antenna circuit. The methods used differ somewhat, however, depending on whether the apparatus is that of a stationary high-power station, or of a small low-power set, a portable field, or airplane radio set, for instance.

High-power Sets.—In the instance of high-power sets, a comparatively low-power vacuum-tube oscillation generator of one of the forms described above is used, operated under conditions which insure as pure a sinusoidal output current as possible, even though this does not permit the most efficient operation. But since the oscillator is of low power, efficiency considerations do not matter much at this point. The high-frequency alternating electromotive force thus generated serves as the input electro-

[1] MESNY, R., *loc. cit.*, *Onde Electrique*, p. 28, January 1924.

motive force of a power amplifier which, in its last stage, feeds the antenna circuit. This power amplifier is operated at maximum efficiency, which involves the use of high plate potentials, negatively charged grids, and, in the last stages, filter circuits for eliminating the harmonics which, as explained before, are unavoidable when the vacuum-tube fundamental frequency output is made very large.

Keying, the function of which 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, may be effected in two different manners, depending on whether the key, when open, entirely stops the oscillations in the antenna, or simply changes the wave length of these oscillations.

The first method is accomplished in placing a simple make-and-break key or relay in the plate or grid circuit of one of the amplifier tubes, in one of the first low-power stages. Opening either one of these circuits stops the operation of the amplifier, and the currents which are interrupted by the key are small (particularly when the key is in the grid circuit of the tube).

The second method requires the key to operate in the oscillation generator circuit. The key is then so connected that, when closed, it short-circuits a part of the inductance of the oscillatory circuit (or connects a small additional capacity in parallel with the main circuit condenser), thus changing the frequency of the generated oscillations. With the modern receiving methods, the amount of *detuning* required is extremely small, so that the antenna current and radiated energy are practically the same at the signal and detuned frequencies.

This permits the simultaneous transmission of several messages over the same antenna circuit, utilizing for all messages the same oscillation generator, and each message being transmitted with the full power of the station.[1] Thus consider the case of two messages being transmitted simultaneously. Two keys will be installed, in such a manner that, when closed, they will each detune the oscillation generator by different amounts, the frequency of the generated oscillations being, say, *A* when the first key is closed alone, *B* when the second key is closed alone, and *C* when the two keys are closed together. Radiation at these three frequencies will be practically the same if these frequencies differ only by a small percentage. At the receiving station, three

[1] Abraham, H. and R. Planiol, *Onde Electrique*, pp. 381–386, July 1922.

receiving circuits are used, tuned respectively to the frequencies A, B, C, and so arranged that one recorder will be operated equally well from the receivers A and C, while another recorder is operated equally well from the receivers B and C.

Low-power Sets.—Low-power sets may, like the higher-power sets, comprise a "master oscillator" supplying the input electromotive force to a small power amplifier. But they may also simply employ a vacuum-tube oscillator for the direct excitation of the antenna circuit.

Thus, Fig. 171 is essentially the same circuit as that of Fig. 156, 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 feed-back coupling adjustment over a fairly wide range of frequencies. The only adjustment then required is that of tuning the antenna circuit so that it will have a natural frequency equal to that which is to be transmitted. This is done by means of a variable inductance or capacity inserted in the antenna circuit. The key may be inserted at point a in the plate circuit, or b in the grid circuit, the opening of the key opening the plate or grid circuit, as the case may be, and stopping the operation of the oscillation generator, which then resumes when the key is closed. The detuning method of keying may also be used by connecting the key across one or two turns of the antenna inductance.

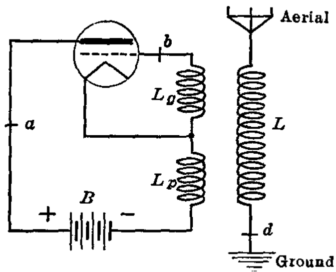


FIG. 171.

The advantage of this circuit is its simplicity, the only adjustments required being that of the antenna wave length and 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 continuous 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, as explained in detail in a previous section.

Now, in the present case, the oscillatory circuit continually loses energy by radiation into space. In fact, the greater the amount of energy this circuit radiates, the better it performs its function as a transmitting antenna. It follows that, on account of the large energy losses in the antenna circuit, this transmitting circuit requires, for satisfactory operation, a vacuum tube of large power capacity and large amplification factor. 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.

In order to overcome this tendency of the circuit of Fig. 171 to refuse to oscillate, an intermediate oscillatory circuit may be used, as shown in Fig. 172. Temporarily considering the antenna circuit as non-existent, the vacuum-tube oscillator circuit is seen to be the same as that of Fig. 147. For suitable adjustments, strong continuous oscillations are then set up in the plate oscillatory circuit L_pC. If now the antenna circuit, tuned to the frequency of the circuit L_qC, is coupled *loosely* to it, a small alternating current will be set up in it, involving correspondingly small radiation. As the coupling is made tighter, an increasing amount of energy is removed from the closed oscillatory circuit L_pC and radiated by the antenna, and the coupling may thus gradually be increased until as much power is radiated as the tube capacity permits. A further increase of antenna coupling will then prevent the generation of oscillations, as in the case considered before. This method permits the use of tubes of as low power as desired (less than 10 watts alternating-current power capacity), a desirable feature for short-range sets used in crowded districts.

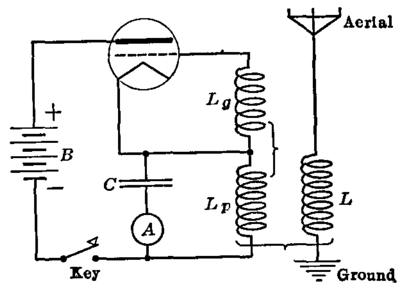


FIG. 172.

It should be noted however that in many cases, as will be explained presently, the antenna may not be coupled to the circuit L_pC as closely as just stated. Suppose once more the antenna to be *loosely* coupled to the vacuum-tube oscillator circuit. The key being then closed, with the circuit L_pC set to the desired wave length, and the grid-plate coupling adjusted for maximum amplitude of the generated oscillations, the antenna circuit may be tuned in the usual way, by adjusting its variable condenser, and the current in the antenna circuit will pass through a well-defined maximum when the antenna and tube circuits have the same natural frequency.

If the antenna, instead of being loosely coupled, is coupled *closely* to the vacuum-tube oscillator circuit, the process becomes more complicated due to the fact, mentioned in a preceding chapter, that the antenna circuit and the oscillatory circuit L_pC now constitute a system having *two* natural frequencies, irrespective of whether or not the two circuits are, individually, tuned to a same natural frequency.

It was shown in Chap. II that such a system oscillates freely at its two natural frequencies simultaneously: these are in every respect equally possible and the resulting oscillation current varies in amplitude at a certain "beat" frequency, as explained before. However, in the case of the three-electrode vacuum-tube oscillator, oscillations of *one* frequency only are generated, due to the fact that, although the oscillations of the system may be considered as "free" (since they are not due to any external alternating electromotive force, but are generated through the action of the circuit itself), certain limiting factors and conditions are introduced by the internal operating characteristics of the tube, and by the manner in which the tube is related to the oscillatory circuit system (that is, by the values and nature of the plate and grid couplings), which make one oscillation frequency more probable or possible than the other.

Thus, with the tube oscillating and the antenna circuit closely coupled to it, if an attempt be made to tune the antenna circuit to the oscillator circuit, (by adjusting its variable condenser in such a manner as to vary its natural frequency ω from a large value to a small one), the current in the antenna circuit, as measured by a hot-wire ammeter inserted in it, will be found to increase, somewhat as shown by the curve AA (Fig. 173) even after the resonance frequency ω_0 (for which the natural frequencies of the two circuits, considered individually, are equal) has been reached, until the natural frequency of the antenna circuit reaches a value ω_1, smaller than the resonance frequency ω_0, where the antenna current *suddenly* drops to a smaller intensity, continuing then to decrease along curve B as the natural frequency of the antenna circuit is further reduced.

If, on the other hand, the antenna circuit's natural frequency is increased, by varying the antenna condenser in opposite direction, the antenna current will increase as shown by curve B, until a frequency ω_2, greater than ω_0, is reached, when the antenna current intensity *suddenly* falls back to a small value, continuing then to decrease along curve A when the natural frequency of the antenna circuit is further increased.

If finally the antenna circuit's natural frequency has been adjusted to some intermediate value ω_3, the antenna current being then as given by the curve A, the opening and then closing of the telegraph key will bring the current down to the value corresponding to curve B.

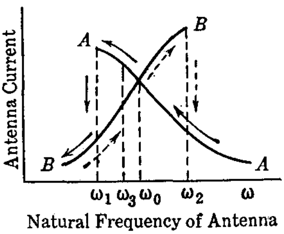


FIG. 173.

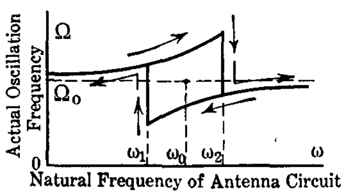


FIG. 174.

The curves A and B may, in other cases, have shapes different from those shown in Fig. 173, as will be seen in a later paragraph.

Corresponding results are obtained if the actual frequency of the generated oscillations is measured (for instance by means of a wavemeter *loosely* coupled to the system), as represented in Fig. 174. Thus, starting with a small value of the antenna circuit natural frequency ω, the oscillation frequency Ω will be greater than the resonance frequency Ω_0 of the individual circuits, and increases until the value ω_2, greater than ω_0 is reached, when the operating frequency suddenly decreases below the value Ω_0. If the natural frequency of the antenna circuit be now decreased, the oscillation frequency decreases until a value ω_1, smaller than ω_0, is reached, when the oscillation frequency suddenly rises to a value greater than Ω_0. These sudden frequency changes, like the sudden amplitude changes, occur after the circuits have been adjusted, through the opening and closing of the telegraph key.

An understanding of these conditions may be obtained by considering[1] the two coupled circuits A, B as equivalent to a *single* (fictitious) oscillatory circuit the inductance, capacity and resistance of which are variable and may be calculated in function of the constants and mutual coupling of the actual circuits A and B, and of the operating frequency.[2] To this equivalent circuit may then be applied all that has been said above concerning the ordinary single-circuit vacuum-tube oscillator.

[1] For a detailed mathematical development of the method described here, see HARMS, *Jahrb. d. draht. Tel. u. Tel.*, Vol. 15, No. 6, pp. 442–457, June, 1920.

[2] In other words, the secondary load created by the coupled circuit B is transferred to the primary circuit A. This reduction of a two-circuit system of fixed electrical constants to a single oscillatory circuit the characteristics of which are functions of the frequency may readily be understood as follows: When an alternating electromotive force of gradually increasing frequency is applied to a simple oscillatory circuit of constant inductance, capacity and resistance, the reactances of the coil and condenser vary in opposite directions, as explained in Chap. II, and become equal to each other at a given frequency value of the applied electromotive force, when resonance is said to obtain.

If, now, a second circuit is coupled magnetically to the coil of the "primary" circuit, the current flowing in this primary circuit induces a "secondary" current in the second circuit, which, in turn, induces an electromotive force in the coil of the first circuit. The electromotive force across the inductance coil of the first circuit thus comprises the self-induced electromotive force due to the primary current (as in the case of a single circuit) and the electromotive force induced in the coil by the secondary current. Now, the amplitude and phase of the electromotive force induced back into the primary inductance coil by the secondary current will, if the secondary circuit comprises inductance, resistance and capacity, depend on the operating frequency. The ratio E/I of the effective electromotive force across the primary inductance coil and current in the primary circuit, which defines the *reactance* of the coil in the case of a single circuit, hence, now defines an *impedance*, since the electromotive force and current in the coil are no longer 90 deg. out of phase with each other. The two-circuit system may therefore be considered as equivalent to a single oscillatory circuit like the original primary circuit but in which the original inductance is replaced by an impedance, as just defined, that is, by a series arrangement of a resistance (which adds to the resistance of the original primary circuit) and an inductance or capacity (which correspondingly alters the total inductance or capacity of the circuit). The electrical characteristics (inductance, resistance and capacity) of the single circuit equivalent to the two-circuit system are, thus, functions of the inductances, resistances and capacities of the two coupled circuits and of the operating frequency, as stated in the text.

In particular, the operating frequency will be substantially equal to the natural frequency of the circuit. Only, as pointed out in Chap. II, the system constituted by the two coupled circuits A and B, and hence also its equivalent single circuit, possesses *two* natural frequencies, and the question arises as to which one of these two frequencies will be the actual operating frequency of the vacuum-tube oscillator.

Now, for each capacity value of the actual secondary circuit B, we may compute the natural frequencies Ω' and Ω'' of the coupled circuit system, and for each of these two frequencies may be calculated the corresponding inductance, capacity and resistance L', C', R', and L'', C'', R'', of the equivalent single circuit. Putting each of these two sets of values respectively into the equation expressing the condition for oscillation generation, as given at bottom of page 182 or top of page 183, it becomes possible to find which set satisfies the operating condition of the tube.

It is thus found that the first set of values L', C', R', which corresponds to the higher frequency Ω', satisfies the equation for all values of secondary (or antenna) circuit capacity *greater* than a certain critical value C_1, while the second set L'', C'', R'', which corresponds to the lower frequency Ω'', and satisfies the equation for all values of secondary circuit capacity *smaller* than some other value C_2.

Now, these two limiting values C_1 and C_2 of the secondary circuit capacity, above or below which one or the other of the two oscillation frequencies becomes a possible operating frequency of the system, depend on the tightness of the mutual coupling of the primary and secondary circuits A and B, on the adjustments of the oscillator tube circuits (mutual couplings of the grid and plate circuits to each other and to the primary circuit A, etc.) and on the internal characteristics of the particular tube used.

Different cases are thus possible, depending on the relative values of C_1 and C_2. Thus, in the cases illustrated in Figs. 175 and 176, the value C_1 is *smaller* than C_2. This creates an overlapping range $C_1 - C_2$ within which both oscillation frequencies Ω' and Ω'' are possible operating frequencies for the vacuum-tube oscillator. On the other hand, if, as in Fig. 177, the value C_1 is *greater* than C_2, operation is impossible at either frequency, within the range $C_1 - C_2$, and the tube will, hence, refuse to oscillate within this range of values of the secondary circuit capacity.

Consider again the instances of Figs. 175 and 176. It should be noted that, while both frequencies Ω' and Ω'' are possible operating frequencies for the vacuum-tube oscillator, for values

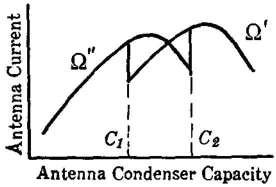


FIG. 175.

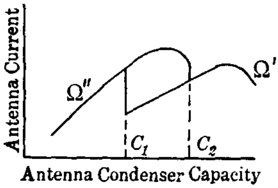


FIG. 176.

of the secondary circuit capacity comprised between C_1 and C_2, these two frequencies are not equivalent to each other, hence, not equally possible, and the amplitude of the generated oscillations will generally not be the same at the two operating frequencies for any given value of the secondary capacity setting, nor will the stability of the two oscillations be the same. This is primarily due to the fact that the circuit system is associated to a vacuum tube, the operating characteristic of which is curved, and should be contrasted with the case of free oscillations, as described in Chap. II, where the two oscillation frequencies are generally equally possible, and the two oscillations take place simultaneously.

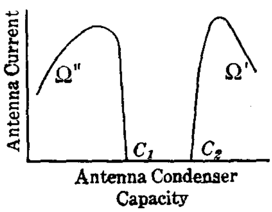


FIG. 177.

In the present case, it may generally be stated that, within the range $C_1 - C_2$, the stable oscillation frequency is the one corresponding to the smaller oscillation amplitude, and this is the oscillation which for instance will set in when the key, after being opened, is closed again. The other oscillation may then be regarded as an unstable continuation of the oscillatory condition of the system, obtained when the secondary circuit capacity is gradually varied or adjusted from some value outside the range $C_1 - C_2$ to some value comprised within this range. And, indeed, this oscillation will cease if some disturbance sets in, such as the opening of the telegraph key, for instance, as stated before.

However, this is not an absolute rule for all cases, particularly when the two possible operating frequencies are comparatively much different from each other, the stable oscillation being then that which most rapidly reaches its final amplitude and substantially in the smallest number of cycles.

The building-up speed of the oscillations thus enters into play, and together with it must be considered the negative resistance which may be developed by the tube, at each of the two possible operating frequencies, during the building-up process of the oscillations. A theory of these phenomena may thus be established by considering the decrements of the circuit system at the two frequencies, which decrements may be regarded as compensated by the "negative decrement" developed by the tube in each case. This standpoint, however, is implicitly included in the above mentioned equations of pages 182 and 183 expressing the condition for oscillation generation.

It will be noted, also, from the nature of the phenomena, that similar conditions will obtain not only for the vacuum-tube oscillator but also for all systems in which direct current (from a battery or generator) is transformed into alternating current or continuous oscillations by means of some variable negative resistance device, such as the carbon arc for instance, associated with a circuit combination having a plurality of "free" oscillation frequencies; or, in other words, for all circuit systems having a plurality of free oscillation frequencies and in which the resistance is a function of the current amplitude in the circuit and of the operating frequency.

As stated before, all these irregularities disappear when the antenna coupling is made sufficiently small, the reaction of the antenna current on the vacuum-tube circuit being, then, negligible.

Of course, similar phenomena take place if the closed oscillatory circuit is being tuned to the antenna circuit instead of the antenna circuit being tuned to the closed circuit. Also, the fact that one of the circuits (the antenna circuit) is an open oscillator does not matter particularly; the property applies to any system of coupled oscillatory circuits having a plurality of oscillation frequencies.

MULTIVIBRATOR AND NEGATIVE-RESISTANCE CIRCUITS

If the input and output circuits of a multistage transformer or choke-coil coupled cascade amplifier are mutually coupled to each other, closing the amplifier cascade or chain upon itself, continuous oscillations will be generated for a proper polarity and degree of the coupling. The frequency of the generated oscillations is, then, determined by the inductance and distributed capacity of the amplifier interstage transformer or coupling coil windings.

Conditions are different in the case of a resistance-capacity or resistance-battery coupled amplifier, for the interstage coupling elements do not then constitute oscillatory circuits. There being, however, important applications of the properties of such resistance coupled cascade amplifiers in which the output electromotive force of the last stage is made to excite the input circuit of the first stage, these will be briefly discussed below.

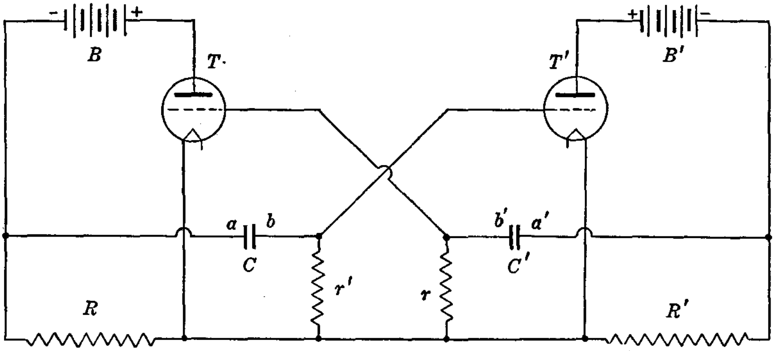


Fig. 178.

Multivibrator.[1]—Consider a resistance-capacity coupled cascade amplifier, such as described in connection with Fig. 137, and suppose that the electromotive force developed in the plate circuit of the *second* tube, instead of being impressed upon the grid circuit of a third tube, is impressed back upon the grid circuit of the first tube. The resulting circuit arrangement, shown in Fig. 178, comprises then two tubes T and T' the plates circuit of which comprise, respectively, the batteries B and B' and resistances R and R'. The filament and grid of each tube are connected to the terminals of the plate circuit resistance R and R' of the other tube, this connection being, further, made through condensers C and C' which prevent the high continuous potential drop across the resistances R and R' from being impressed onto

[1] Abraham and Bloch, *Ann. de Physique*, 9° serie, XII, p. 237.

the grids. The latter, being then insulated from the rest of the circuit by the condenser dielectric, are conductively connected to the filaments through resistance r and r', as explained in the preceding chapter.

If then the resistances R and R', resistances r and r', condensers C and C', batteries B and B', and tubes T and T' are, respectively, identical to each other, the system will be in equilibrium if the plate currents in the two tubes have the same intensity. This state of equilibrium is unstable, however, as will be shown presently. The plate current of tube T, flowing through the resistance R, produces a potential difference between the ends of the latter directly proportional to the plate-current intensity and of such polarity as to make the filament end of the resistance positive with respect to the battery end. Now, the armatures of the condenser C being connected to the ends of the resistance R, the condenser C becomes charged to this potential difference, and when the system is in equilibrium, the armatures a and b of the condenser C are at the same potentials, respectively, as the battery and filament ends of the resistance R. There is thus no potential difference across the resistance r' connecting the armature b to the filament end of the resistance R, and, consequently, the grid of the tube T', being connected to the filament through this same resistance r', is, hence, also, substantially at the same potential as the filament. The same remarks apply to the corresponding circuit elements of the other tube.

Now suppose that in some manner the plate current of tube T is varied slightly from its equilibrium intensity: decreased, for instance. The potential difference across R decreasing correspondingly, the condenser C will discharge in order that the voltage across its armatures may remain equal to that across the resistance R. This means that an electric current will flow in the condenser circuit in the direction $br'Ra$, and a potential difference will be developed across the resistance r' the condenser end b of which will become positive with respect to its filament end. But the grid of tube T' being connected to the filament through this same resistance r', a positive potential will thus be applied to it, and the plate current of tube T' will correspondingly increase.

This in turn increases the potential drop across R'. The condenser C' being connected to the ends of the latter, a charging current will flow in its circuit in the direction $a'R'rb'$ in order to raise the condenser voltage to the new value of potential difference across R'. But this current, flowing through the resistance r, produces a potential difference across it, which makes its condenser end b' negative with respect to its filament end. And the grid of tube T being connected to the end b' of the resistance therefore becomes also negative, which produces a decrease of the plate-current intensity in tube T.

The original small decrease of plate current in tube T is thereby made larger, and the process continues, the plate current decreasing in tube T and increasing in tube T', until limited by the lower and upper bends of the tube, characteristic curves being reached. But a stoppage of the operation producing also a stoppage of the condenser currents, hence, of the current flow through the resistances r and r', there will no longer be any potential difference across the ends of these resistances, and

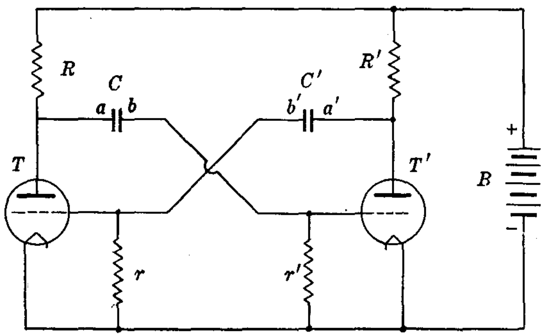


FIG. 179.

the grids of the tubes will suddenly be at the same potential as the filaments. This, in turn, produces a corresponding increase of the current in the tube T and decrease of the current in the tube T', so that the operation described above repeats itself in opposite direction.

The system thus continuously oscillates, the condensers C and C' alternately charging and discharging through their resistances Rr and $R'r$. The *frequency* of the oscillations is determined by the values of the capacities C,C' and resistances $(R + r)$ and $(R' + r')$ of each of the circuits CRr, $C'R'r'$. The *amplitude* depends on the ratio $C/(R + r)$ and also on the "feed-back ratio" R/r of the system.

Instead of using the circuit of Fig. 178 it is generally employed under the form shown in Fig. 179, which differs from the preceding one simply in that the plate circuits of the two tubes are energized by a single common battery B. This battery, it will be noted, is then included in the condenser circuits, but this does not alter the mode of operation, which is the same as described above.

A property of these devices is that the generated oscillations are not sinusoidal but, on the contrary, contain a very large number of multiple-frequency components or harmonics of large amplitude, hence the name of *multivibrator* given to the devices. This property finds application, particularly, in the calibration of wavemeters. Thus, a multivibrator of low fundamental frequency (say of the order of 1,000 cycles per second) may have its fundamental frequency adjusted to equal exactly the oscillation frequency of some standard pendulum or tuning fork—or better, the standard pendulum or tuning fork may, in some suitable manner, be made to generate electrical impulses of its own frequency and drive the multivibrator fundamental frequency into step with its own frequency. The fundamental frequency of the multivibrator having then an accurately known value, its harmonic frequencies are at once known also, being exact multiples of this fundamental frequency, and may thus be used for exciting the wavemeters under calibration.

Negative Resistance Circuits.[1]—Consider a resistance-battery coupled cascade amplifier, such as shown in Fig. 136, and suppose that the electromotive force developed in the plate circuit of the *second* tube, instead of being impressed upon the grid circuit of a third tube, is impressed back upon the grid circuit of the first tube. The resulting circuit arrangement, shown in Fig. 180, comprises, then, two tubes T and T' the plate circuits of which comprise, respectively, the batteries B and B' and resistances R and R'. The filament and grid of each tube are connected to the terminals of the plate circuit resistances R or R' of the other tube, this connection being, further, made through batteries C and C' of such voltage and polarity that the grid voltage of the tubes (respectively equal to the direct-current potential drop across R or R' plus the voltage of the battery C or C') is substantially equal to zero, as measured with respect to the filaments.

If, then, the corresponding circuit elements of the two tubes are respectively identical, the system will be in equilibrium if the plate currents in the two tubes have the same intensity. But this state of equilibrium is unstable, as will be understood presently.

[1] British Patent 148582 to Eccles and Jordan.

Thus, suppose the operating equilibrium to be disturbed, even only very slightly and temporarily, for instance, by giving a slight positive impulse to the grid of tube T, thereby increasing the plate current of this tube. This will increase the potential drop across the resistance R and, hence, make the grid of tube T' more negative. This, in turn, decreasing the plate-current intensity in tube T', reduces the potential drop across the resistance R' and renders the grid of tube T more positive. The small initial posi-

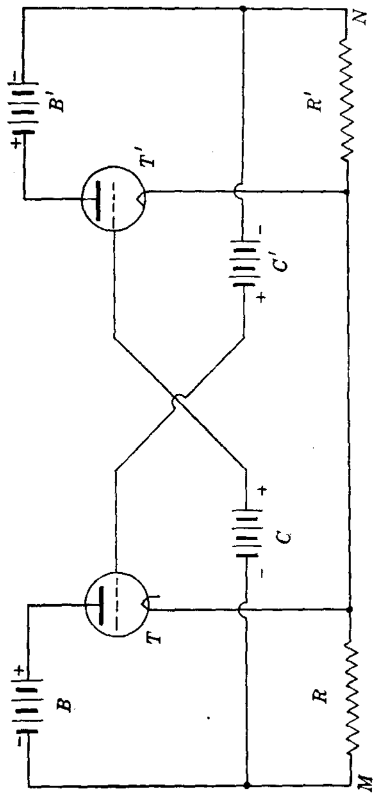


FIG. 180.

tive impulse on the grid of tube T is thereby reinforced, and the process continues until the plate currents in the tubes T and T' reach a maximum and minimum value, respectively, as determined by the bends of the characteristic curves. The operating condition thus reached constitutes a *stable* equilibrium.

If, however, an impulse in the opposite direction reverses this equilibrium, even if only slightly, then the process described above repeats itself in opposite direction, until the plate current in tubes T and T' are, respectively, a minimum and a maximum, a stable equilibrium being then reached.

Between these two stable operating conditions, the device operates like a negative resistance and may be used for amplifying small electromotive forces. It is simply necessary that the electromotive force to be amplified be sufficiently large to *start* the reversal of operating conditions, the reversal proceeding then all

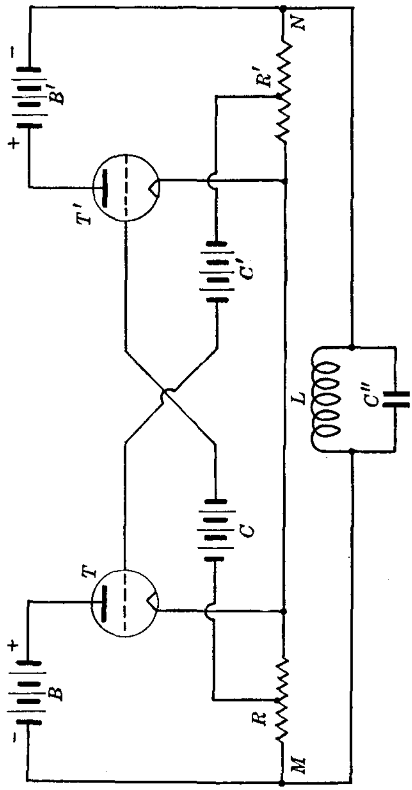


Fig. 181.

by itself until the power capacity of the tubes is reached. The minimum electromotive force required for such operation may be reduced by connecting the grids of the tubes to an intermediate point of the resistances R and R', as shown in Fig. 181.

The amplifying property of the device may be used in many different manners. Figure 181 illustrates its use for oscillation generation in the circuit LC''. A small oscillation started in this circuit sets up an alternating potential difference between points M and N and, hence, produces an alternating-current flow in the circuit branch $MRR'N$, correspondingly varying the grid potentials of the tubes. An amplified and in-phase reproduction of this electromotive force is then developed between points MN, through the operation of the device as described above, which serves to build up and sustain the oscillations in the circuit LC''.

CHAPTER IX

THE THREE ELECTRODE VACUUM-TUBE RECEIVER

THE SIMPLE THREE-ELECTRODE VACUUM-TUBE DETECTOR

Two-electrode Vacuum-tube Detector.—The operation of the three-electrode vacuum tube detector may be better understood by first considering the two-electrode or Fleming valve detector.[1]

The circuit used, as shown in Fig. 182, is the same as that of Fig. 87, except that the crystal detector is here replaced by the plate-filament space of a two-electrode vacuum tube. A battery *A* is used for heating the filament, but no battery is connected

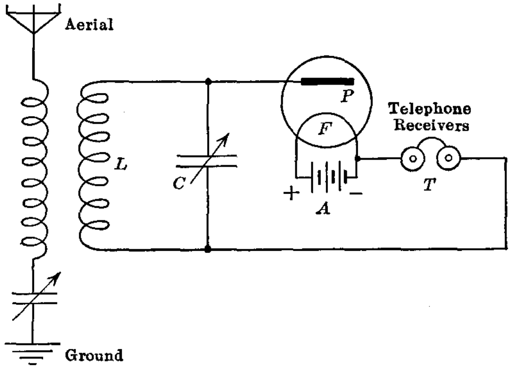


FIG. 182.

between the plate and filament, so that normally there is no electron flow from filament to plate, hence no current in the plate circuit *PLTFP*.[2]

When a signal is being received, the incoming waves set up a high-frequency alternating current in the antenna circuit, and a

[1] This constitutes, historically, the first application of thermionic phenomena to radio: British Patent 24850, November 1904 to J. A. Fleming.

[2] This is true only when the plate is connected to the negative end of the filament, as shown in the figure. 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 this negative portion of the filament.

high-frequency alternating electromotive force is developed across the secondary coil L and condenser C, and is applied between the plate and filament of the tube. The plate then becomes alternately positive and negative, and an electric current flows through the tube during those half-cycles (and those only) when the plate is positive with respect to the filament.

This current, as shown in the second curve of Figs. 84 and 110, is thus unidirectional and pulsating, and may be considered as composed of a direct current, equal to its average value, superimposed on high-frequency alternating current of frequency equal to that of the incoming oscillations.

If, as in continuous-wave telegraphy, these oscillations are unmodulated, that is, of unvarying amplitude during a dot or a dash, the direct and alternating components of the rectified or "detected" current are also constant, and no sound is produced in the telephones. The direct-current component may, however, actuate a telegraph relay connected in place of the telephone receiver, and serve to record the received signals.

If, on the other hand, the incoming waves are of variable amplitude (damped or modulated waves), the direct and alternating components of the rectified current vary correspondingly, and the telephone receivers will be actuated and produce a sound.

Of course, in order that the high-frequency electromotive force developed across the coil L may be applied upon the detector tube, the telephone winding must be shunted by a small capacity by-pass condenser, as explained in the case of the crystal detector. The high-frequency alternating component of the detected current then passes through this condenser, while the direct current and the low or audio-frequency current components pass through the telephone windings.

Three-electrode Vacuum-tube Detector. Grid Detection.— Instead of connecting the telephone receiver directly in the detector circuit, as shown in Fig. 182, the received signals may, after detection, be amplified by means of a three-electrode vacuum tube. A high impedance (resistance or low-frequency choke coil) is then connected in the detector circuit in place of the telephone receiver (shunted by a by-pass condenser S for permitting the high-frequency electromotive force to be impressed upon the detector D, as explained before). The direct and low-frequency components of the detected (rectified) current, passing through this high impedance, set up a potential difference between its ends, constituting the output electromotive force of the detector circuit. This output electromotive force is then applied between the grid and filament of a three electrode vacuum tube operating as a low-frequency amplifier and energizing the telephone receiver connected in (or coupled to) its plate circuit. For this purpose the grid and filament of the three-electrode tube are respectively connected to the ends of the detector-circuit impedance.

Thus (Fig. 183) the antenna circuit *AG* is coupled to the tuned secondary circuit *LC* across which is connected, as before, the two-electrode detector tube[1] *D* in series with the high impedance *Z* (connected in place of the telephone receivers) shunted by the high-frequency by-pass condenser *S*. One end of the impedance *Z* is connected to the grid of the three-electrode low-frequency

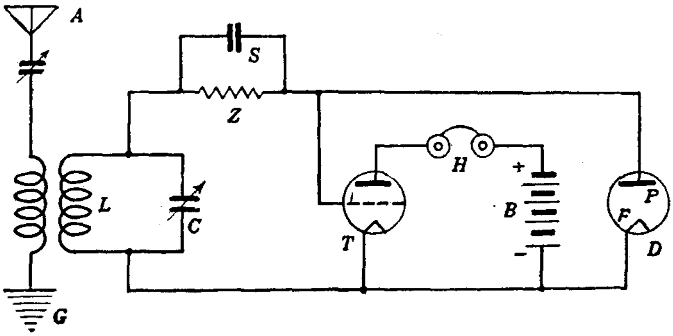


FIG. 183.

amplifying tube *T*, the other end being connected to the filament of this tube *T*. In the figure, this latter connection is made through the coil *L* which, having low resistance and small low-frequency reactance, is practically equivalent to a direct connection between the filament and impedance *Z*. The plate circuit of the tube *T* contains the energizing battery *B* and telephone receiver *H*.

As explained in the preceding section, incoming waves set up a high-frequency alternating current in the antenna circuit, inducing a high-frequency electromotive force in the tuned secondary circuit *LC*, which is impressed between the plate and filament of the detector tube *D*, through the by-pass condenser *S*. The resulting unidirectional rectified electric current (as opposed to the *electron* current) then flows through the detector circuit

[1] Filament battery omitted for simplicity.

in the direction *FLZPF*, giving rise to a potential difference between the ends of the impedance element *Z*, hence, also between the grid and filament of the tube *T*, and producing corresponding variations of the plate current of tube *T*.

From the direction of current flow *FLZPF* as just mentioned, and with the connections as shown in the figure, the *ohmic-resistance potential drop* across the impedance *Z* produced by the rectified current is of such polarity as to make the grid of the tube *T* negative with respect to the filament by an amount which is a direct function of the amplitude of the received signals, which results in a corresponding *decrease* of the plate-current intensity in tube *T*. The *reactive potential drop* across the impedance *Z*, on the other hand, is proportional to the amplitude *variations* of the received signals. Applications of these remarks will be described further below.

Now, it will be noted that the grid and filament of the tube *T* are respectively connected to the plate and filament of the detector tube *D*. Like the plate of the tube *D*, the grid of tube *T* becomes alternately positive and negative with respect to the filament when a signal is being received, and the grid-filament combination of the three-electrode tube *T* will operate in the same manner as the plate-filament combination of the two-electrode tube *D*, and like the latter constitutes a rectifying detector. The two-electrode tube *D* may hence be done away with, its function being accomplished in exactly the same manner by the grid and filament of the tube *T*, which thus operates simultaneously as a detector and a low-frequency amplifier. The only difference is, that the rectified current being now much smaller, the impedance *Z* must be made correspondingly larger in order that the rectified current may produce across it a potential difference of sufficient magnitude. In other words, the internal grid-filament resistance of the tube being, in general, greater than that of the usual two-electrode vacuum-tube detector, the impedance *Z* must be increased also in order to be, as required, of the same order of magnitude as this internal grid-filament resistance—generally from 1 to 5 megohms, depending upon the particular type of tube used.

The circuit thus reduces to that of Fig. 184. The impedance *Z* is here shown as an ohmic resistance (constituted, for instance, by a strip of paper coated with india ink or graphite, or a carbon or metal film deposit on a glass rod, etc.). The rectified current flows, in the *external* grid circuit, from the filament through the coil L, resistance Z, on to the grid, and produces a potential drop across Z which, as stated before, makes the grid end of this resistance Z more negative than its filament end. A received signal thus produces a *decrease* of the plate-current intensity, proportional to (or a direct function of) the received-signal intensity, and, hence, variable in the case of modulated or damped-wave reception, producing in this case a sound in the telephone receiver connected in the plate circuit of the tube.

Summarizing the above, rectification of the received signal current takes place in the grid circuit of the tube. The grid must hence be connected, through the resistance Z and tuned circuit

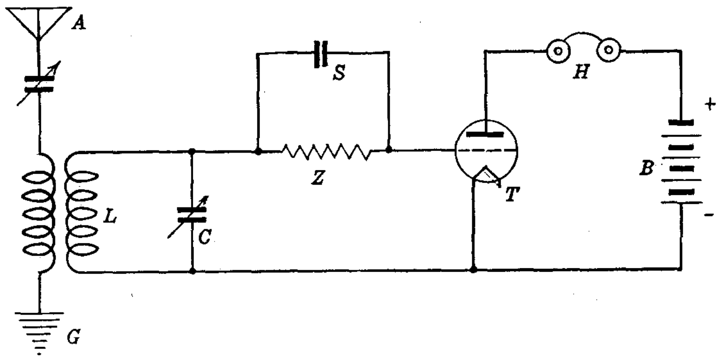


FIG. 184.

LC, to the negative end of the filament, in order that the incoming oscillations may make it alternately positive and negative and become rectified. On the other hand, the tube operating also as an audio-frequency amplifier, its plate-battery potential must be so chosen that it will operate over the straight portion of its plate-current, grid-voltage characteristic curve. And since all received signals produce a decrease of the plate-current intensity, maximum operating range will be obtained if the operating point of the tube, when no signals are being received, is around the upper part of the straight portion of the curve, which can be obtained by utilizing a fairly low plate potential. A somewhat higher plate voltage may also be used, but this requires a slight positive potential to be given to the grid, such as obtained in connecting it to the positive instead of the negative end of the filament, as just stated, which brings the grid-operating point near the sharpest bend of the grid-current, grid-voltage characteristic curve.

On the other hand, instead of an ohmic resistance, the impedance Z may be constituted by a low-resistance coil of large inductance[1] (low-frequency iron-core choke coil), as shown in Fig. 185. The circuit is then limited to the reception of damped or modulated waves.[2] Thus, the rectified current, which flows in the grid circuit when incoming oscillations make the grid alternately positive and negative, consisting of a succession of unidirectional high-frequency impulses of slowly variable amplitude (due to the audio-frequency modulation), may be considered as constituted of a high-frequency alternating component, which is by-passed

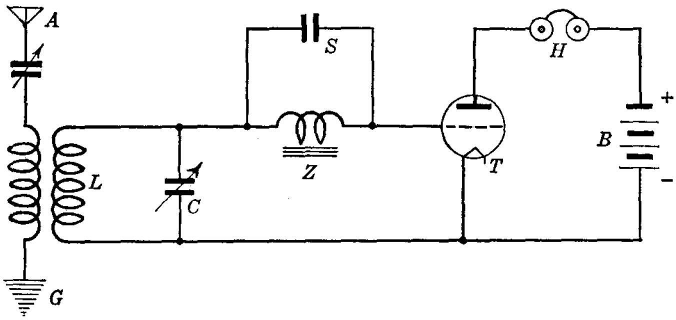


FIG. 185.

by the condenser S, a direct-current component, which produces no appreciable potential drop across the 1,000 or 2,000 ohms resistance of the coil Z, and an audio-frequency alternating component which develops an alternating potential difference of same frequency across the coil Z. The resulting grid-potential variations operate the tube, which therefore operates as a low-frequency amplifier. Unlike the preceding case, the grid-potential variations being here *alternating*, the operating point of the tube, in the absence of any received signals, must preferably be in the center part of the straight portion of the grid-voltage, plate-current characteristic curve.

[1] BARTHELEMY, R., "La detéction par self," in "*La T. S. F. moderne,*" Paris, November 1924.

[2] Under modulated waves are comprised the voice-modulated waves of radio telephone sets, as well as continuous waves modulated at some audible or superaudible frequency, or continuous waves modulated at the receiving station through some heterodyne method.

NOTE.—The operation of the three-electrode vacuum tube as a detector through grid rectification and employing a shunted resistance in the grid circuit (Fig. 184) may also be explained in a different manner.[1] Thus, consider the circuit of Fig. 184, and suppose the resistance Z to be infinitely great, or disconnected. When, then, no oscillations are being received, 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 S. The grid then

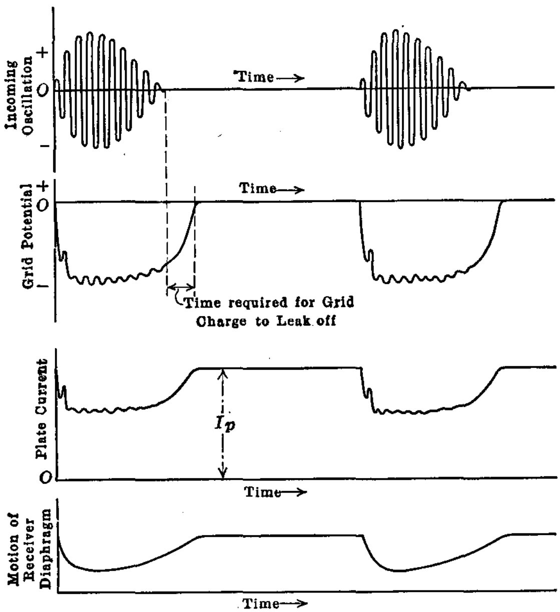


FIG. 186.

assumes the potential corresponding to zero current in its circuit. In the case of a tube free from all gas, as assumed here,[2] this grid potential is zero. When the antenna circuit is then energized by incoming oscillations, say, a succession of damped waves emitted by a spark transmitting set, an oscillatory current is set up in

[1] This second manner is less general, in that it does not lead to the principle of using a low-frequency impedance, such as a choke coil, as done above. But it permits an easier understanding of the "blocking" of the tube for too great a grid-leak resistance.

[2] The case of a tube containing an appreciable amount of gas is discussed by Ralph Bown, *Phys. Rev.*, Vol. X, No. 3, p. 253, September 1917.

the circuit LC, and an alternately positive and negative electromotive force appears at the terminals of the coil L. This alternately charges the grid positively and negatively through the condenser S. In the positive half-cycle, the grid attracts some of the electrons present in the tube. In the negative half cycle, however, it does not lose these electrons again,[1] 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 of a tube containing gas, this charge automatically leaks off through the gas of the tube. In the case of a highly evacuated tube, it is necessary to shunt the condenser S with a high resistance Z of 1 to 5 megohms in order to provide a leakage path for the charge. The operation of the tube is illustrated by the curves of Fig. ,186 which are self-explanatory.

Plate Rectification.—The three-electrode vacuum tube may also be used as a detector by taking advantage of the fact that the

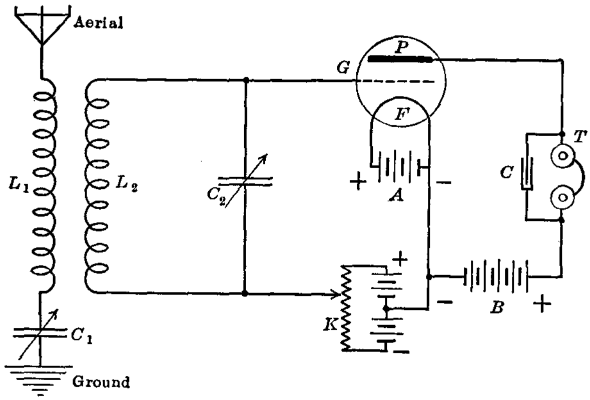


FIG. 187.

plate-current, grid-voltage characteristic curve is not a straight line, especially at its upper and lower ends. Thus, consider a three-electrode tube connected, as shown in Fig. 187, 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

[1] This is because the grid is a cold electrode, and only a heated body, like the filament, can emit electrons.

energized by the battery B which sets up 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 potential divider. By means of the latter, the grid potential may be adjusted to such a value E_g that, under the steady conditions prevailing when no oscillations are being received, the operating point of the tube is some point M or N of either bend of the characteristic curve[1] (Fig. 188).

Thus, assume that when no oscillations are received the plate potential furnished by the battery B and the grid potential E_g result in a steady current I_p flowing in the plate circuit, the oper-

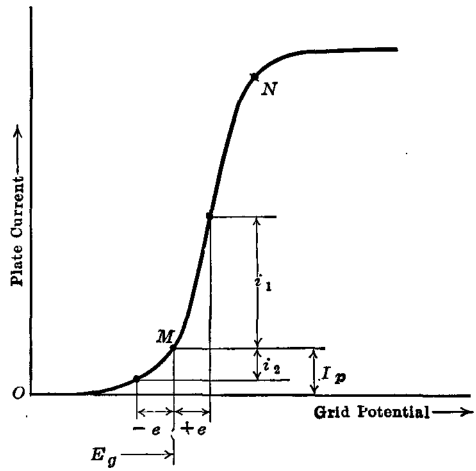


Fig. 188.

ating point M being on the lower bend of the curve (Fig. 188). If an oscillation is then received by the antenna, an oscillatory current will be set up in the circuit L_2C_2 (Fig. 187) and an alternating electromotive force will appear across the coil L_2. This alternating electromotive force is impressed between the filament and grid of the tube, producing at every cycle substantially equal and opposite variations $\pm e$ of the steady grid potential E_g. Plotting these in Fig. 188, the resulting variations i_1 and i_2 of the plate

[1] If the plate potential furnished by the battery B is suitably chosen, it is possible to do away with the grid battery and potentiometer. Thus, referring to Fig. 126, it is seen that for a plate potential of 100 volts, the lower bend of the curve corresponds to a grid potential of about zero, for the particular tube considered, which makes the use of a grid battery unnecessary.

current I_p are seen to be 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. That is, whereas the operation of the three-electrode vacuum tube as an amplifier about the center portion of the straight part of its characteristic curve simply results in a substantially undistorted amplified reproduction of the grid electromotive force, asymmetrical amplification as used here gives rise, in addition to the high-frequency plate current

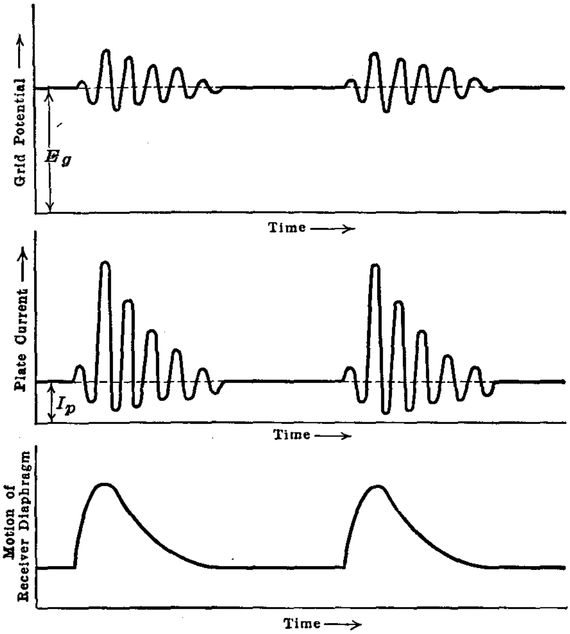


FIG. 189.

component, to a direct-current (rectified) component proportional to (or direct function of) the amplitude of the received signals, hence variable if these are damped or modulated, producing then a sound in the telephone receiver.

These conditions are illustrated in Fig. 189. The upper curve shows the symmetrical variations of the grid potential about its normal steady value E_g under the effects of the incoming oscillations. The resulting plate current corresponding to the various instantaneous values of grid potential is found by referring to the characteristic curve of Fig. 188, and the plate-current curve of Fig. 189 is thus obtained, showing the symmetrical variations of grid potential producing 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.

Obviously 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 intensity, instead of a large increase and small decrease, as in the case studied above. The operation about the lower bend is generally a better practice, as 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.

An advantage of the plate-rectification method is that, being not based on the use of any current in the grid circuit, the steady plate and grid voltages may be so chosen that the grid remains always negative with respect to the filament during reception of the signals and absorbs no power whatever from the tuned receiving circuits, which are hence in no way damped by the presence of the detector. This advantage is offset, however, by the fact that a grid-potential adjusting battery is required, which complicates the apparatus. The grid-rectification method described before is therefore more widely used, particularly as the damping of the circuits resulting from the grid current may, if required, be compensated for through regenerative amplification, as explained in the following sections.

THE NON-OSCILLATING REGENERATIVE DETECTOR

Simple Regenerative Amplification.—In the preceding discussion of the simple three-electrode vacuum-tube detector, the detected or rectified current was seen to comprise a direct and low-frequency component, and a high-frequency component. Only the direct and low-frequency components were used, while the high-frequency component appeared as useless and was either by-passed by a suitable shunt condenser, or choked out by the reactance of the telephone receiver winding, or other circuit components.

The high-frequency component may, however, serve for the radio-frequency amplification of the received signals. Referring to Chap. VIII, it is simply necessary to provide a negative "feed-back" coupling between the plate and grid circuits of the tube, as in the case of a three-electrode vacuum-tube oscillation generator, but making this coupling loose enough so that the tube will not of itself generate continuous oscillations.

When signals are being received, the high-frequency plate-current component then induces an electromotive force in the input circuit of the tube, through this feed-back coupling, which is in phase with the high-frequency electromotive force set up in this input circuit by the incoming signals, and these are thereby amplified. Any one of the vacuum-tube oscillation generator circuits of Chap. VIII may be used for this purpose (provided the feed-back coupling is loosened as just stated).

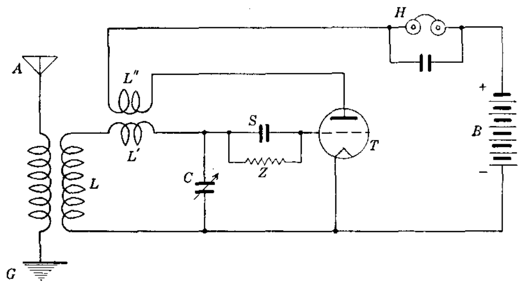


Fig.. 190.

Thus for instance, the circuit of Fig. 190 comprises an antenna circuit AG, coupled to a vacuum-tube detector circuit which is essentially the same as that of Fig. 184, except for the fact that coils L' and L'' are inserted in the grid and plate circuits, respectively, and loosely coupled to each other. Considering then, for example, a damped incoming oscillation, this sets up in the antenna circuit an oscillating current of same frequency and general damping characteristics. Through the coupling between the antenna and secondary tuned circuit $LL'C$, energy is transferred to the latter, building up a damped, high-frequency oscillation in it and operating the detector tube as described before. In addition to the low-frequency impulse, this produces in the plate circuit of the tube a high-frequency alternating current which, flowing through the coil L'', induces an electromotive force in the coil L' which, for a negative coupling polarity of the two coils, is in phase with the oscillating electromotive force operating in the circuit $LL'C$. Energy is thus synchronously supplied by the plate circuit to the oscillatory grid circuit, partly compensating for the resistance losses in that circuit, and thereby increasing the amplitude of the oscillations, and hence, also, the strength of the signals heard in the telephones.

The circuit here described was chosen merely for the simplicity of explanation. Other circuits, such as that of Fig. 159, are actually better suited, as they permit a very smooth and fine adjustment of the feed-back coupling, enabling the tube to be operated very near to its oscillation point, without reaching it, however.

As explained in the case of the three-electrode vacuum tube oscillation generator, in Chap. VIII, the regenerative operation of the tube is equivalent to inserting in the oscillatory grid circuit a negative resistance which partly neutralizes the positive ohmic resistance of the circuit. This neutralization grows as the feed-back coupling between plate and grid circuits is tightened. Now, as the resistance of the circuit is thus being decreased, the damping of the circuit is also lessened, and a point is reached when this damping is so small that the oscillations set up in the circuit by incoming waves can no longer follow the modulations accurately enough to permit their faithful reproduction in the telephone receiver.[1] This, together with the fact that for a tighter coupling the tube breaks into oscillation, limits the amount of amplification which may usefully be obtained by this method.

Superregenerative Amplification.—As just explained, the amplification of received signals through the regenerative method cannot be increased indefinitely by tightening the feed-back coupling, due to the fact that a point is reached when the negative resistance developed by the tube becomes greater, in absolute value, than the positive ohmic resistance of the circuit, in which continuous oscillations are then generated. The so-called super-regenerative method of reception,[2] however, by preventing such oscillation generation in a manner described further below, permits the feed-back coupling to be given a much tighter value, greater than normally required for oscillation generation, and

[1] This may be explained in considering that a modulated wave, as set forth in the chapter on radio telephony, comprises a plurality of frequency components, and the smaller the circuit decrement, the sharper its frequency selecting action, so that for a tight feed-back coupling, the circuit singles out one or a few only of the frequency components of the modulated wave, correspondingly distorting the modulations.

[2] ARMSTRONG, E. H., *Proc. Inst.* of Radio Eng., 10, 244, 1922.

extremely powerful amplification of the received signals is thus obtained without the objectionable distortion mentioned in the preceding section.

In order to understand the underlying principle of this method, consider again the regenerative receiving circuit of Fig. 190. For a very loose coupling between the coils L' and L'', the resistance of the tuned input circuit $LL'C$ is positive. If the coupling of L' and L'' is gradually made tighter, the regenerative action of the tube enters into play more and more, introducing a negative resistance in the circuit and reducing its total resistance, until a point is reached when this total resistance becomes zero, after which, for still greater values of the coupling, the circuit resistance becomes negative and the tube starts generating continuous oscillations in it.

But the value of negative resistance introduced by the tube in the circuit $LL'C$ may be adjusted in a variety of other possible manners. Thus, for instance, the feed-back coupling being kept unchanged and fairly tight, the voltage of the plate-circuit energizing battery B may be given a smaller or a greater value. For a small plate-battery voltage the tube then develops only a small negative resistance, and operates simply as a regenerative amplifier. For a large plate voltage, the regenerative action increases, and for still larger voltages the tube starts generating oscillations in the circuit.

Still a different way is to increase or decrease the positive resistance of the grid oscillatory circuit $LL'C$ (by means of a rheostat inserted in it, for instance). The tube then will or will not generate oscillations in this circuit, depending on whether the negative resistance developed by it is greater or smaller, respectively, than the positive resistance of the circuit.

Now, suppose an alternating electromotive force E, of frequency equal to the natural frequency of the grid oscillatory circuit $LL'C$, to be impressed upon the system, such as when a signal is being received. If the total resistance R of the circuit is positive (positive resistance greater than the negative resistance developed by the tube, as determined by the operating conditions), this electromotive force E rapidly establishes in the circuit an alternating current which, since resonance conditions are established (circuit tuned to the frequency of the energizing electromotive force), is equal to

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

In the case of a received signal, this current is, at best, very small, and if the electromotive force E be suddenly removed, the current is rapidly damped out by the positive resistance, as represented by a curve of the form of Fig. 33.

If the resistance R of the circuit is nearly equal to zero[1] (positive resistance of the circuit but little different from the negative resistance introduced in it by the tube), the impressed electromotive force E produces a current in the circuit which ultimately reaches a very great value. But due to the fact that the damping of the circuit is nearly zero, the time required for the current to build up to this large ultimate value is correspondingly great, as explained before, so that during the first few cycles (sometimes as many as 50 or 100 cycles), the current intensity remains small.

Finally, if the resistance R is large and negative (positive resistance of the circuit markedly smaller than the negative resistance developed by the tube), a starting impulse given to the circuit produces the growth of a free oscillation in it, the amplitude of which increases according to an exponential law. That is, exactly as in the case of a circuit having positive resistance (positive damping), and in which a free oscillation was seen (Chap. II) to be damped out at a constant rate, or decrease logarithmically, or in other words, in which every cycle of the free oscillation had an amplitude equal to that of the preceding cycle *less* a certain constant *percentage* or *fraction* determined by the positive resistance of the circuit, the free oscillation of a circuit having negative resistance (and, so to say, "negative damping") builds up at a constant rate, determined by the value of the negative circuit resistance R, so that every cycle of the free oscillation has an amplitude equal to that of the preceding cycle *plus* a certain constant percentage or fraction. This growth of the oscillation continues until the power capacity of the tube is reached, as determined by the curvature of its characteristic curve. It follows that the maximum amplitude of the oscillation is reached sooner or later, depending, respectively, on whether the initial starting impulse is great or small, but the actual process of building up of the oscillation is independent from this initial impulse and determined solely by the negative decrement of the circuit, that is, ultimately, by the negative resistance R of the circuit, other factors remaining constant.

[1] Irrespective of whether positive or negative, as pointed out by David, Dufour and Mesny, *Onde Electrique*, Paris, No. 41, pp. 175–200, May 1925.

Applying these principles to the reception of radio signals, a superregenerative receiving circuit is essentially a regenerative circuit, as shown in Fig. 190, but with a fairly tight feed-back coupling, and in which the resistance is rapidly and alternately made positive and negative, in one of the manners described presently, at a frequency sufficiently greater than the modulation frequency of the received signals as not to disturb or prevent their proper reproduction in the telephone receivers, and sufficiently smaller than the signal frequency (radio frequency) to permit the building up of oscillations during the periods of negative resistance of the circuit.[1]

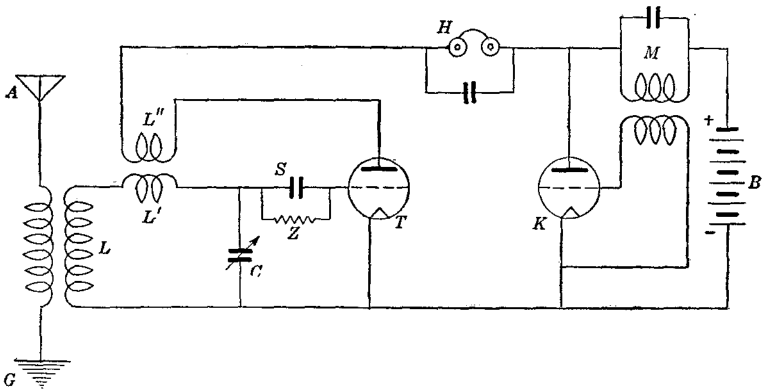


FIG. 191.

Before explaining the operation of such a circuit, a few of the possible manners of effecting this rapid resistance change will be described.

A first method consists in alternately varying the plate voltage of the receiving tube of Fig. 190 above and below the plate-battery voltage value, making the negative resistance developed by the tube alternately greater and smaller than the positive resistance of the oscillatory grid circuit, and hence making the effective resistance R of the receiving circuit alternately negative and positive. For this purpose, a generator of alternating electromotive force of suitable frequency (10,000 to 30,000 cycles) is inserted in the plate circuit of the regenerative detector tube in series with its plate-circuit energizing battery B. The most

[1] For radio telephony reception, this resistance variation frequency is generally chosen between 10,000 and 30,000 cycles per second. For radio telegraphy, it may be made somewhat smaller.

convenient generator is the three-electrode vacuum-tube oscillator, and the circuit of Fig. 190 thus modified becomes as shown in Fig. 191. The only difference is that the plate circuit of the detector tube *T* comprises, in addition to the battery *B*, telephone receivers *H* and feed-back coupling coil *L''*, an oscillatory circuit *M* having a natural frequency of about 10,000 to 30,000 cycles, in which an auxiliary tube *K* generates continuous oscillations, having for this purpose its grid circuit suitably coupled to this oscillatory circuit *M*. The resulting alternating electromotive force developed across the condenser and coil of this circuit *M* thus adds to or subtracts from the electromotive force of the battery *B*, and varies the plate voltage of the receiving tube *T*.

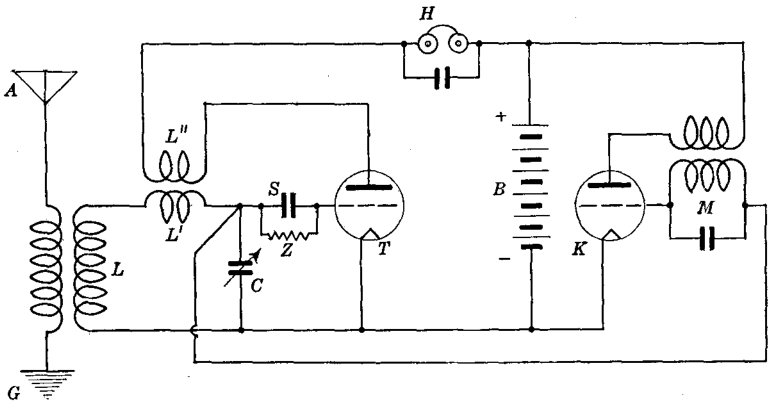


FIG. 192.

The same battery *B* is here shown energizing the plate circuit of the oscillatory tube *K*, which simplifies the apparatus.

A second method is illustrated in Fig. 192. The receiving tube *T* is here connected in exactly the same manner as in Fig. 190, and hence its operation need not be described again. However, the ends of the inductance *LL'* of the tuned receiving circuit *LL'C* are connected respectively to the filament and grid of a second tube *K*. If the grid of this tube *K* is given some positive potential (for instance by means of a battery, not shown in the figure), then the alternating electromotive force set up by an incoming signal across the inductance *LL'* will produce a current flow in the grid circuit of tube *K*, and the resulting energy expenditure in the internal grid-filament resistance of this tube will correspondingly damp out the received signal current. If, on the other hand, the grid is negative, no current will flow in the grid circuit, and the signals will not be weakened by the presence of the tube K. They will, on the contrary, be strongly amplified through the retoractive action of the tube T. The operation of the circuit of Fig. 192 is thus based on an alternately positive and negative potential being given to the grid of tube K at a frequency of 10,000 to 30,000 cycles, which is done by coupling the grid and plate circuits of the tube K to an oscillatory circuit M tuned to this frequency, the tube then operating as a generator, which of course involves the production of an alternately positive and negative potential on the grid of this tube, as required. The positive resistance of the circuit $LL'C$ then varies correspond-

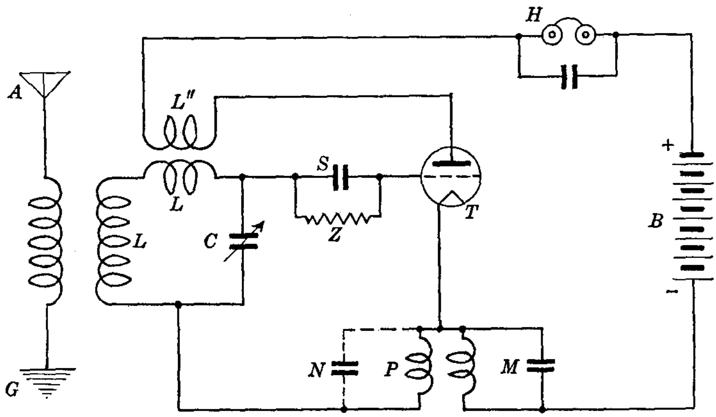


FIG. 193.

ingly, and becomes alternately greater and smaller than the negative resistance developed by the tube T.

It is also possible to use a single tube for the three functions of detector, regenerative amplifier, and resistance-varying oscillator. Thus (Fig. 193) the same receiving circuit is used as shown in Fig. 190, but the plate circuit of the receiving tube T comprises, in addition, an oscillatory circuit M tuned to a frequency of 10,000 to 30,000 cycles, and in which the tube itself sustains continuous oscillations, its grid circuit being, for this purpose, coupled to the circuit M by means of a coil P. The latter is shunted by a high-frequency by-pass condenser N, permitting the radio-frequency currents of the received signals to operate in the grid circuit without being choked out by the impedance of the coil P. The electromotive force generated in circuit

M is thus, as in Fig. 191, impressed upon the plate of the tube, in addition to the battery electromotive force, and the desired resistance variations of the circuit $LL'C$ are obtained.

Referring now to the mode of operation of these superregenerative receiving circuits, consider an incoming signal energizing the antenna and setting up a high-frequency alternating electromotive force in the circuit $LL'C$ tuned to its frequency. This frequency being much higher than that of the locally generated resistance variations, there will be a number of cycles of this signal electromotive force for every cycle of the local resistance varying oscillation.

Now, analyzing this resistance variation,[1] when the receiving circuit resistance is at first large and positive, the incoming signal produces in it a high-frequency alternating current which, as explained at the beginning of this section, rapidly builds up to a small value directly proportional to the signal intensity. As the circuit resistance, under the effect of the locally generated oscillations, gradually decreases and approaches zero, the continued action of the high frequency signal electromotive force tends to increase the amplitude of the oscillation thus built up in the circuit. But this increase is a slow one, due to the continually decreasing damping of the circuit (and the consequent increasing of its time constant). The circuit resistance thus reaches an appreciable negative value, with a current of small, finite amplitude flowing in the circuit, and directly proportional to the signal intensity. This current constitutes the starting impulse for a free oscillation of the circuit, which now rapidly builds up, as explained, at a rate determined by the value of the negative resistance of the circuit. If this resistance were remaining negative for some time, this free oscillation would soon reach the maximum intensity possible with the particular tube used. But the amount of time during which the circuit resistance is negative being limited, the oscillation growth is stopped and the oscillation rapidly dies out, when the circuit resistance again becomes positive. And since during its growth the free oscillation always remains proportional to the amplitude of the starting impulse, as explained before, the final amplitude reached by the oscillation during the negative resistance time interval is also proportional to the signal intensity. The circuit resistance being now again positive, the entire process repeats itself.

[1] See David, Dufour and Mesny, *loc. cit.*

It is thus seen that the current in the receiving circuit $LL'C$ is made up of from 10,000 to 30,000 free high-frequency oscillations per second, the final amplitude of each of which is proportional to that of the signal at the moment of its beginning growth. The rectified or detected plate current is thus constituted by a similar series of 10,000 to 30,000 impulses per second the envelope of which forms a much amplified reproduction of the modulations or envelope of the received signal. The fact that this reproduction is thus cut up into 10,000 to 30,000 impulses per second does not disturb the reproduction of the modulations in the telephone receivers, since this frequency is beyond the limit of audible vibrations.

Of course, the maximum amplitude of each of the 10,000 to 30,000 free oscillations being a function of both the received signal intensity and the rate of increase of the oscillations, this maximum amplitude is seen to depend upon the value of negative resistance of the circuit, hence, on the local operating conditions, such as the intensity of the locally generated oscillations, the tightness of the feed-back coupling between coils L' and L'', heating of the filament, etc.

In particular, in order that the operation of the circuit may be as described above, that is, in order that the final amplitude of each of the 10,000 or 30,000 free oscillations be proportional to the signal intensity at the moment of their starting growth, as required for radio telephony reception, it is necessary that the free oscillations be cut off *during* their growth and that they be not allowed to reach the maximum possible value afforded by the power capacity of the tube. The resistance variation frequency being fixed, this requires the negative resistance of the circuit to be limited to a not too large absolute value, which in turn requires the locally generated oscillations to be of comparatively small amplitude.

On the other hand, if the received signals are strong, or the negative resistance of large absolute value, all of the oscillations may reach the maximum value afforded by the tube power capacity, and the signal modulations will not be reproduced in the plate circuit. Such an adjustment of the circuit-operating conditions is hence unsuitable for radio telephony reception, but it permits a strong reception of weak signals, and is particularly well suited to the reception of radio telegraph signals.

It should be noted that, when no signals are being received, substantially no high-frequency oscillations are generated in the circuit, although this is periodically in a condition which permits such oscillation generation. The reason is that some impulse is required in the circuit for starting the oscillations. In the absence of received signals, such an impulse can be due only to an occasional disturbance (static, stray, or an irregularity of the electron emission from the filament, etc.), and an oscillation thus started cannot last longer than one of the negative resistance periods of the circuit, being damped out in the following positive resistance period. Such a disturbing oscillation is hence of so short a duration that it does not seriously interfere with the reception of the signals.

Finally, it should be stated that the superregenerative method of reception is particularly well suited to short-wave (very high-frequency) reception, it being necessary, for the proper performance, that there be a sufficiently great number of cycles of the received signal current to permit the building up of an oscillation during each period of the circuit-resistance variation. If the signal frequency becomes so small that only a few cycles can take place during one period of negative resistance of the circuit, then the free oscillation cannot build up to a sufficient amplitude, and the circuit operates practically no better than the ordinary regenerative receiver.

THE OSCILLATING THREE-ELECTRODE VACUUM-TUBE RECEIVER

Autodyne Receiver.—As shown in Chap. V, one of the most efficient methods of undamped wave reception is 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 Fig. 109.

It is possible, however, to use a single vacuum tube both for detecting the incoming oscillations and generating the local oscillations. It is necessary, for this purpose, simply to tighten the feed-back coupling of any regenerative detector circuit, that of Fig. 190 for instance, so that the tube will, in addition to detecting the incoming oscillations, generate oscillations of its own.

The antenna circuit being then tuned to the frequency of the signal waves to be received, these waves set up strong oscillations in it, which induce a corresponding alternating electromotive force of their own frequency in the grid circuit 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 $LL'C$ (Fig. 190). This is adjusted to a frequency *slightly different* from that of the antenna circuit. There are thus two alternating electromotive forces of slightly different frequencies impressed upon the grid, which, combining as explained in Chap. V, result in a high-frequency alternating grid electromotive force of periodically variable amplitude, the frequency of this amplitude variation (beat frequency) being equal to the difference between the radio frequency of the received signal and the frequency of the locally generated oscillation.

As in the case of an incoming wave modulated at the transmitting station, the rectifying or detecting action of the tube then produces corresponding beat-frequency pulsations of the plate-current intensity, as explained for the simple three-electrode vacuum-tube detector circuit (grid rectification), resulting in a note being heard in the telephone receivers.

This method of self-heterodyne reception using a single tube for detection and local oscillation generation is also called the *autodyne* method of reception. Its main advantage is its simplicity and ease of handling. It has, however, certain disadvantages, as outlined presently, which make it preferable to resort to different circuit arrangements, to be described later under the section on heterodyne receivers.

A first drawback is that the tube, generating oscillations continuously, excites the antenna circuit, which hence radiates energy, sometimes as far out as several miles. This may be avoided by inserting one or several stages of radio-frequency amplification between the oscillating tube and the antenna, the radio-frequency amplifier acting as a one-way relay or connection, permitting received impulses to be transmitted from the antenna to the detector-tube circuit, but preventing the oscillations generated in this latter circuit from reaching the antenna.

A second point is that the grid oscillatory circuit, in which the local oscillations are generated, must be detuned slightly, in order that the local oscillations may have a frequency different from that of the received signals and produce "beats" with these. The circuit being then not in resonance with the incoming signal frequency, the intensity of the signals is correspondingly reduced.

Finally, a phenomenon known as "automatic synchronization" may take place, which entirely prevents the locally generated oscillations from "beating" against the received oscillations. This property will be studied in detail presently. Circuits and methods whereby it may be avoided in view of accomplishing heterodyne reception will be described thereafter.

Automatic Synchronization.[1]—Consider again the circuit of Fig. 190, with the feed-back coupling between plate and grid circuits so tightened that the tube will be generating continuous oscillations. The frequency F of these oscillations is equal to the natural frequency Ω of the grid oscillatory circuit $LL'C$. Suppose that undamped wave signals of frequency f are to be received. The adjustable tuning condenser C of the grid oscillatory circuit will be varied, varying the natural frequency of the circuit and frequency F of the locally generated oscillations, until the beat frequency F-f becomes small enough to come within the range of audible vibrations, when a high-pitched whistling note will be heard in the telephone receivers of the set. As the circuit, is further adjusted toward resonance, the beat frequency F-f decreases, the pitch of the note becomes lower, and the note disappears entirely when resonance is achieved, the local frequency F being then equal to the frequency f of the received signals. If the tuning condenser is varied further in the same direction, the frequency F differs again more and more from the frequency f, the beat note becomes audible again, increasing in pitch until this becomes so high as to exceed the limit of audibility.

Such, at least, are the conditions in heterodyne receivers of the kind described in a later section of this chapter. In the present case, however, the alternating electromotive force of frequency f due to the incoming signals operates directly in the local oscillator circuit, giving rise to the so-called synchronization phenomena. Thus, instead of decreasing continuously from a very high pitch to zero and increasing again gradually to a very high value when the receiving circuit is being tuned as just described, the pitch of

[1] The property of automatic synchronization, studied here in connection with the three-electrode vacuum-tube oscillator, is quite general, and may be brought forth as well in mechanical oscillators, a tuning fork, organ pipe, pendulum, etc.

the note heard in the telephone receivers starts decreasing as before, and then *suddenly disappears*, well before the resonance point is reached, as if the receiving tube had stopped oscillating. Adjusting the tuning condenser further in the same direction, well past the resonance point, the beat note *appears again suddenly* with the same pitch it had when it disappeared, the pitch then increasing normally as described above, until it exceeds the limit of audibility.

There is thus a frequency range, on either side of the resonance point, wherein *no* beats are produced in the circuit, which can be explained only by the fact that, within this range, the oscillations in the receiving circuit $LL'C$ have a frequency equal to that f of the incoming oscillations, the local oscillation frequency being hence no longer determined by the local tuning adjustments, contrary to the conditions obtaining when no signals are being received (since the frequency of the generated oscillations is then equal to the natural frequency of the oscillator circuit $LL'C$). In other words, the local oscillations are *synchronized* with the incoming oscillations.[1]

Similar conditions obtain when the frequency f of the incoming oscillations is varied at the distant transmitting station while the local receiver tuning adjustments remain unchanged: the phenomena depend essentially on the relative values of the local and signal frequencies and intensities.

The study of the phenomena thus consists in recording the frequency and intensity of the current in the receiving circuit in function of the frequency and intensity of the incoming oscillations. This has been done in the case of the circuit of Fig. 190, and the results[2] are given graphically in the curves of Fig. 194, which will be explained presently.

[1] Synchronization of the three-electrode vacuum-tube receiver was described experimentally and theoretically by H. G. Möller, *Jahr. d. Tel. u. Tel.*, Vol. 17, No. 4, pp. 256–287. April 1921. The theory was also developed from different standpoints, by E. V. Appleton, *Proc. Cambridge Phil. Soc.*, Vol. 21, Part 3, pp. 231–248, 1922; and by B. van der Pol, Phil. Mag., 3, January, 1927, pp. 65–80. Reference should be made to these three papers, for a complete mathematical discussion. Experimental results are given by J. Golz, *Jahrb. d. Tel. u. Tel.*, Vol. 19, pp. 281–289, No. 4, April 1922; and by F. Rossmann and J. Zenneck, *Jahrb. d. Tel. u. Tel.*, Vol. 23, No. 3, pp. 47–52, 1924.

[2] The *current intensity* in the oscillatory circuit $LL'C$ is derived from the readings of a galvanometer connected in series with a rectifying detector and loosely coupled to the circuit $LL'C$. The *frequency* of the currents in the

But consider first the circuit of Fig. 190, and suppose the filament heating battery of the receiving tube to be disconnected, or else the filament current to be reduced by means of a rheostat to such an extent that there will be no appreciable electron emission from the filament.[1] The tube will then be inoperative, and the only current in the oscillatory circuit *LL'C* will be that due to the incoming signals. This current hence has the same frequency as the transmitter current, and if the latter be varied (its amplitude remaining constant), the current in the receiving circuit will be a maximum when the transmitter frequency is equal to the natural frequency of the circuit *LL'C*. Thus, the

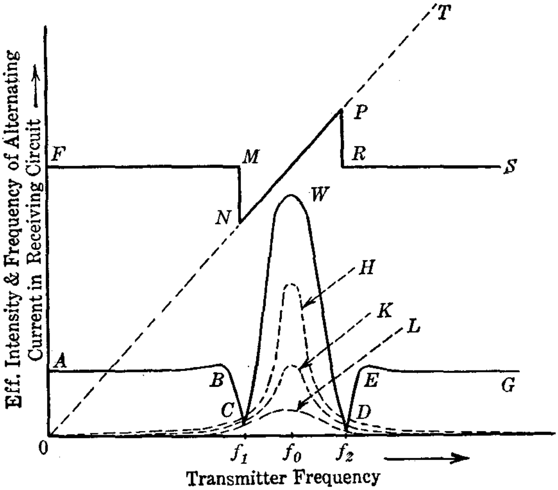


FIG. 194.

well-known resonance curve *L* (Fig. 194) is obtained, giving the current intensity in the circuit *LL'C* in function of the transmitter frequency *f*. At the same time, the frequency of the current in the circuit *LL'C* being always equal to the transmitter frequency, it is represented in function of the latter by the straight line *OT* (Fig. 194).

receiving circuit is measured by means of a heterodyne wavemeter coupled *loosely* to the receiving circuit and adjusted to a frequency differing from both the local and transmitter frequencies. The local and signal currents then each produce a beat note in the heterodyne receiver circuit, and these two notes, which are readily distinguished from each other, are compared with a suitable audio-frequency generator of known frequency.

[1] ROSSMANN, F., and J. ZENNECK, *loc. cit.*

Suppose now the filament current to be turned on, permitting the operation of the tube. If the coupling between the coils L' and L'' of the grid and plate circuits is *negative*, the operation of the tube introduces a negative resistance in the circuit $LL'C$, partially or completely neutralizing the positive (ohmic) resistance of the circuit. Assuming first the negative resistance thus introduced in the circuit to be smaller (in absolute value) than the ohmic resistance, then the tube operates as a regenerative amplifier, and the circuit $LL'C$ behaves like an ordinary oscillatory circuit of very low ohmic resistance, its resonance curve being highly peaked, with a sharp maximum when the transmitter frequency is equal to the natural frequency of the circuit (curve K or H, Fig. 194).

In all the above cases, the current in the circuit $LL'C$ is solely due to the incoming signals, and has a frequency equal to the transmitter frequency, as represented by the straight line OT.

If, on the other hand, the negative resistance introduced by the tube is greater, in absolute value, than the ohmic resistance of the circuit $LL'C$, the tube will generate oscillations of its own in this circuit. And *in the absence of any received signals* the effective intensity and frequency of these locally generated oscillations, represented respectively by the ordinates OA and OF (Fig. 194), are determined by the local operating conditions (battery voltages, tightness of the feed-back coupling, etc.) and by the natural frequency of the circuit $LL'C$.

Now, when signals are being received, the transmitter frequency f being first assumed to be much different from the local frequency F, an alternating current will flow in the circuit $LL'C$ *in addition* to the locally generated current, and this signal current will have a frequency equal to the transmitter frequency f. But, on account of the resonance properties of the circuit $LL'C$ (illustrated by the curve H, for instance), this signal current will have a very small intensity only, its frequency being far removed from the resonance frequency of the circuit; and in combining with the local current it will not materially change the local operating conditions, producing simply slight "beats" with the local current. In other words, the locally generated current retains its original frequency F, substantially equal to the natural frequency of the circuit $LL'C$, while its amplitude, which was constant when no signals were being received, varies slightly above and below this steady value by an amount equal to the amplitude of the received signal current and at a frequency equal to the difference F-f between the local and signal frequencies. This amplitude variation or modulation does not change the average amplitude of the current in the circuit $LL'C$, and increases but very slightly its effective intensity.

Thus, when the transmitter frequency f is gradually increased toward the resonance frequency F of the receiving circuit, or decreased from a large value toward this same frequency F, it sets up in the receiving circuit a current of its own frequency f, as represented by the line OT (Fig. 194), but of small amplitude, while the locally generated oscillations retain their original frequency, as shown by the lines FM and SR, and their original effective intensity, as shown by the curve portions AB and GE. At the same time a pure whistling "beat note" is heard in the telephone receivers, of gradually decreasing pitch as the frequency difference between the local and signal currents decreases.

Actually, when the frequency f of the transmitter is thus varied toward the receiver frequency, the signal current increases gradually in the receiving circuit (see curve H), so that the amplitude of the beats increases as their frequency decreases. These increased amplitude variations of the current in the circuit $LL'C$, while not changing the average value of the latter, increase slightly its effective intensity, which explains the rising tendency of the curve portions AB and GE toward points B and E.

If, now, the transmitter frequency f is adjusted further toward the resonance frequency of the receiver, the intensity of the current in the circuit $LL'C$ suddenly decreases, as shown by the curve portions BC and ED, while at the same time the note heard in the telephone receivers becomes harsh and ragged. And when the transmitter frequency reaches the value f_1 and f_2 corresponding to the minimum points C and D of the curve, the sound in the telephone receivers ceases completely, while the frequency of the local current jumps from its original value to that of the transmitter frequency, as shown respectively by the straight-line portions MN and RP.

Finally, as the transmitter frequency is varied between the values f_1 and f_2, the current intensity in the circuit $LL'C$ rises sharply to a very large maximum W when the transmitter frequency is given a value f_0 equal to the natural frequency (resonance frequency) of the receiving circuit, while the frequency of the locally generated current has within this entire interval f_1f_2 the same value as the transmitter frequency, as shown by the portion *NP*, which coincides with the line *OT*, no "beats" being then produced within this frequency range f_1f_2.

These various conditions may be explained quite simply by referring to Fig. 195, in which the curve *SQPHNT* is the grid-potential, plate-current operating characteristic of the tube. Suppose, first, that no signals are being received. The tube then simply operates as an oscillation generator, and its grid voltage oscillates, as shown by curve 1, with a constant amplitude *A* and at the natural frequency of the circuit *LL'C*. The operating point of the tube then travels back and forth along the characteristic curve between points *Q* and *N*.

Now, as shown in Chap. VIII, in order that stable oscillations may build up in the circuit, it is necessary that the average slope of the curve portion *QN*, which is substantially equal to the slope g of the straight line *QN*, fulfil the requirement

$$g > \left(\frac{CR + \frac{1}{r}}{M}\right)$$

(the letters having the same meanings as in Chap. VIII), as this will make the negative resistance developed by the tube greater (in absolute value) than the ohmic resistance of the oscillatory circuit.

It was shown also that this slope g is greatest when the oscillations are small, and decreases when the oscillation amplitude increases, the oscillations thus building up until their amplitude reduces the slope g of the straight line joining the extreme positions of the operating point of the tube on the characteristic curve to the critical value

$$\frac{\left(CR + \frac{1}{r}\right)}{M}$$

Thus, if *A* (Fig. 195) is the stable oscillation amplitude in the case considered here, the slope g of the straight line *QN* is just equal to the above critical value, and the negative resistance then developed by the tube equals the ohmic resistance of the circuit *LL'C*, the overall resistance of the circuit thus being zero, and the free oscillation maintaining itself in it with this constant amplitude *A*.

If a signal of appreciably different frequency is received, the signal and local currents produce "beats" and the grid voltage will be of the general shape of curve 2 (Fig. 195), which differs from curve 1 in that the amplitude of the alternating grid voltage, instead of being constantly equal to A, varies periodically between values $A + B$ and $A - B$, where B is the amplitude of the received signal electromotive force. The operating point of the tube, instead of oscillating continuously between points Q and N of the curve, will consequently oscillate between points S and T when the "beat current" amplitude is a maximum, between points P and H when it is a minimum, and between intermediate points, such as Q and N for instance, at other times.

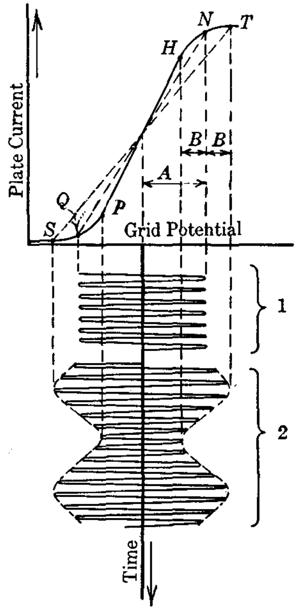


FIG. 195.

Now, although the beat modulation B of the grid voltage extends equally above and below the average value A, the shape of the characteristic curve of the tube is such that the angle between the straight lines QN and PH is smaller than that between the lines QN and ST. It follows that with a modulated grid voltage of the form of curve 2, the *average slope* g' of the operating characteristic curve (taken over a whole "beat" cycle) is *smaller* than the slope g of the line QN corresponding to the operating curve when no signals are being received.

Thus, the fact that an incoming signal is received and produces beats with the locally generated current results in a decrease of the average slope of the operating characteristic curve of the tube. But this, in turn, according to the above remarks, reduces the value of the negative resistance developed by the tube, which thus becomes smaller than the ohmic resistance of the oscillatory circuit $LL'C$. The overall resistance of the circuit thus becomes positive, the free (locally generated) oscillation dies out, and the system behaves like an ordinary tuned receiving circuit, in which, therefore, the current and electromotive force have the frequency of the received signals. Of course, in view of the very tight feed-back coupling, and considering the particular mode of operation of this process, the received signals are very greatly amplified, but no beats are produced, since the local frequency current is no longer generated.

It should be noted also that this synchronization process as just described does *not* take place if the beat-modulation amplitude B is smaller than a certain critical value. Thus, if the beat amplitude B increases gradually from zero upward, the lines PH and ST gradually swing away from the line QN until the line PH coincides with the straight portion of the characteristic curve (in which position it is shown in Fig. 195). From this point on, the line ST alone continues to separate from the line QN, the angle between these two lines then becoming markedly and increasingly larger than that between QN and PH, and permitting synchronization to take place, as described, through a reduction of the *average* slope of the operating curve[1] and a consequent reduction of the negative resistance developed by the tube.

The minimum beat amplitude B required for bringing about this condition is hence dependent on the initial angular position of the line QN (which in turn is determined by the amplitude A of the locally generated oscillations when no signals are being received), being smaller the closer the line QN comes to the straight portion of the characteristic curve, hence the smaller the amplitude A of the locally generated oscillations. Indeed, the mathematical discussion shows that synchronization takes place only if the ratio B/A is greater than a certain critical value.

Now, B is equal to the electromotive force induced in the receiving circuit by the incoming signals. It is therefore directly proportional to the signal intensity, and is greater the smaller the difference between the signal and local frequencies (as shown, by curve H (Fig. 194). If the received signals are so weak that the electromotive force B set up by them in the receiving circuit makes the ratio B/A smaller than the critical value, even when the natural frequency of the circuit is equal to the transmitter (or signal) frequency, then synchronization will not take place. If, on the contrary, the ratio B/A is *greater* than the critical value when the receiving circuit is exactly tuned to the transmitter frequency, then the receiving circuit may be more or less detuned with respect to the signal frequency without

[1] This average slope is the slope of the bisector line of the angle between ST and PH.

destroying the synchronization of the local and signal electromotive forces, until the detuning is made so large that the resulting reduction of B makes the ratio B/A smaller than the critical value, when the local oscillation frequency separates from the signal frequency. A synchronization frequency range is thus defined, the width of which increases and decreases with the received signal intensity.

On the other hand, the value of the ratio B/A may be altered by changing the intensity A of the locally generated oscillations, which may be done by varying the feed-back coupling between coils L' and L'' (see Fig. 149). Thus, for a very weak feed-back coupling, the generated oscillations A are small and the ratio B/A has a correspondingly large value. It follows that even a weak signal B, or a strong signal of greater detuned frequency, will produce synchronization of the receiver. This same condition occurs also when the feed-back coupling is made so tight that the locally generated oscillations nearly stop (coupling coefficient M_2 (Fig. 149)). But if the coupling is given some average value, the generated oscillations A become large, as shown in Fig. 149, and the ratio B/A is correspondingly reduced. Synchronization then takes place only when the incoming signal sets up a strong electromotive force in the receiving circuit, or in other words, when the signal has a frequency very close to the natural frequency of the receiving circuit. The width of the synchronization frequency range is thus a function of the tightness of the feed-back coupling, being large for weak couplings, small for tighter couplings, and larger again for very tight values of the coupling, as shown approximately by the curves of Fig. 196.

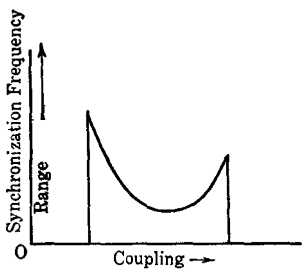


FIG. 196.

There remains to be explained why the current intensity decreases in the receiving circuit when the signal frequency is just outside the synchronization frequency band, as shown by the curve portions BC and DE (Fig. 194). This may be understood as follows, by referring again to Fig. 195. When the signal and local currents combine and produce "beats" in the receiving circuit, the maximum amplitude of the current in this circuit alternately increases and decreases above and below its original value A. Consider first one half-cycle of the "beat" when the amplitude varies from A to $A + B$ and comes back to the value A. The average operating curve of the tube then changes from QN to ST and then returns to QN. During this entire half-beat cycle the average slope of the operating curve is then *smaller* than the slope of the line QN. The negative resistance developed by the tube is hence smaller, in absolute value, than the ohmic resistance of the circuit $LL'C$ (Fig. 190), and the locally generated free oscillation in this circuit is *damped*. If the beat amplitude B is large, such as when the signal frequency approaches the resonance frequency of the circuit, the local oscillation may thus be damped out completely. During the following half-cycle of the "beat," the average operating line swings from QN to PH, and then back to QN. The average slope is thus *greater* than the slope of the line QN, and the negative resistance developed by the tube is consequently greater than the ohmic resistance of the circuit $LL'C$. A free oscillation therefore builds up in the circuit at a rate proportional to its negative resistance. But this negative resistance being, in absolute value, very near zero, the building-up process is slow, and the oscillation thus does not reach a large value when the following half-beat cycle begins, damping out this free oscillation as explained before.

This process of course becomes noticeable only when the signal frequency is sufficiently close to the local frequency to make the beat amplitude B large, for the damping action of the positive resistance developed during one half-beat cycle is then sufficiently greater than the building-up process, due to the negative resistance developed during the following half-beat cycle to offset the latter markedly. This increasing damping action results in the decreased current intensity in the circuit represented by the curve portions BC and DE (Fig. 194). When the action is complete, synchronization takes place, as described before, the local oscillation being completely damped out, and the resonance curve then developing normally to a high peak value under the action of the incoming oscillations.

An application[1] of the synchronization properties of the autodyne receiver consists in tuning the receiving circuit to the frequency of the signals to be received, and making the adjustments in such a manner that synchronization will take place. Incoming signals then do not produce "beats" with the locally genera-

[1] MÖLLER, *loc. cit.*

ated oscillations, but change the amplitude of these oscillations from the value OA (Fig. 194) to the value corresponding to the maximum ordinate W (Fig. 194). As explained before for the simple detector operation of the tube, this increase of the oscillations produces a corresponding decrease of the plate current intensity during the entire signal (dot or dash), capable of operating a direct-current recording device (telegraph relay, etc.). On the other hand, an interfering signal of slightly different frequency will produce "beats" with the locally generated oscillations, provided its frequency lies outside the synchronization frequency range, and hence produces a beat-frequency alternating current in the plate circuit, which either will not operate the direct-current receiving instrument, or else may readily be separated from the direct current representing the desired signal.

Multiple-circuit Regenerative and Autodyne Receiver.[1]— Instead of the simple circuit arrangement of Fig. 190, more complicated circuits are sometimes used for special purposes, two examples of which will be given here.

Thus, the sharp resonance curve which characterizes the circuit of Fig. 190, when operating as a regenerative or as an autodyne receiver, may make this circuit unsuitable for the reception of signals comprising a number of frequency components spread over a more or less wide frequency band. The arrangement of Fig. 197 may then be used, in which the antenna circuit AG is coupled to the grid circuit of the detector tube T, the grid and plate circuits of this tube being, furthermore, coupled, through coils L_g and L_p respectively, to a complex oscillatory circuit system, which is here made up of two coupled circuits K and H, tuned independently to a same frequency F. As known, such a circuit system HK has *two* natural frequencies F' and F'', respectively smaller and larger than the natural frequency F of each circuit considered separately.

In the arrangement of Fig. 197, the circuit H is coupled to the circuit K, and the latter is coupled to both the plate and grid circuits of the tube T. Under these conditions, it is simply necessary that the coupling polarities of the plate and grid circuits to the circuit K be properly chosen, as in the case of a single tuned circuit, irrespective of the coupling polarity of the circuits H and K, which does not play any part in the feed-back process. The

[1] See Hazeltine, L. A., *Proc. Inst. Radio Eng.*, 1918, and also F. Rossmann and J. Zenneck, *loc. cit.*

receiver then behaves exactly like that of Fig. 190 as a regenerative or autodyne receiver, with the only difference that it has *two* resonance frequencies, as just pointed out. As explained in Chap. II, if the coupling between the two circuits H and K is *loose*, these two frequencies F' and F'' come very close to each other and the resonance curve of the system is flat-topped, the system then operating as a frequency band filter.

There is a second manner in which the circuits may be arranged, as shown in Fig. 198, in which the entire circuit chain HK now

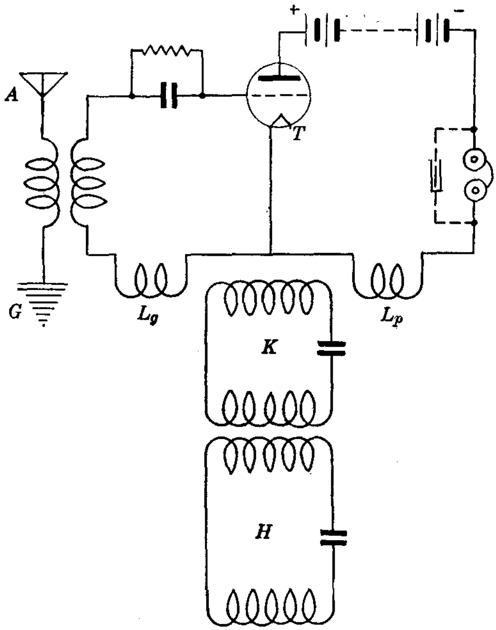


FIG. 197.

forms the feed-back coupling link or channel between the plate and grid circuits of the tube. The properties of this arrangement are distinctly different from those of the former, due to the fact that the oscillation-sustaining impulses are here transferred from the plate to the grid circuit through the medium of the currents in *both* circuits H and K.

Now, as known, the alternating currents in the circuit H and K are in phase with each other at one of the resonance frequencies F' or F'', and in phase opposition (180 deg. out of phase) at the other of these two frequencies. If, therefore, the polarities of the couplings between L_p and H, H and K, and K and L_g are such that the tube develops a *negative* resistance at one of these two frequencies, say F' for instance, it will develop a *positive* resistance at the other frequency F''. It follows that the system will operate as an amplifier at the one frequency F', and as a choking or damping-out device at the other frequency F''. These two frequencies being adjustable by varying the coupling between circuits H and K, it is thus possible to amplify desired signals of frequency F' and reduce interfering signals of frequency F''.

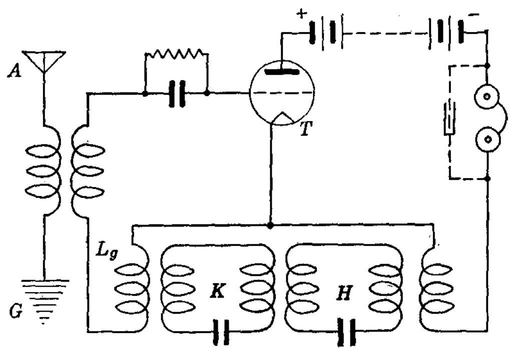


FIG. 198.

VACUUM-TUBE HETERODYNE RECEIVERS

Many heterodyne receiving circuits have been devised with the object of avoiding radiation of the locally generated high frequency oscillations and avoiding also possible interaction between the oscillatory circuit tuned to the frequency of the received signals and the circuit in which the local heterodyning oscillations are generated.[1] Since interaction is due primarily to the fact that the electromotive force which operates in either one of these circuits also operates in the other, or in other words, that the two circuits are to some extent coupled to each other. The methods used therefore either aim at uncoupling the two circuits or else detuning them with respect to each other to such an extent that the electromotive force operating in one of them will have a frequency so different from the natural frequency of the other

[1] Such interaction either may produce automatic synchronization, as explained in the preceding sections, or may lead to the phenomena described in Chap. VIII, p. 202, in which the tuning of one of the circuits affects that of the other.

circuit that it will produce only a negligibly small current or reaction in it.

A first uncoupled circuit arrangement[1] is shown in Fig. 199. The grid circuit of the three-electrode vacuum tube T comprises a first oscillatory circuit LC, coupled or connected to the antenna circuit AG (either directly or through some high-frequency amplifier) and tuned to the frequency of the incoming signals, and a second oscillatory circuit $L'C'$ coupled to a coil L'' of the plate circuit and in which the local oscillations are generated. One side a of the coil L and condenser C being connected to the filament F of the tube, the other side b is connected to the middle point d of the coil L'. One side h of the coil L' and condenser C' is connected to

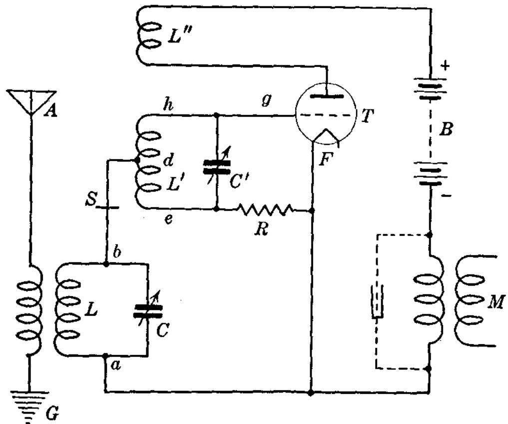


Fig. 199.

the grid g and may hence be considered as connected to the filament of the tube through the internal grid-to-filament resistance of the tube. The other side e of the coil L' and condenser C' is then connected to the filament F through a resistance R which is adjusted to be equal to the average value of this internal grid-filament resistance of the tube.

Under these conditions, when an incoming signal sets up an electromotive force across the coil L and condenser C (hence also across the filament and grid of the tube through the portion dh of the coil L'), the resulting currents in the two circuit branches $bdhgFa$ and $bdeRFa$ will be equal, and produce equal and opposite elec-

[1] A circuit of this general type is used, in particular, in the so-called "tropadyne" superheterodyne receiver.

tromotive forces across the two coil halves *dh* and *de*, so that no electromotive force appears between the ends *e* on *h* of the coil *L'*. The incoming signals thus do not set up any current in the circuit *L'C'*, and hence do not in any way disturb or affect the local oscillation generation. Conversely the local oscillations do not appreciably energize the antenna circuit.

Tracing out the circuit in a similar manner for the locally generated oscillations, these are seen to set up no electromotive force across the terminals *a* and *b* of the circuit *LC*. The resonance adjustments of the two circuits *LC* and *L'C'* are thus entirely independent from each other.

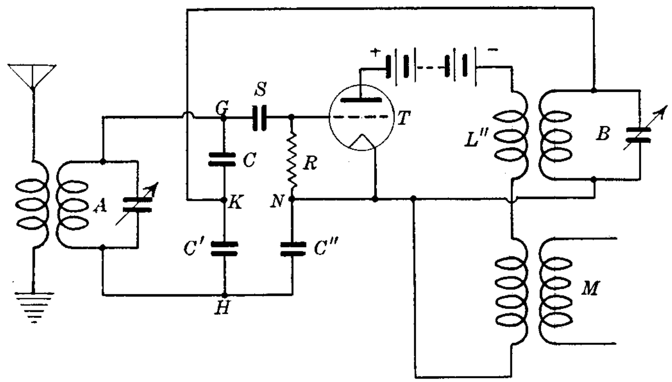


FIG. 200.

The arrangement may be considered as an alternating-current bridge circuit, the four corners of which would be points *e*, *d*, *h*, and *F*. The electromotive force, due to the incoming signals, is applied to this circuit system at points *d* and *F*; that, due to the locally generated oscillations, is applied at points *e* and *h*.

If the tube is to operate also as a detector, a grid condenser is inserted at point *S*, and the resistance *R* serves then also as a grid leak. The beat frequency output electromotive force of the tube is then developed across the secondary terminals of the transformer *M*.

A second bridge receiver circuit[1] is shown in Fig. 200, in which the internal grid-filament capacity of the tube *T*, in series with the condenser *S*, constitutes one arm of a capacity bridge arrangement, the other arms of which are constituted by condensers *C*, *C'* and *C''*. A first oscillatory circuit *A* is coupled to the antenna

[1] RUNGE, W., *Jahrbuch d. Tel. u. Tel.*, Vol. 27, p. 171, No. 6.

circuit and tuned to the frequency of the received signals, and connected to two opposite corners G and H of the system. The other corners K and N are connected to a second circuit B in which the local oscillations are generated, this circuit being, for this purpose, coupled to the coil L'' of the plate circuit of the tube. Both the signal and locally generated electromotive forces are thus impressed upon the grid, but the electromotive force of either circuit A or B does not operate in the other, if the capacities of the bridge arms are suitably adjusted and balanced. The grid is, further, connected to the filament through a high leak resistance R, and the tube, operating as a detector, will furnish a beat

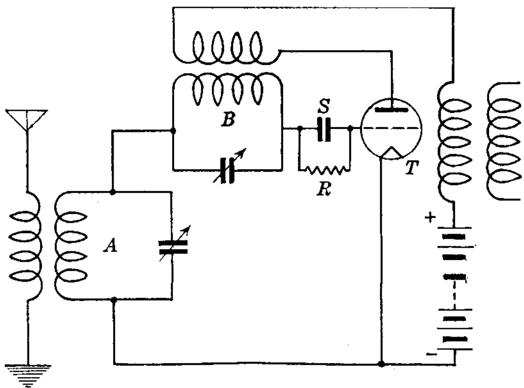


FIG. 201.

frequency electromotive force at the secondary terminals of the output transformer M.

A different principle is used in the circuit of Fig. 201. The grid circuit of the tube T again comprises an oscillatory circuit A coupled to the antenna and tuned to the frequency of the incoming signals, and a second oscillatory circuit B, coupled to the plate circuit, and in which the local oscillations are generated. Only, this circuit is tuned to a frequency equal to nearly one-half or one-third, or some other fraction of the frequency of the incoming signals, and the tube operating conditions are so chosen that, in addition to this low fundamental frequency, harmonics will be generated as well. The second, third, or higher harmonic frequency will then differ from the signal frequency by a certain amount, and the resulting beats produce an electromotive force of their own frequency across the output terminals of the transformer M. Thus, if the incoming signals have a frequency of, say, 1,000,000 cycles and the local frequency generated in circuit B is 500,500 cycles (that is, one-half the signal frequency plus 500), the second harmonic of this local current will have a frequency of 1,001,000 cycles and produce beats of 1,000 cycles frequency with the incoming signals.

Finally, a still different method[1] is shown in Fig. 202. The grid circuit of the receiving tube T comprises simply an oscillatory circuit A tuned to the frequency of the incoming signals and coupled to the antenna circuit. The plate circuit of this tube T comprises the output transformer M, but *is not energized by any battery*. Instead, it comprises an oscillatory circuit B in which

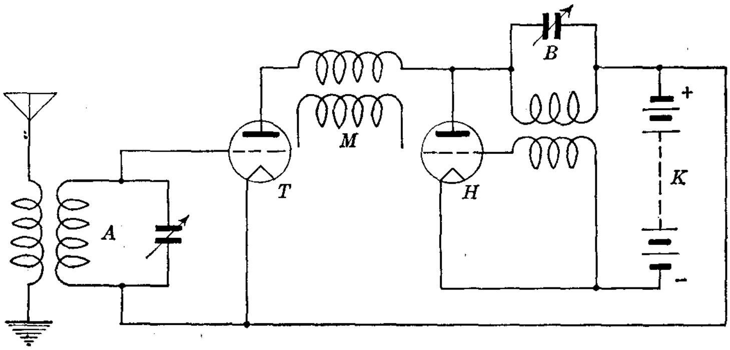


FIG. 202.

oscillations are generated locally by a second tube H, energized by a battery K and having its grid circuit coupled to the plate circuit B, in the usual manner, for oscillation generation.

The alternating electromotive force generated across the coil and condenser of circuit B is thus impressed between the plate and filament of the tube T. The plate thus becomes alternately positive and negative, so that a current will flow in the plate circuit of tube T during the positive cycle halves only of the electromotive force generated in circuit B. If, then, an electromotive force of slightly different frequency from this is developed by incoming signals in the grid circuit A, the plate current of the tube T will have an average value varying at a frequency equal to the frequency difference between the currents in circuits A and B. The local oscillations being rectified in tube T (through its unidirectional conductivity between filament and plate), the incoming

[1] JOUAUST, R., *Onde Electrique*, Paris, Vol. 1, No. 1, pp. 26–33, Jan. 1922.

oscillations need not be "detected" or rectified in any way to produce a beat-frequency electromotive force in the transformer M. No grid leak or condenser is hence shown in the figure. With this method, interaction between the two circuits A and B is, of course, impossible (except through the very small internal capacities of the tube).

Superheterodyne Receivers.—It was shown in a previous section that the rectifying efficiency of a detector increases with the amplitude of the alternating electromotive force to be detected, and that high-frequency cascade amplification of the received signals before their detection thus permitted a considerable increase in the distance range of the receiving station.

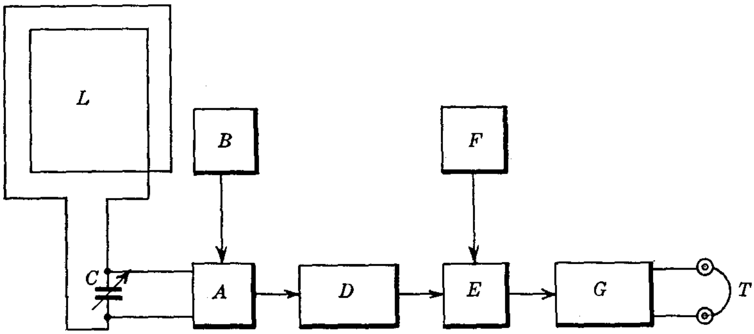


FIG. 203.

Now, a multistage high-frequency cascade amplifier operates most effectively over a certain limited frequency range only, the impedance balance between the tubes and the coupling elements of the various amplifier stages varying with the frequency of the received signals. In addition to this limitation of the frequency range, within which a given amplifier may be used with good efficiency, arises the difficulty of operating amplifiers at the very high frequencies corresponding to short and very short wave signals.

For these reasons, the heterodyne method of reception may be used for transforming the high-frequency alternating current of the received signals into an alternating current of lower radio frequency, through combination of the received-signal current with a locally generated current of suitable frequency, this lower transformed frequency being then better suited to the characteristics of the cascade amplifier used. The method, primarily designed for the reception of continuous-wave telegraph signals, involves the use of several heterodyning and detecting stages, and is therefore called the superheterodyne or multiple-heterodyne method of reception.[1] It may also be adapted for radio telephone reception, as pointed out later, but the last (audio-frequency) heterodyning stage must then be omitted.

The principle of the method is illustrated in Fig. 203. The antenna circuit, here constituted by a loop *L*, is tuned by means of an adjustable condenser *C* and connected to a first detector circuit *A*, which also receives the output electromotive force of a local heterodyne generator *B*. The detector *A* will then yield an alternating output current having a frequency equal to that of the "beats" of the local and signal currents, that is, equal to the difference between the radio frequency of the signals and the frequency of the local generator *B*.

This beat frequency being adjustable at the convenience of the receiving operator through a suitable choice of the frequency of the local generator *B*, is given some radio frequency, generally comprised between 15,000 and 50,000 cycles, for which the amplifier *D* will operate with good efficiency.

After this amplification, the beat or intermediate frequency current is combined with the current produced by a second local generator *F* and sent through a second detector *E*. The operating frequency of this second generator *F* is so chosen that the beat frequency with the output current of the amplifier *D* will be of audible frequency. The audio-frequency output electromotive force of the detector *E* is then impressed upon the input terminals of a low frequency amplifier *G*, which then actuates the telephone receivers or loud speaker *T*.

Of course, if the receiver is intended for radio telephony reception, the second oscillator *F* must be omitted, since the received signals are already modulated at audio frequency at the transmitting station.

The apparatus may be simplified, as shown in Fig. 204, in having a single tube perform the operations of the detector *A* and local generator *B*, and similarly a single tube in place of the independent detector *E* and generator *F*, these tubes being then connected up in one of the manners described in the preceding section.

Also, the intermediate frequency developed at the output terminals of the detector *A* and the low-frequency output of the detec-

[1] E. H. Armstrong, French Patent 501511, Dec. 30, 1918 and *Proc. Inst. Radio Eng.*, 1924.

tor E being markedly different, the amplifier D may be "reflexed" and serve for both intermediate and low-frequency amplification. An additional low-frequency amplifier G may nevertheless be added after this, as shown in the figure.

The superheterodyne receiver, thus permitting the efficient use of cascade amplifiers, constitutes a very powerful receiving apparatus, and has the advantage of being both selective and very easily adjusted and tuned. Since the frequency of the currents to be amplified in D is adjustable at the receiving station, and independent from the frequency of the received signals, the amplifier may be built for a given frequency at which it will operate most effectively, the "intermediate" or first beat frequency being

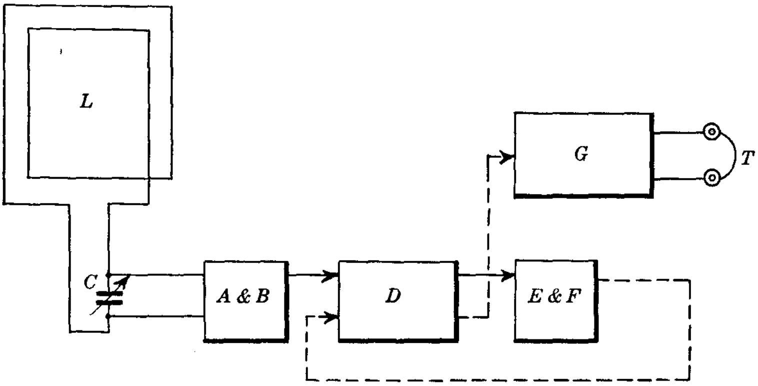


FIG. 204.

then always adjusted to that value by properly choosing the operating frequency of the local generator B.

The operation of adjusting the apparatus for receiving a given station then reduces to the tuning adjustment of the antenna circuit (which is done by means of the condenser C) and the frequency adjustment of the local generator B (which may be done by means of a second variable condenser, the generator B being a three-electrode vacuum tube oscillator and therefore comprising a frequency determining oscillatory circuit). These two adjustments may, in fact, be effected simultaneously by means of a single knob, by connecting the two variable condensers mechanically (through gears or by mounting them on a same shaft) and giving such shapes to their movable plates that the frequency difference between the antenna and local generator circuits remains substantially constant and equal to the operating frequency of the amplifier *D*.

The great selectivity of the superheterodyne receiver may be explained in the following manner. Consider an incoming signal of, say, 1,000,000 cycles frequency and a disturbing signal of 1,005,000 cycles frequency. The *per cent* frequency difference between the two signals is extremely small (0.5 per cent) and the ordinary tuning methods will hardly be capable of separating these properly unless multiple tuned circuits be used (tuned cascade amplifiers, for instance). With the superheterodyne method, however, a locally generated current of, say, 1,040,000 cycles frequency is superimposed upon the received currents, so that the output electromotive forces of the first detector corresponding to the two received signals will have frequencies of 30,000 and 25,000 cycles respectively between which a difference of 20 per cent is thus established. The two signals hence become readily distinguishable, and may be easily separated by means of tuned filter circuits.

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 shown in Chap. 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 at audio frequency, such as those produced by a telephone transmitter actuated by the voice, a sound having the same frequency and characteristics as the transmitter-current variations 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 sound or speech producing these variations at the transmitting station.

A radio telephone transmitting circuit thus consists essentially of a radiating or antenna circuit in which oscillations are set up by some generator of continuous oscillations, and an arrangement whereby the amplitude of these oscillations may be varied or modulated by voice or sound vibrations. The most important methods for generating continuous oscillations have already been studied; namely, the arc, the alternator and the three-electrode vacuum-tube oscillator. The study of radio telephone transmitting circuits will therefore be confined mostly 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 connected or coupled to a rectifying detector and telephone-receiver circuit and such auxiliary tuned and amplifier circuits as may be required. Certain limitations of these circuits will be explained later, but they do not affect the fundamental characteristics just stated.

However, before describing these radio telephone methods and circuits, a brief study will be made of the low-frequency modulated high-frequency alternating current which permits radio telephone communication to be achieved.

The Modulated High-frequency Current.—Consider an alternating current

$$A \cos \Omega t$$

of amplitude A and radio frequency Ω flowing in an antenna circuit. This current is said to be *modulated* or unmodulated according to whether or not its amplitude A is varied.

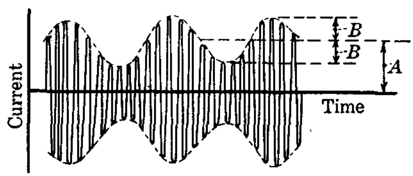


FIG. 205.

Suppose, for simplicity, that this amplitude is modulated sinusoidally by an amount B above and below its value A at a frequency ω, so that the amplitude, instead of remaining constantly equal to A, is equal to

$$A + B \cos \omega t.$$

The modulated high-frequency antenna current, represented in Fig. 205, will then be expressed by the relation

$$(A + B \cos \omega t) \cos \Omega t,$$

which may also be written

$$A \cos \Omega t + B \cos \omega t \cos \Omega t,$$

or, through a simple trigonometric transformation,

$$A \cos \Omega t + \frac{B}{2} \cos (\Omega + \omega)t + \frac{B}{2} \cos (\Omega - \omega)t \qquad (a)$$

Thus, the fact that the high-frequency alternating current $A \cos \Omega t$ is modulated sinusoidally by an amount B and at a frequency ω results in two continuous, undamped alternating currents, of constant amplitude $B/2$ and of respective frequencies $\Omega + \omega$ and $\Omega - \omega$ flowing in the antenna circuit *in addition* to the original unmodulated current $A \cos \Omega t$, generally called the *carrier current.*

Conversely, generating these two constant amplitude, constant frequency alternating currents in the antenna circuit in addition to the original carrier current $A \cos \Omega t$, results in the sinusoidally modulated current of Fig. 205. It is the function of the modulating apparatus to generate these two high-frequency currents when energized or actuated by the low-frequency modulating current $B \cos \omega t$.

A simple sinusoidal modulation of this kind is actually but seldom encountered. It represents a pure musical note, such as the steady whistling sound of a flute. Speech vibrations are much more complex and may be considered as a combination of a number of simultaneous sinusoidal vibrations of different amplitudes and frequencies, the latter being comprised approximately between the useful limits of 200 and 3,000 vibrations per second. Each one of these component vibrations gives rise, in the radio telephone transmitting circuit, to *two* high-frequency alternating currents, as just explained, of frequencies respectively above and below that of the high-frequency carrier current, and differing from this by an amount equal to its own (component) frequency.

The voice-modulated antenna current thus comprises the carrier current frequency F and two frequency *side bands* extending respectively from $F + 200$ to $F + 3{,}000$ cycles, and from $F - 200$ to $F - 3{,}000$ cycles per second.

Since, as will be shown further below, the component currents of at least one entire side band must all be transmitted and received with approximately the same relative strength, it follows that the tuning of the circuits cannot be made as sharp as for radio telegraphy, where only one single frequency is involved.[1]

[1] In radio telegraphy, the high-frequency alternating current is periodically established and interrupted to form the dots and dashes. In other words, its amplitude is thus varied between a constant maximum when the key is closed and zero when it is open. Strictly speaking, this variation may hence be considered as a modulation of the high-frequency current, according to the definition of modulation given above. But the sending or keying speed results in a modulation frequency which is of a small order as compared to the ordinary useful speech frequencies, say, of the order of 100

This requires the tuned circuits to have a certain amount of damping, or better, the simple oscillatory circuits should be replaced by filter circuits having a flat-topped resonance curve.

Referring again to the first case of simple sinusoidal modulation, the curve of Fig. 205 may be taken as representing the current in the receiving antenna as well as in the transmitting antenna circuit. And if a simple rectifying detector receiving circuit be used, as shown in Fig. 87 or 88, only the positive or the negative cycle halves of this current will pass through the telephone receivers, producing a pulsating current varying in the same manner as the current in the transmitting circuit, and producing a sound which follows the original modulations.

Like the curve of Fig. 205, the above expression (*a*) may also be considered as representing the current in both transmitting and receiving circuits. In this expression, only the last two terms are functions of the modulating current, the first term $A \cos \Omega t$ being the carrier current, which is thus seen to remain unchanged, irrespective of whether there is a modulating current operating in the circuit or not. Now suppose that in some manner the carrier current is suppressed and prevented from flowing in the transmitting antenna, and suppose also that a local generator at the receiving station produces a current of carrier frequency Ω and of proper amplitude. Then the currents in the receiving circuits will be the same as if the carrier current had been received from the transmitting station, and reception will occur in a normal manner.[1]

These remarks illustrate the actual function of the carrier-frequency current in the receiving circuits. In exactly the same manner as for heterodyne reception, the carrier-frequency current combines with each component of each side band, producing

---

cycles per second for a transmitting speed of 300 words per minute. The frequency side-band width thus extends over only 100 cycles on either side of the carrier frequency, and permits a reasonably sharp tuning of the circuits.

[1] This *suppressed-carrier current transmission method* has the advantage of reducing the necessary power at the transmitting station in a considerable proportion, the amplitude of the carrier current being *at least* equal to one-half the total modulated current. However, it has the disadvantage of requiring the current of carrier frequency generated locally at the receiving station to be *absolutely* of the same *frequency* and *phase* as the carrier current suppressed at the transmitting station, any departure in frequency or phase resulting in a more or less frequent interruption of the sound in the telephone receivers, rendering actual reception impossible.

"beats" of a frequency equal to the frequency difference between the carrier and the particular component considered. Thus, in the above case of simple sinusoidal modulation, the carrier current of frequency Ω combining with the upper side-band current of frequency $\Omega + \omega$, produces[1] a beat current of frequency ω. And combining with the lower side-band component of frequency $\Omega - \omega$, it produces beats of this same frequency ω. These two beat currents are in phase with each other and add their effects in the telephone receiver, reinforcing each other.

But this also shows that a current of modulation frequency ω may be obtained in the receiver circuit just as well if *one side band* only had been transmitted, the other being suppressed. This reduces the received-signal intensity, but has the double advantage of permitting the suppression of the carrier current without creating the difficulties mentioned in the preceding footnote, and of reducing the overall frequency band width required to that of one single side band, permitting a correspondingly sharper tuning of the circuits.[2]

In the following sections, therefore, a study will be made, successively, of the ordinary radio telephone transmitting circuits (sending out the two side bands and the carrier current), the suppressed carrier and single side-band methods, and the corresponding receiving methods.

Ordinary Modulating Methods.—The modulating devices described in the present section yield a modulated high-frequency current which comprises the carrier frequency current as well as the two side bands.

Ferromagnetic or Detuning Modulation Methods.—When a transmitting antenna is connected to a high-frequency alternating current generator of fixed operating frequency, a high-frequency alternator, for instance, or a three-electrode vacuum-tube oscillator followed by a one-way power amplifier, the current intensity in the antenna depends upon the tuning of the latter. It is a

[1] After detection.

[2] In the case of single-band transmission, the *phase* of the carrier frequency current generated locally at the receiving station is of no practical importance (it changes the relative phases of the various sound components, but this is generally not detected by the ear). The *frequency* of this local current may also be *slightly* different from the actual carrier frequency without unduly altering the pitches of the various sound components. For speech transmission, the permissible difference between local and actual carrier frequency may be as high as about 40 cycles per second.

maximum when the natural frequency of the antenna circuit is equal to the generator frequency, and decreases when the antenna-circuit frequency differs from this value, or in other words, when the antenna circuit is detuned, the current being smaller the greater the amount of detuning.

If the antenna inductance is a coil wound over an iron core (or better, a finely laminated steel core of high electrical resistivity), carrying also a second winding connected to a direct-current generator, the inductance value of the first coil will depend upon the amount of magnetization provided by this second coil, the magnetic permeability of the core (and hence also the coil inductance) decreasing when the direct-current magnetization is increased.

Assuming, then, the antenna circuit to be tuned to the alternator frequency when the direct-current magnetization is zero, the flow of a direct current in the second coil will detune the antenna

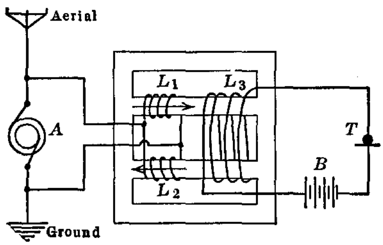


FIG. 206.

circuit, correspondingly reducing the intensity of the high-frequency antenna current. If, then, the direct current in the magnetizing coil is given such a value that the antenna alternating current is decreased to about one-half its maximum value, a variable resistance connected in series with the direct-current generator may be used for varying the direct-current magnetization, and, therefore, also the detuning of the antenna circuit and the intensity of the high-frequency antenna current. This variable resistance may be constituted by a telephone transmitter (microphone), and the current variations will then follow the modulations of the sound vibrations actuating the microphone.

A particular form of this so-called ferromagnetic modulator is the Alexanderson magnetic amplifier, schematically shown in Fig. 206. This shows a high-frequency alternator A connected to the transmitting antenna, in which a continuous high-frequency alternating current is thereby established. The antenn a inductance is 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 high-frequency currents from being induced in the direct-current magnetizing coil L_3.

This third coil L_3 is 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 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 permeability of iron is a function of its magnetization, 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 the device is properly designed, 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, hence the name of "ferromagnetic amplifier" given to the device. This can be readily understood by considering that the reactance of the coils L_1 and L_2 is directly proportional to the permeability of the iron, which in turn is 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 the coils L_1 and L_2. It is thus possible by producing a variation of 0.2 amp. in the current of coil L_3 to vary the antenna power from 5.8 to 42.7 kw.[1]

Vacuum-tube Modulation Methods.—There are many different manners in which the three-electrode vacuum tube may be used for controlling or modulating the amplitude of a high-frequency alternating current. These vacuum-tube modulation methods are generally used in connection with vacuum-tube oscillation generators, but may also be used with other types of high-frequency generators.

A first method is illustrated in Fig. 207. The plate circuit of the three-electrode vacuum tube M, energized by the battery B, is coupled to the antenna circuit AG. The grid circuit of the tube comprises the battery C, and the secondary windings of transformers F and E through which, respectively, the output electromotive forces of the high-frequency alternating current generator H

[1] See GOLDSMITH, A. N., *Radio Telephony*, p. 203.

(which may be a three-electrode vacuum-tube oscillator) and of the microphone circuit D are impressed upon the grid. A condenser K of small capacity is connected in series with the secondary winding of the high-frequency transformer F, in order that this winding shall not form a low impedance short-circuit across the low frequency transformer E.

Under these conditions the high-frequency generator H energizes the antenna circuit AG through the medium of the tube M, which operates as a high-frequency amplifier. But at the same time, the audio-frequency electromotive force due to the operation of the microphone D, varies the average grid potential about which the high-frequency grid-potential variations, due to the generator H, are taking place. The operating point of the tube thus oscillates along the characteristic curve about an average

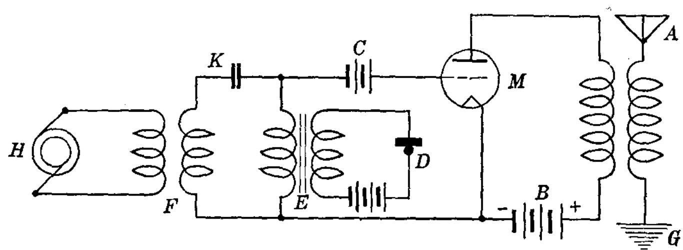


Fig. 207.

position, which changes at audio frequency according to the microphone modulations.

Now, if the high-frequency electromotive force e operating in the grid circuit, due to the generator H, is not too large, the resulting high-frequency alternating plate-current component i is given by the relation

$$i = ge,$$

where g is the slope of the curve at the average operating point[1] (in other words, g is the mutual conductance of the tube at that point).

Assume, then, the grid circuit battery C to be of such polarity and voltage that the average operating point, in the absence of any audio-frequency modulation, is somewhere on the upper or lower bend of the characteristic curve of the tube. The

[1] This relation follows directly from the definition of the mutual conductance of the tube, as given in Chap. VI.

audio-frequency modulations, in alternately shifting the average operating point above and below this position, will then correspondingly increase and decrease the average operating slope g, and consequently also the high-frequency current intensity i as expressed above.

These conditions are illustrated graphically in Fig. 208. Curve A being the lower part of the grid-voltage, plate-current characteristic curve of the tube, let P represent the direct-current operating point in the absence of any modulation, as determined by the plate- and grid-battery voltages. A high-frequency grid electromotive force of constant amplitude, as shown by curve B, will then produce an equally constant amplitude high-frequency alternating current in the plate circuit, as shown by curve N. If, now, a low-frequency modulating electromotive force is also applied to the grid, the constant amplitude high-frequency electromotive force will no longer vary the grid potential above and below an average value equal to the voltage of the battery C, but

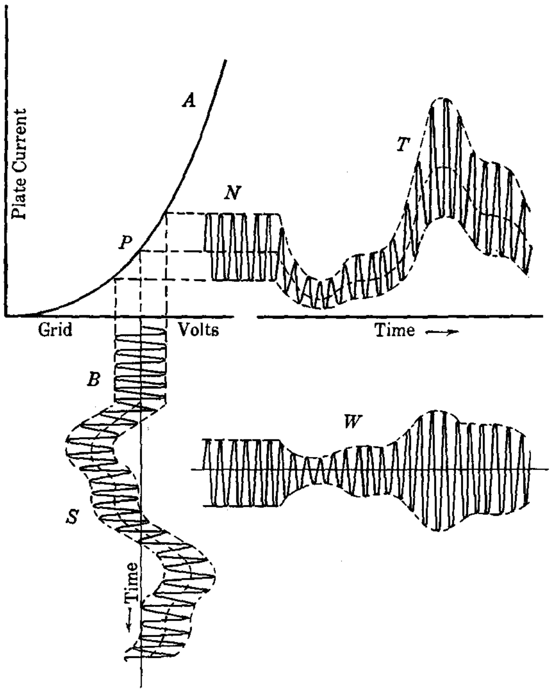


FIG. 208.

about a comparatively slowly variable value equal to the sum of this battery voltage and the low-frequency modulating electromotive force (curve S). The resulting plate current is shown by the curve T. It comprises a low-frequency component, which is an amplification of the modulating microphone current, and a high-frequency component of *variable* amplitude (although the high-frequency grid electromotive force is of *constant* amplitude, as supplied by the generator H), the amplitude being smaller or larger, according to whether the average operating point is shifted toward the more horizontal or the more vertical portions of the characteristic curve by the low-frequency modulating electromotive force.

Now, the low-frequency component of the plate current produces no useful current in the antenna circuit, which has a practically infinite impedance at audio frequencies. Only the high-frequency component will set up a current in the antenna circuit, and this antenna current will be proportional to the high-frequency plate-current component. This high-frequency plate-current component is shown by curve W, which may also be taken as representing the antenna current. It is seen to have an *envelope* which has the same shape as the low-frequency modulating electromotive force. And if a rectifying detector is used at the receiving station, cutting off all the negative (or all the positive) cycle halves of the received current, the rectified current will have an average value which will follow the original audio-frequency modulations.

Actually, the modulation of the high-frequeny current will have exactly the same shape as the modulating grid electromotive force only if the portion of the characteristic curve described by the operating point of the tube is an arc of a parabola (the mutual conductance g being then a linear function of the modulating electromotive force), which is approximately the case over a limited portion only of the operating characteristic curve of the tube. It follows that distortionless operation requires both the high- and low-frequency grid electromotive forces to be of comparatively small amplitudes. The modulated high-frequency output of the tube (curve W) may then be amplified by means of a high-frequency cascade amplifier before being impressed onto the antenna circuit.

Instead of applying both the audio- and radio-frequency electromotive forces to the grid circuit of the modulator tube M, one only of these electromotive forces may be applied to the grid circuit, the other being then impressed onto the plate circuit. Thus, Fig. 209 shows an arrangement belonging to the so-called "absorption modulation" circuits, in which the plate circuit of the tube *M*, energized by the battery *B* and coupled to the antenna circuit *AG*, is also coupled to the high-frequency generator *H* through a transformer *F*, while the grid circuit of the tube is coupled to the microphone circuit *D* through the transformer *E*.

Under these conditions, the generator *H* sets up a high-frequency alternating current in the plate circuit of the tube, which induces a proportional high-frequency electromotive force in the antenna circuit *AG*. The intensity of the high-frequency alter-

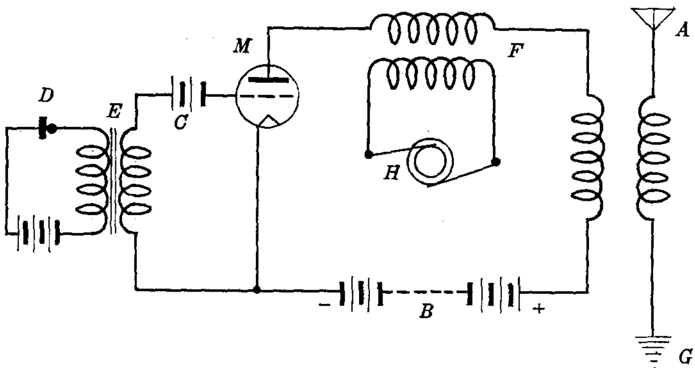


FIG. 209.

nating plate current depends, of course, upon the *resistance* of the plate circuit. This resistance is, for a large part, constituted by the internal plate-filament resistance of the tube *M* and depends, therefore, upon the value of grid potential of the tube (being greater or smaller according to whether this potential is more or less negative). This potential being, in turn, varied at audio frequency by the action of the microphone *D*, the alternating-current intensity in the plate circuit of the tube, and hence in the antenna circuit, is correspondingly varied or modulated.

Of course, the high-frequency electromotive force of the generator *H*, instead of operating in the plate circuit of the tube *M* in series with the battery *B*, may operate in parallel with this, as shown in Fig. 210. The battery *B* is here connected in series with a choke coil *N*, which prevents the flow of a high-frequency alternating current through the battery. The small-capacity condenser K prevents the flow of direct and audio-frequency currents through the generator H. The high-frequency output of the generator H thus divides between the radiation resistance of the antenna circuit AG and the parallel connected internal plate-filament resistance of the tube M. The latter being varied at audio frequency by the variable grid potential due to the operation of the microphone D, the high-frequency current distribution between the antenna circuit and the tube M is correspondingly altered, and the antenna current thereby modulated.

An application of this arrangement where the generator H is constituted by a vacuum-tube oscillator is shown in Fig. 211. The tube H, energized by the battery B, has its grid and plate circuits coupled to each other and to an oscillatory circuit, and also (directly or through a high-frequency amplifier) to the

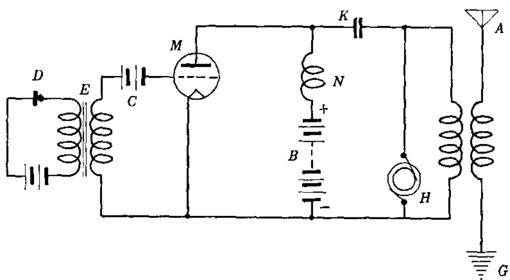


FIG. 210.

antenna circuit, and operates as an oscillation generator. The modulating tube M, with its grid circuit connected to the microphone circuit D as before through the transformer E, is connected in parallel with the oscillator tube H and energized by the same battery B. The variable internal plate-filament resistance of the tube M thus shunts the output terminals of the high-frequency generator H, as in the case of Fig. 210, and operates as explained before.

Of course, the audio-frequency output electromotive force of the microphone circuit D might be impressed onto the grid circuit of the oscillator tube H, and the tube M dispensed with altogether. This method, however, is generally not as good as that illustrated in Fig. 211, for the grid voltage, under the low-frequency variations due to the operation of the microphone, may then become so highly negative or positive that the oscillator tube H suddenly ceases to oscillate. The antenna current then passes suddenly from some large value to zero, and the operation is consequently interrupted. The operation of the set may even be entirely stopped.

Finally, as explained before, the audio-frequency electromotive force may be made to operate in the plate circuit of the tube M,

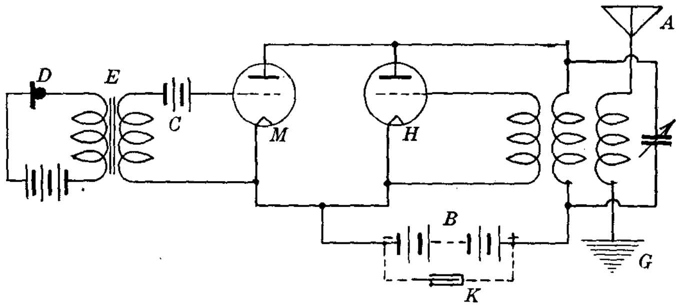


FIG. 211.

as shown in Fig. 212, which represents a so-called "power modulation" arrangement. The output electromotive force of the high-frequency generator H is impressed onto the grid circuit of the tube M, and the resulting alternating plate-current component excites the antenna circuit AG. The amount of high-frequency power developed in the plate circuit of the tube and trans-

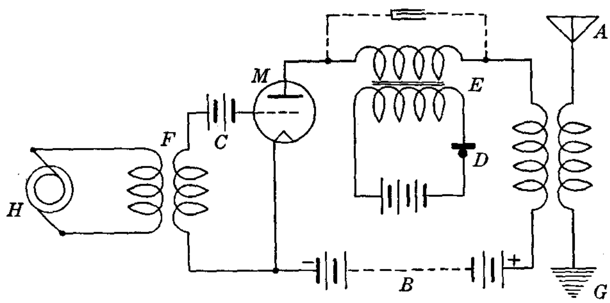


FIG. 212.

ferred to the antenna circuit depending upon the plate voltage of the tube, this is varied or modulated by connecting, in series with the battery B, the secondary winding of the transformer E through which the audio-frequency output electromotive force of the microphone circuit D is introduced into the plate circuit, adding to or subtracting from the battery voltage.

Figure 213 shows a similar "power modulation" circuit arrangement, in which the tube *H* operates as an oscillation generator. Its plate circuit is energized by the battery *B* connected in series with the choke coil *N*. The plate and grid circuits are coupled electrostatically to each other through the condensers *P* and *T*, and continuous oscillations are thus generated in the circuit *PTL*, coupled to the antenna circuit *AG*. (The grid is connected to the circuit through a condenser *S* which prevents the high potential of the battery *B* from being impressed onto it through the coil *L*. It must then be connected conductively to the filament through a high impedance *R*).

In parallel with the energizing battery *B* and its choke coil *N* is connected the secondary winding of the transformer *E*, through which the audio-frequency electromotive force produced by the

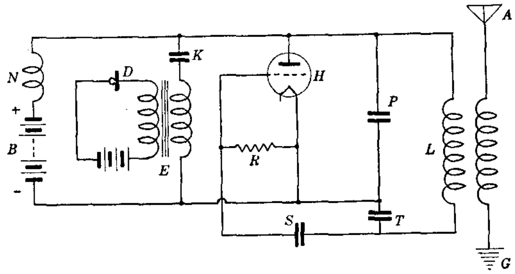


Fig. 213.

operation of the microphone *D* is impressed between the plate and filament of the tube *H*. A condenser *K* is connected in series with this secondary winding to prevent the direct current from the battery *B* flowing through it. The audio-frequency electromotive force thus varies the operating plate voltage of the tube, and correspondingly changes the amplitude of the high-frequency oscillations generated by it in the circuit *PTL* and in the antenna circuit.

More effective operation is obtained by first amplifying the audio-frequency electromotive force produced by the microphone, before impressing it onto the oscillator tube *H*. This is shown in Fig. 214, which differs from Fig. 213 in that the audio-frequency output electromotive force of the microphone circuit *D* is impressed onto the grid circuit of a tube *M*. The plate circuit of this tube being energized by the battery *B* in series with the audio-frequency choke coil *N*, the audio-frequency grid potential variations cannot produce corresponding plate-current variations, in view of the large impedance of the coil N. But they develop a much amplified audio-frequency electromotive force between the plate and filament of the tube M (as explained in Chap. VII, in the section on Voltage Amplification). This is then impressed between the plate and filament of the oscillator tube H, varying its operating plate voltage and hence the amplitude of its generated oscillations.

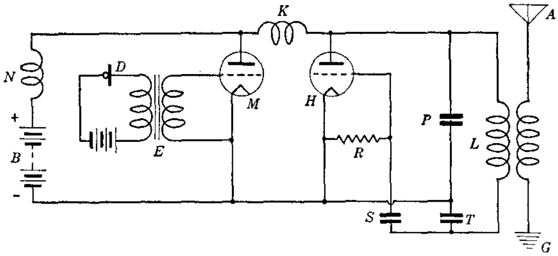


FIG. 214.

The radio-frequency choke coil K connected between the two tubes H and M prevents the oscillations generated by the former from being shunted off by the plate-filament resistance of the latter.

Ordinary Radio Telephone Reception Methods.—The current in a radio transmitting antenna, as obtained by the methods described above, is a high-frequency alternating current of varying maximum amplitude, the variations following the current variations of the 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 thus be had from curve W (Fig. 208).

If this current is rectified by a suitable detector, such as a crystal or vacuum tube, only the negative or positive half-cycles are permitted to flow in the receiving circuit. The current thus rectified then has an envelope and an average value which vary in proportion to the microphone current at the transmitting station. The rectified current received will therefore reproduce the speech when made to flow through telephone receivers.

It follows 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 receiving circuits making use of such detectors can therefore be used substantially 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, pointed out at the beginning of this chapter, that the transmitting antenna does not radiate a wave of one single frequency, but one comprising a large number of component frequencies which, through their combinations, constitute the modulated high-frequency wave, and which, for a speech-modulated wave, form a frequency band of about 6,000 cycles width (3,000 for each frequency side band).

It follows that the circuits, instead of being sharply tuned, must have a flat-topped resonance curve, as explained before. Failure to receive all of the transmitted frequency components equally well, or approximately so, results in a distortion of the low-frequency detected current, which is no longer an exact reproduction of the original speech modulation.

Another cause of distortion, which applies to all frequency components alike, lies in the fact that the operating characteristic curve of the detector differs from a straight line. That is, the current through the detector is not directly proportional to the voltage applied to the detector, but is rather proportional, say, to the square of the voltage. This introduces double-frequency harmonics in the low-frequency detected current, which do not exist in the original modulating vibrations nor in the envelope of the received modulated high-frequency current, and hence unduly alter the shape of the telephone receiver current.

This is illustrated in Fig. 215, where *S* represents the voltage-current characteristic curve[1] of the detector. Let, then, the solid line curve *W* represent the high-frequency alternating voltage set up across the detector by an incoming wave. The amplitude of this alternating electromotive force varies in accordance with the

[1] In case of a three-electrode vacuum-tube detector utilizing the grid-detection principle and employing a shunted grid condenser, curve *S* should be taken as representing the grid-voltage, grid-current curve of the tube. In case the plate-detection principle is used (asymmetrical amplification), curve *S* is the grid-voltage, plate-current curve of the tube.

speech modulations, and it is the function of the detector to furnish a low-frequency current having as closely as possible the same shape as the "envelope" (dotted line $MNPQ$) of this received high-frequency electromotive force.

Suppose, as an illustration, that the modulations are such that the amplitude of the received signals varies periodically by equal amounts above and below the unmodulated amplitude A, as shown in the figure. A distortionless rectification should then yield a low-frequency detected current, as shown by the dotted

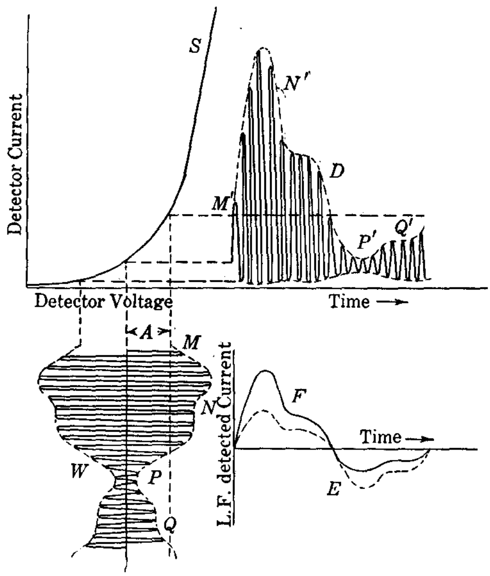


FIG. 215.

curve E, which is an exact reproduction of the envelope curve $MNPQ$, with its positive and negative halves of identical shape.

Actually, however, the rectifying action of the detector is based on the fact that the operating characteristic S of the detector is *curved,* in other words that the detector current (represented by the solid line curve D) is a *distorted* reproduction of the received current (or electromotive force) shown by curve W, as explained before. It follows that the envelope $M'N'P'Q'$ of the detected current, corresponding to the envelope $MNPQ$ of the received input detector electromotive force is also distorted, so that finally the low-frequency output current of the detector (curve F)

is a correspondingly distorted reproduction of the original modulations.

Now, the amount of distortion depends essentially on the so-called modulation depth of the received signals. In the case of a sinusoidally modulated signal (Fig. 205), this modulation depth is defined as the ratio B/A of the modulation amplitude B to the unmodulated amplitude A of the high-frequency carrier current. The greater this ratio, that is, the deeper the modulation, the greater is the amount of distortion resulting from the process of detection as just described.

This may readily be shown mathematically as follows. Let

$$i = i_0 + ae + be^2$$

represent the detector voltage-current characteristic curve S (Fig. 215), i and e being, respectively, the detector current and voltage, and i_0, a and b being three constants. If then

$$e = A \cos \Omega t + \frac{B}{2} \cos (\Omega + \omega)t + \frac{B}{2} \cos (\Omega - \omega)t$$

represents a sinusoidally modulated electromotive force, as explained before, A and Ω being the carrier amplitude and frequency, B and ω the modulation amplitude and frequency, the term ae in the above expression becomes a high-frequency term, which is of no interest here, while the term be^2 comprises, in addition to high-frequency terms, two low-frequency terms:

$$bAB \cos \omega t \text{ and } b\frac{B^2}{4} \cos 2\omega t,$$

the first of which is of the original modulation frequency ω and the second, the distorting, undesired double-frequency term.

Now, the relative amplitude of this double-frequency term, as compared to that of the fundamental frequency term, is equal to the ratio

$$\frac{bB^2/4}{bAB} \text{ or } \frac{B}{4A}.$$

Hence, the greater the value of the ratio B/A, called, as stated before, the modulation ratio or depth, the greater also the amplitude of the second harmonic as compared to that of the fundamental component of the detected current, and consequently the greater the resulting speech distortion in the receiving apparatus.

In order to minimize such distortion, it is thus necessary to make the ratio B/A small. This may be done at the transmitting station by keeping the modulation amplitude B small as compared to the carrier amplitude A, that is, to modulate the carrier current but weakly at the transmitting station. This, of course, has the disadvantage of reducing the transmission efficiency, since radio telephone transmission is precisely based on the modulation of the carrier current; in other words, it requires the major part of the energy to be transmitted unmodulated in the form of a large carrier current, while the modulated part of the current, which is ultimately the useful current for telephone communication, must be kept small.

A second and better method therefore consists in using a rather large modulation ratio B/A and increasing the carrier current amplitude A at the receiving station by generating at the receiving station a local high-frequency alternating current of same phase and frequency as the received carrier current, and impressing this local current onto the detector in addition to the received signal current. The ratio B/A may thus be decreased in the receiving circuit as much as required while at the same time the modulation ratio B/A may be made substantially as large as desired at the transmitting station, thereby giving a good transmission efficiency.[1]

It may at first seem difficult to thus provide at the receiving station a locally generated current of *identically* the same phase and frequency as the received carrier current. Actually, this is a simple operation, it being necessary merely to use an oscillating receiving circuit so adjusted that it will become synchronized by the incoming signals. As explained in the preceding chapter, a circuit like that of Fig. 190 might be used, or any other type of regenerative amplifier circuit, with the feed-back coupling so tightened that the tube will generate continuous oscillations of its own in the absence of any received signals. The frequency of these locally generated oscillations is simply made *approximately* equal to the carrier-current frequency of the signals to be received. When these are being received, automatic synchronization takes place, as described in the preceding chapter, the locally generated current falling into step with the received carrier cur-

[1] With the modulating arrangements described in the preceding Section, the ratio B/A is never greater than unity (which corresponds to 100 per cent modulation), at least as long as undistorted modulation is effected. Values of B/A greater than unity may be obtained by special methods described later.

rent, and adding to the latter, thereby reducing the ratio B/A, as required.

However, this receiving arrangement does not only greatly amplify the carrier current, but also those frequency components of the complex received current which are in the neighborhood of the carrier current frequency (and which, in each of the two side bands, correspond to the low-frequency components of the modulating sound vibrations), while it does not in the same proportion amplify the received-current components the frequencies of which are further away from the carrier frequency, and which correspond to the higher frequency components of the modulating sound vibrations. This results in a distortion of the detected currents, producing a sound in the telephone receivers which is more guttural than the original. This may readily be corrected in a number of different manners, by inserting suitable filter circuits in the radio-frequency amplifying circuits (this method is particularly effective when applied to the intermediate-frequency amplifier circuits of a superheterodyne receiver), or by using an audio-frequency amplifier after the detector, having such operating characteristics as to amplify the high-pitched notes or sounds to a greater extent than the lower-pitched ones. The distortion is thus compensated by a distortion in the opposite direction, provided as just described.

Single Frequency-band Radio Telephony.—As explained before, radio telephone communication may be effected by transmitting one only of the two side bands of the modulated high-frequency current, and combining this single side-band current in the receiving circuit with a locally generated alternating current of carrier frequency. The latter then produces beats with each one of the component currents of the received side-band current, which after rectification or detection give rise to low-frequency alternating currents reproducing the various components of the original modulating vibrations.[1]

The reception method thus constitutes a special case of the heterodyne reception method, and use may be made of the heterodyne receiving circuits described in the preceding chapter.

[1] In thus combining with the locally generated current of carrier frequency, each component of the received side-band current also gives rise to harmonics, the envelope of the "beats" being not sinusoidal. However, these harmonics are sufficiently small not to produce any objectionable distortion of the speech.

In fact, the main problem in single side-band receiving apparatus is that of keeping the frequency of the local current as constant and as nearly equal as possible to the frequency of the carrier current suppressed at the transmitting station.

On the other hand, special circuit arrangements are required at the transmitting station for suppressing the carrier current and one of the two side bands, permitting one only of the two side-band currents to energize the antenna circuit.

It will be recalled that F being the carrier-current frequency, and the useful speech frequencies being comprised between, say, 200 and 3,000 cycles per second, the speech-modulated high-frequency current in an ordinary transmitting circuit comprises, in addition to the carrier current (of frequency F) the two side-band currents the frequencies of which are respectively comprised between $F + 200$ and $F + 3{,}000$ cycles (upper side band) and between $F - 200$ and $F - 3{,}000$ cycles (lower side band). The *relative* importance of the frequency differences between the various components of such a modulated high-frequency current, or in other words the *per cent* differences between the various frequency components, is hence greater the lower the carrier frequency F. The separation of the one side-band frequencies is therefore most easily achieved when as small a carrier frequency F as possible is used, say of the order of 30,000 cycles per second. Of course, in most cases an appreciably greater carrier frequency is required or desirable for establishing radio telephone communication. A second modulation operation is then required at the transmitting station at this higher carrier frequency, as described further below. The process of single side-band transmission thus involves two distinct operations, which will be studied presently.

Consider an ordinary radio telephone transmitting circuit as previously described, comprising a radio-frequency generator and a modulating arrangement. As pointed out in a preceding section, the modulating action of the microphone results in two side-band currents arising in the output circuit *in addition* to the carrier current, the latter flowing in the output circuit of the set irrespective of whether the microphone is actuated by the sound vibrations or not. If then the radio-frequency generator be also used for energizing an auxiliary circuit, the constant carrier-frequency alternating electromotive force set up in the latter may be opposed to the equally constant carrier-frequency electromotive force operating in the actual output circuit of the transmitting set.

This carrier-frequency electromotive force may thus be neutralized, and *no* current of carrier frequency will then be set up in the actual output circuit. The output circuit of the radio set will therefore carry the two side-band currents only, and only when the microphone is actuated by the modulating sound vibrations. If then the carrier frequency used is comparatively small, say 30,000 cycles, it will be an easy matter as explained above, to tune out one of the two side-band currents by means of a simple filter circuit, leaving only one side-band current operating in the output circuit, the per cent frequency difference between the two side bands being then very large, and a frequency gap of 400 cycles existing between the two bands, owing to the elimination of the carrier-frequency current.

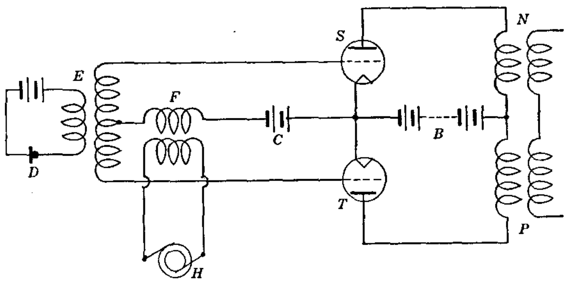


Fig. 216.

The neutralization or elimination of the carrier current is generally done by means of a so-called *balanced modulator*,[1] represented schematically in Fig. 216, which utilizes two similar three-electrode vacuum tubes connected substantially in the same manner as for push-pull amplification (see Fig. 144). Only, the output electromotive force of the high-frequency carrier-current generator H is here impressed onto the grids of both tubes S and T through a transformer F connected in the *common* circuit branch connecting the grids to the filaments of the tubes, while the low-frequency modulating electromotive force developed by the action of the microphone D is impressed differentially upon the grids through a transformer E having the middle point of its secondary winding connected to the filaments of the tubes, and the ends of this same winding connected to the grids.

[1] See "Carrier Current Telephony and Telegraphy," by E. H. Colpitts and O. B. Blackwell, *Trans. Am. Inst. Elec. Eng.*, Vol. XL, pp. 205–300, 1921; and also "Production of Single Side Band for Transatlantic Telephony," by R. A. Heising, *Proc. Inst. Radio Eng.*, June 1925.

Suppose, first, that no speech is impressed onto the microphone D. Only the high-frequency alternating electromotive force of the carrier-current generator H operates then in the grid circuits of the tubes, varying the grid potentials of both tubes simultaneously, and hence producing synchronous, in-phase current variations in the plate circuits of the two tubes. By suitably connecting the secondary windings of the transformers N and P in series with each other, the high-frequency electromotive forces induced in them by these plate-current variations of the tubes S and T, may then be made to oppose each other, and if the tubes, transformers, and other corresponding circuit elements are respectively similar to each other, these electromotive forces will be equal and balance each other out exactly. Thus, *in the absence of any modulating electromotive force, the output electromotive force of the device, as developed across the series-connected secondaries of transformers N and P, is equal to zero, providing the circuits of the two tubes are properly balanced.*

If now the microphone D is actuated by a sound vibration, an alternating low-frequency modulating electromotive force is developed across the secondaries of the transformer E and impressed onto the grids of the tubes, varying the grid potentials of the two tubes in *opposite* directions. And since the average operating grid potential of the tubes, as determined by the voltage of the battery C, is so chosen as to bring the operating point of the tubes on a curved portion of their characteristic curve, these oppositely directed grid-potential variations unbalance the system in periodically opposite directions. The high-frequency alternating plate currents of the two tubes produced by the action of the generator H will then no longer be equal, and the electromotive forces developed in transformers N and P no longer balance each other exactly, so that a high-frequency electromotive force appears at the output terminals of the device.

Thus, *in the presence of a modulating electromotive force in the transformer E, a high-frequency output electromotive force is developed across the series-connected secondaries of transformers N, P.*

Analyzing the action further, the low-frequency alternating electromotive force developed by the operation of the microphone D across the secondary of transformer E first increases the grid potential of one of the tubes while it decreases that of the other, then decreases the grid potential of the first tube and increases that of the second. The high-frequency alternating current developed by the action of the generator H in the plate circuit of tube S will hence be periodically greater and smaller than that developed in the plate circuit of tube T. In other words, the unbalance between the secondary high-frequency electromotive forces of transformers N, P is periodically reversed, and the secondary high-frequency electromotive force of transformer N is greater than that of transformer P during one-half cycle of the low frequency modulating electromotive force due to the microphone D, while that of transformer P is greater than that of transformer N during the following half-cycle of the modulating electromotive force.

It follows that the high-frequency output electromotive force of the device, developed across the ends of the secondary circuit branch NP varies periodically in amplitude between zero and a maximum proportional to the modulation amplitude, at a frequency equal to twice the modulation frequency (since the periodicity of the unbalance reversals is equal to that of the successive half-cycles of the modulating electromotive force, as just stated), and also that the polarity of this output electromotive force reverses at every half-cycle of the modulating electromotive force. These characteristics denote that *the output electromotive force of the device does not comprise the carrier frequency of the generator H, but comprises the two side-band frequencies.*

Having thus isolated these two frequency bands (which in the present example extend from 30,200 to 33,000 cycles and from 29,800 to 27,000 cycles, respectively), a filter circuit is connected in series with the output terminals of the balanced modulator, stopping one of the two frequency bands, the lower one, for instance, and passing only the electromotive forces and currents belonging to the other frequency band.

The single frequency band electromotive force thus obtained (extending from 30,200 to 33,000 cycles) is now used as a *modulating* electromotive force in a *second* modulator, actuated by a second high-frequency alternating-current generator, the frequency of which is the actual transmission-carrier frequency, say, for instance, 100,000 cycles. As in the case of audio-frequency modulation just described, the resulting modulated high-frequency current then comprises two frequency side bands the limits of which are respectively equal to the carrier frequency (100,000 cycles) plus and minus the modulating frequencies (30,200 to 33,000 cycles), hence extending, in the present instance, from 130,200 to 133,000 cycles (upper side band) and from 67,000 to 69,800 cycles (lower side band). These two side bands are thus separated by a frequency gap of 60,400 cycles, and may very easily be separated by means of tuned filter circuits, so that only one of these two side-band electromotive forces, the upper one for instance, will energize the transmitting antenna circuit. The modulated waves thus occupy a frequency band of only 2,800 cycles width (equal to the frequency-band width of the modulating sound vibrations).

At the receiving station, the received current will, correspondingly, be a complex current the component frequencies of which extend from 130,200 to 133,000 cycles, and it is necessary merely to combine this received current with a locally generated alternating current having a frequency of 130,000 cycles (equal to the sum of the frequencies of the suppressed intermediate and high-frequency carrier currents, *viz.*, 30,000 and 100,000 cycles) to produce beats which, after rectification or detection, result in a current reproducing the original audio-frequency modulations. Thus, although the modulation at the transmitting stations is effected in several steps, demodulation at the receiving station may be achieved in one single operation.

Mathematical Theory of the Balanced Modulator.—It may be of interest to outline the operation of the balanced modulator in mathematical terms, in a manner showing the relation of this device to the push-pull amplifier, and also leading to the theory of the modulator-type heterodyne receiver, as pointed out further below.

Following the same method as on page 173, let i_1 and i_2 represent the alternating plate currents set up in the plate circuits of the tubes S and T respectively. Since the output windings of transformers N, P are so connected that the electromotive forces induced in them oppose each other, as explained in the preceding section, the electromotive force developed at the terminals of the series-connected secondaries of these transformers is proportional to the difference $i_1 - i_2$ of these alternating plate currents. This was found, on page 174, to be equal to

$$i_1 - i_2 = A\left[2g - 6c^2m - \frac{3A^2}{2}\right]\sin \omega t + \frac{mA^3}{2}\sin 3\omega t$$

where $A \sin \omega t$ represents in the present instance the electromotive force of amplitude A and audio frequency ω developed by the operation of the microphone D (Fig. 216), pure sinusoidal modulation being here assumed for simplicity. Now, the factor c in the above expression, representing the electromotive force operating in the common branch of the grid circuits of the two tubes, is not constant as in the case of the push-pull amplifier, but is equal to the voltage c of the battery C (Fig. 216) *plus* the alternating voltage

$$h = H \sin \Omega t$$

produced by the generator H, of carrier frequency Ω.

The above expression thus becomes

$$i_1 - i_2 = A\left[2g - 6m\left(c + h\right)^2 - \frac{3A^2}{2}\right] \sin \omega t + \frac{mA^3}{2} \sin 3\omega t,$$

which, after some simple handling, becomes

$$\begin{aligned} i_1 - i_2 = A\left[2g - 6m\left(c^2 + \frac{H^2}{2}\right) - \frac{3A^2}{2}\right] &\sin \omega t \\ &+\frac{mA^3}{2} \sin 3\omega t \\ &+\frac{3mAH^2}{2} \sin (2\Omega + \omega)t \\ &-\frac{3mAH^2}{2} \sin (2\Omega - \omega)t \\ &+6mAcH \cos (\Omega + \omega)t \\ &-6mAcH \cos (\Omega - \omega)t. \end{aligned}$$

It appears at once that all of these six terms being multiplied by A, which is the amplitude of the audio-frequency modulating electromotive force, the device will not develop any output electromotive force when the modulation is absent, that is, when the microphone is not operating. This result was already pointed out in the preceding section.

When A is different from zero, that is when the microphone is actuated by sound vibrations of frequency ω, the six terms of the above expression appear in the output electromotive force of the modulator.

The first and second terms represent audio-frequency currents of frequencies respectively equal to ω and 3ω. They do not play any useful part in the radio-frequency output circuits of the transmitting set, which have a practically infinite audio-frequency impedance.

The third and fourth terms represent two side-band currents of frequencies respectively equal to $(2\Omega + \omega)$ and $(2\Omega - \omega)$, differing from *twice* the carrier-current frequency Ω by an amount ω equal to the modulating audio frequency. They are of comparatively small amplitude and may be suppressed completely by means of filter circuits, or simply choked out by the high impedance of the output circuit, which is designed for an output frequency of the order of the carrier frequency Ω.

Finally, the last two terms represent the two side-band currents of frequencies respectively equal to $(\Omega + \omega)$ and $(\Omega - \omega)$, that is, equal to the carrier frequently Ω plus or minus the modulating frequency ω.

It will be noted that there is *no* output current of carrier frequency Ω, which brings out the essential property of the balanced

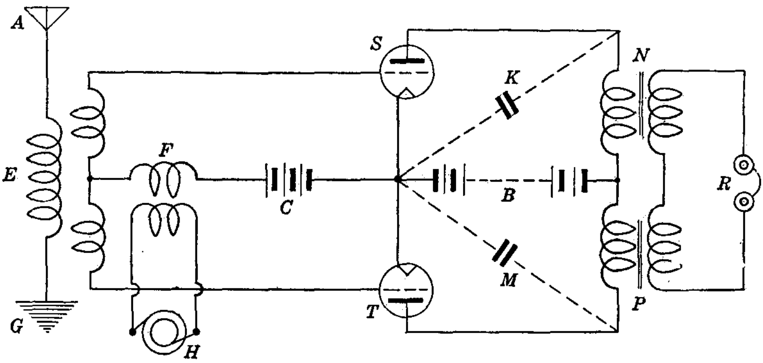


FIG. 217.

modulator, which is to eliminate the carrier-frequency current from the output circuit.

Note on the Modulator-type Heterodyne Receiver.—The balanced modulator circuit just described may be adapted for use as a heterodyne receiving circuit, serving to combine a locally generated current of frequency Ω with the radio-frequency current due to incoming signals. A receiving circuit of this kind is shown in Fig. 217, which is essentially the same as that of Fig. 216, except that the microphone circuit D is here replaced by the receiving antenna circuit AG, and the output transformers N, P are connected to telephone receivers R (or to an intermediate frequency amplifier if the beats are of superaudible frequency as in the superheterodyne receiver). Using the above formulas, the frequency ω then repersents the radio frequency of the received signals while Ω is the frequency of the local generator H. As shown by the last expression, the output current of this circuit system comprises a term of frequency $\Omega - \omega$. If then the local generator frequency Ω is so adjusted as to make this frequency difference $\Omega - \omega$ small enough to come within the range of audible vibrations, a sound will be heard in the telephone receivers R without any further rectification or detection being required.

The transformers N and P being audio-frequency transformers, the radio-frequency terms of the output current (represented by

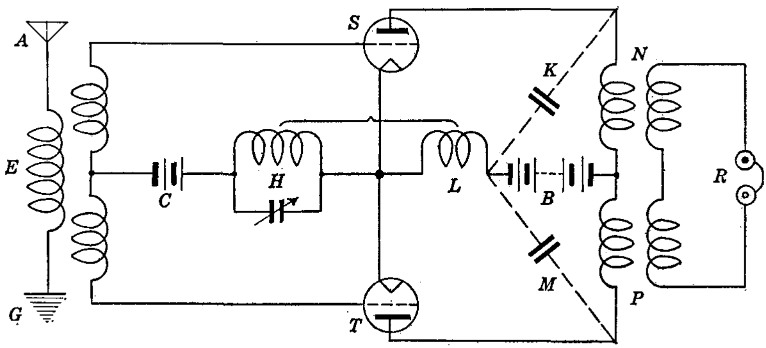


FIG. 218.

the first five terms of the above expression) are by-passed by small capacity condensers K, M (Fig. 217).

The generation of the local heterodyning oscillations may be effected by the very same circuit system, as shown in Fig. 218. It is simply necessary to insert an oscillatory circuit H in the common branch of the grid circuits of the two tubes and couple it inductively to a coil L connected in the common branch of the two plate circuits. It should be noted that in this arrangement the received signals do *not* operate in the circuit H, and that this circuit is not coupled to the antenna circuit, which is therefore not energized by the locally generated oscillations.

CHAPTER XI

MISCELLANEOUS APPLICATIONS

Airplane Radio Apparatus.—Airplane radio apparatus does not differ fundamentally from ordinary radio apparatus. It must be specially 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 designer. The design and construction of all parts must also have great ruggedness in order that the apparatus may operate properly and without changing its adjustment 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 generally 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 line 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 governments engaged in the World war. In all the later airplane sets, except of course the damped-wave transmitting sets, vacuum tubes were used both for transmitting and receiving. The circuits are essentially the same as for ordinary sets designed for ground use. An interesting question, however, was as to the best method of supplying power for the vacuum-tube filament and plate circuits. Storage batteries have been used to quite some extent in cases where high voltages were not required; but the limited time during which such batteries can be depended upon to supply power at 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 on 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 air fan rotating at 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 air fans, which drive the generator at approximately constant speed despite variations in the airplane or wind velocity; special regulating vacuum tubes, iron wire resistances, etc. These various devices, as well as many of the special types of generators used, have been described in the technical press of different countries, and are not treated here further.

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. With regard to the loop type, reference should be made to a section of Chap. III in which the loop radio receiving circuits used on airplanes were fully discussed. These are used mostly for direction finding and navigation purposes, but may, of course, serve for receiving messages in the usual way. Other types of loop antennae permit both transmitting and receiving from an airplane, and have been used, to a limited extent, on airplanes flying in squadron formation.

The type of open antenna which has found the widest application is illustrated in Fig. 219, and constitutes a so-called "trailing antenna." It consists of a 300-ft. length of bare wire which 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 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 radio system. The counterpoise is made up of the metallic parts of the airplane—the engine, stay wires, etc., which are all carefully bonded

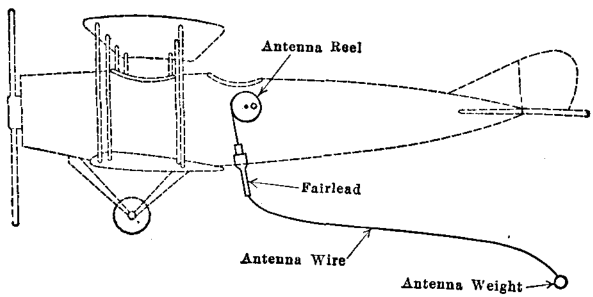


FIG. 219.

together. The radiation characteristic of a trailing antenna is shown in the diagram of Fig. 220, 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. It may be noted from Fig. 220 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 side of the fuselage, and not in 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 electrical 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. 221, 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.

The length of antenna wire used depends upon the wave length of the signals. As a rough rule, the natural wave length in meters is approximately equal to the length in feet of the trailing

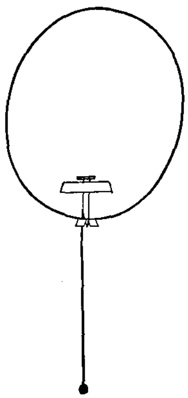


FIG. 220.

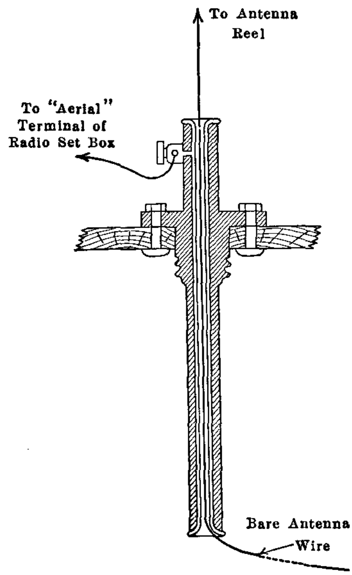


FIG. 221.

antenna, as used on a Curtiss JN-4 airplane. It should be noted that the natural wave length of such a trailing antenna depends not only upon the length of the trailing wire, but also upon 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 a 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. 222, 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. 223, but somewhat 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. 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.

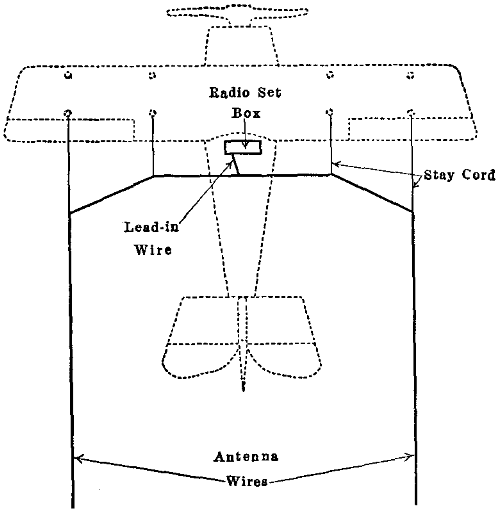


FIG. 222.

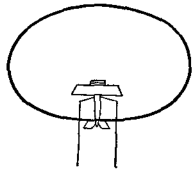


FIG. 223.

Submarine Radio Apparatus.—Loop antennae have been successfully used for submarine radio work, and some of the tests made will be briefly described here.[1]

Referring to Fig. 224, 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 submarine. *EF* is another wire similar to *AB*. Both these wires are of copper, covered with 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

[1] The data on the submarine loop antenna was kindly furnished by the U. S. Bureau of Standards, Washington, D. C.

circuit connected at D. The loop circuit is thus electrically the same as an 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 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

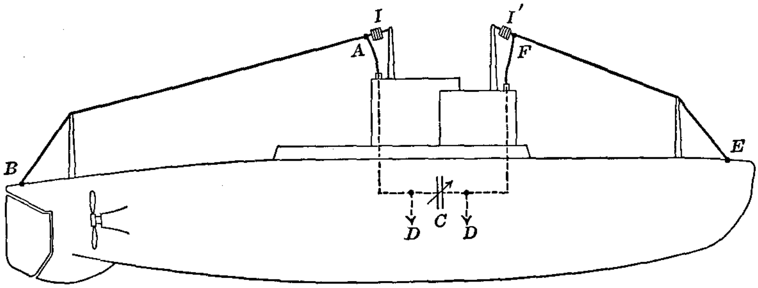


FIG. 224.

the surface. The maximum depth at which the signals are readable is dependent upon the wave length of the transmitting station. This depth increases with the wave length, and is about 21 ft. (for the top of the loop) when receiving signals of 10,000-m. wave length. The receiving apparatus used in these tests comprised a vacuum tube detector and three-stage resistance coupled vacuum tube cascade amplifier.

Transmission of the 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-m. 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 of the water, 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 ordinary antennae 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 the ship is submerged or on the surface, in the manner explained for ordinary loops in a preceding chapter.

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

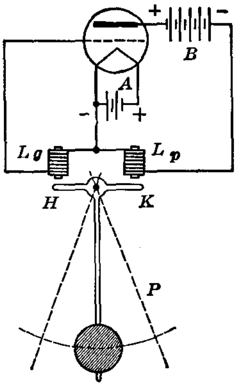


FIG. 225.

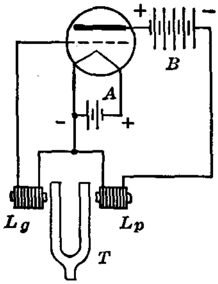


FIG. 226.

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 are of suitable magnitude and phase to sustain the mechanical oscillations.

As an illustration, consider the system of Fig. 225, which shows how a three-electrode vaccum tube may be used to maintain continuous oscillations of a pendulum P. Two coils L_g and L_p are inserted respectively in the grid and plate circuits of the tube and placed in front of a small 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 thus compensates the friction losses which, in the absence of the vacuum-tube device, would damp out the oscillations of the pendulum and bring it to rest.

Another example is given in Fig. 226, where continuous vibrations of a tuning fork T are sustained 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 possessing a natural period of vibration or oscillation of its own. This is identical with the case of electrical oscillations, where the tube circuits are coupled inductively to a common oscillatory circuit having a natural period of electrical oscillation.

Now, in both of the above cases the operation is based on the fact that the motion or vibration of the mechanical oscillator produces *magnetic* field variations in a coil connected in the grid circuit of the tube. Similar effects may be obtained if the motion of the mechanically vibrating body produces *electric* field variations, a condenser being then used for linking the mechanical vibrator to the vacuum-tube circuits. This is most strikingly illustrated in the so-called piezo-electric oscillators which will be described presently.

Piezo-electric Resonators and Oscillators.—These devices make use of the property of certain crystalline dielectric materials, such as quartz, Rochelle salt, and tourmalin, for instance, to become electrically polarized when submitted to a mechanical stress; and, conversely, to become mechanically strained when subjected to an electrical stress, that is, when placed in an electrostatic field.[1]

Thus, consider a condenser made up of two metal plates separated by a thin piece of quartz. If the plates are pressed

[1] These properties obtain only as far as the mechanical or electrical stress applied to the crystal is suitably directed with respect to the axes of the crystal. This makes it necessary in practice to grind opposite faces of the crystal piece used parallel to each other, and in the desired direction.

together, squeezing the quartz slab between them, they will become electrically charged; in other words, a potential difference will appear between them. Conversely, if the condenser is connected to a battery establishing a potential difference between the metal plates, the quartz plate will become slightly flattened, or else its thickness will increase slightly, depending upon the polarity of the metal plates.

This permits electrical vibrations to be converted into mechanical vibrations, or, conversely, mechanical into electrical vibrations. Thus, if an alternating potential difference is established between the metal plates of the condenser, the quartz plate between them will correspondingly and alternately contract and expand at the same rate and frequency, and this mechanical vibration of the quartz plate will be particularly strong if the frequency of the alternating potential difference applied to the

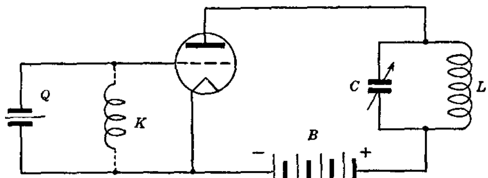


FIG. 227.

condenser armatures is equal to the natural frequency of mechanical vibration of the quartz plate (which may be as high as several hundred thousand cycles per second when the piece of quartz is sufficiently small).

Conversely, if the piece of quartz is struck mechanically and set into free vibration, an alternating potential difference appears at the condenser terminals, having a frequency equal to that of the free mechanical vibrations of the quartz crystal.

Now, if, as in Fig. 227, the condenser with the quartz dielectric Q is connected between the grid and filament of a three-electrode tube,[1] this alternating electromotive force will produce corresponding current variations in the plate circuit of the tube. And by inserting in this plate circuit an oscillatory circuit LC tuned to the frequency of the mechanical vibrations of the quartz piece, a high alternating potential will build up by resonance

[1] In order that the grid shall not find itself insulated from the filament by the condenser Q, it is then necessary to provide a conductive grid-to-filament connection through a choke coil K of large inductance.

across this circuit, and hence also between the plate and filament of the tube.

As a result of the internal grid-to-plate capacity of the tube, a proportional alternating electromotive force will be induced in the grid circuit and applied across the condenser *Q*, adding itself to the electromotive force originally developed across this condenser by the starting vibration of the quartz crystal. And since, as stated before, an alternating electromotive force applied to the condenser *Q* will produce alternate contraction and expansion of the crystal, it may be understood that this arrangement will serve to sustain continuously the oscillations or vibrations of the quartz crystal.

This oscillator does not differ in principle from that of Fig. 163, the only change being that the grid-circuit oscillating element is here a quartz crystal instead of an electrical oscillatory circuit.

The mechanical oscillators described here (pendulum, tuning fork, and piezo-électric oscillator) find their most useful applications as standard frequency oscillation generators, and in all circuit arrangements where a great constancy of the generated frequency is required.

INDEX