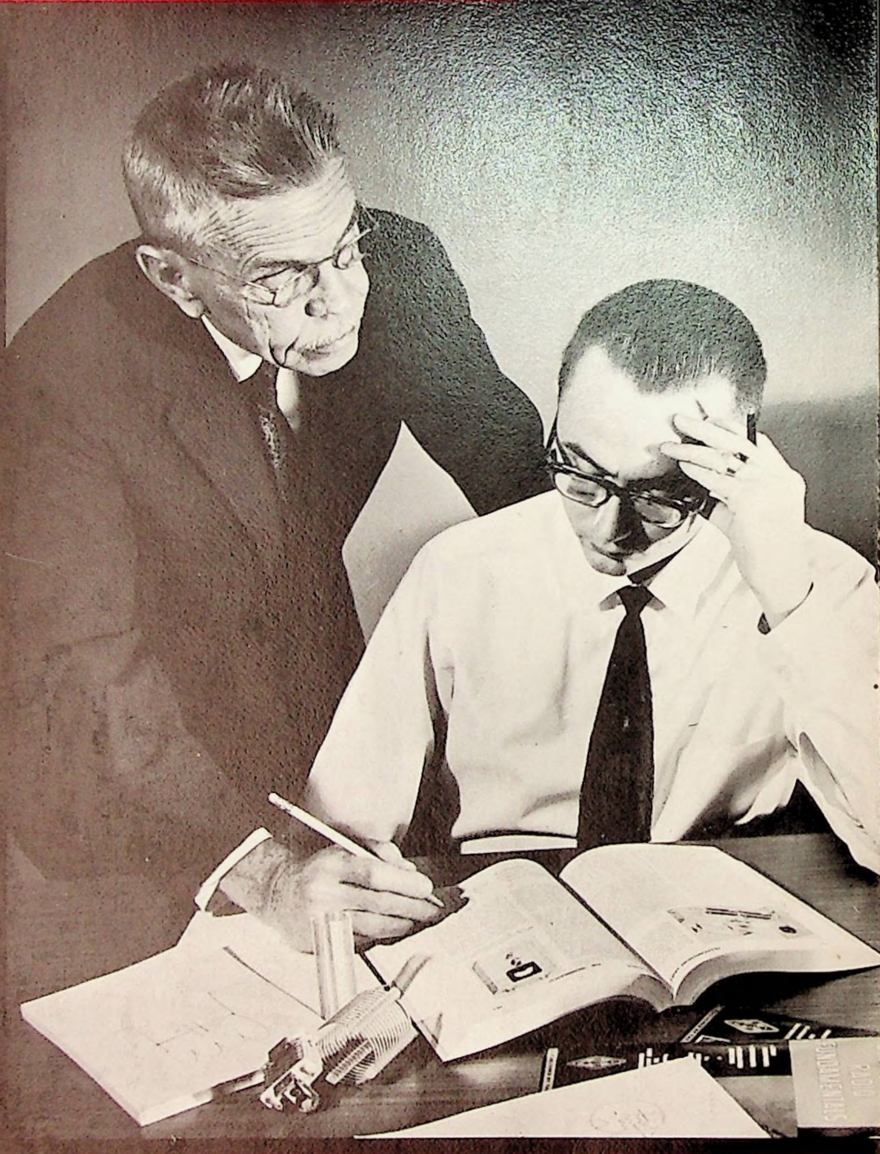
A COURSE IN RADIO FUNDAMENTALS

$4.00

COMPLETELY NEW!

A BASIC TEXT FOR INDIVIDUALIZED STUDY

By GEORGE GRAMMER



PUBLISHED BY THE AMERICAN RADIO RELAY LEAGUE

A COURSE IN RADIO FUNDAMENTALS

By

GEORGE GRAMMER

Technical Director, ARRL
and Technical Editor, QST
(Retired)

Published by

THE AMERICAN RADIO RELAY LEAGUE, INC.

Newington, Connecticut, U. S. A.

CONTENTS

FOREWORD

The original ***A Course in Radio Fundamentals*** was a study-and-experiment program using ***The Radio Amateur's Handbook*** as a text. Written during the early part of World War II, it first ran serially in ***QST*** and later was published in pamphlet form as the demand for it continued. With minor changes it has since gone through many reprintings.

Yearly revisions of the Handbook have made it increasingly impracticable to base the ***Course*** on that volume. This edition, therefore, is complete in itself. It has been written expressly for the purpose of presenting the electrical and electronic principles that are basic to understanding radio circuit operation. The detailed application of those principles is left to other publications which, by their periodical nature, can expose the latest circuit and component developments for the reader's consideration.

The text is at an intermediate level. The treatment is quantitative where simple mathematics, commonly taught in elementary high-school courses, will suffice. However, the novice probably will benefit if, as a preliminary, he first takes the even simpler path offered by ***How to Become a Radio Amateur*** and ***Understanding Amateur Radio***, then later follows with this ***Course***.

It is hoped that the book also will serve as a convenient one-cover technical reference for those who, although familiar with the subject, occasionally need to consult scattered books and publications for refresher information.

JOHN HUNTOON, W1RW
General Manager, ARRL.

Newington, Conn.

Chapter 1

Charge; electrical forces; induction; electric field; conductors and insulators; potential and emf; capacitance; dielectrics; distribution of charge on conductors; shielding.

The Electric Field

THE STUDY of electricity begins with the idea of electric charge – what it is, what it does, how it can be produced, the laws that govern its behavior. Electric charge originates in the atom, the very tiny particle – actually, the atom is an assembly of even smaller particles – from which all substances are built up. The atom's principal sections are the nucleus (the central core) and one or more electrons which move around the nucleus in complicated ways.

The electron is the smallest particle of electricity that can exist. All electrons are alike, and electrical effects external to the atom depend on whether or not an electron can be set free to move about independently.

CHARGE: In a neutral atom – that is, one showing no external electrical effects – the total quantity of electricity or charge represented by its electrons is balanced by an equal quantity associated with the atomic nucleus – balanced because there are two types of electricity, the kind represented by the electrons being termed negative and that associated with the nucleus being called positive. This difference is called polarity.

In many materials, particularly metals, it is possible to dislodge one electron from an atom with relative ease. When an atom loses an electron its electrical balance is upset. The electron removed is a free particle of negative electricity, and since the atom has lost that amount of negative charge it is left with an equal excess of positive charge. The free electron is called a negative ion and the atom lacking an electron is called a positive ion. When some billions of similar ions are collected together the quantity of charge becomes large enough to cause observable effects.

ELECTRICAL FORCES: Electric charges exert force on one another. If two charges have the same polarity they will resist being brought close together; that is, they will repel each other. If the charges have opposite polarity, one positive and one negative, they attract each other.

The attraction and repulsion between charges can easily be demonstrated by a simple experiment. Suspend a small piece of metal foil from a dry thread, then rub one end of a plastic comb on wool cloth and bring it close to the foil, as in Fig. 1-1(A). Rubbing the comb on wool detaches electrons from the wool fibers and deposits them on the surface of the comb; thus there is a negative charge on the rubbed end of the comb. Then if the charged end of the comb is brought near the foil, free electrons in the foil will be repelled by the negative charge on the comb and will congregate as far from it as they can – on the edge farthest from the comb. This leaves the near edge of the foil with a deficiency of electrons – i.e., a positive charge – and it is therefore attracted to the negative charge on the comb. Because the positively charged edge is nearer the comb than the negatively charged edge, the *net* force acting on the foil is attraction, and the foil swings toward the comb.

Both attraction and repulsion are thus demonstrated, the latter by the foil's turning its negatively-charged edge away from the comb, and the former by the movement of the foil toward the comb.

INDUCTION: The opposite charges on the foil are said to be induced on it by the charge on the comb. If the comb is removed from the vicinity of the foil without allowing the two to come into contact, the free electrons on the foil immediately distribute themselves over it and external electrical effects disappear.

However, if the comb does make contact with the foil some of the excess electrons on the comb move onto the foil, giving it a negative charge, (Fig. 1-1(B), and the foil is immediately repelled from the comb (C). This charge remains on the foil, as can be demonstrated by moving the comb away and then bringing it back to the vicinity of the foil. The force is always repulsion. (The effect will eventually disappear because electrons leak off the foil to ground, and thus over a period of time the foil loses its charge.)

The foil can be given a charge of the *opposite* kind by bringing the comb and foil close together, as shown in Fig. 1-2(A), but not allowing them to touch. Then if the foil is touched momentarily by

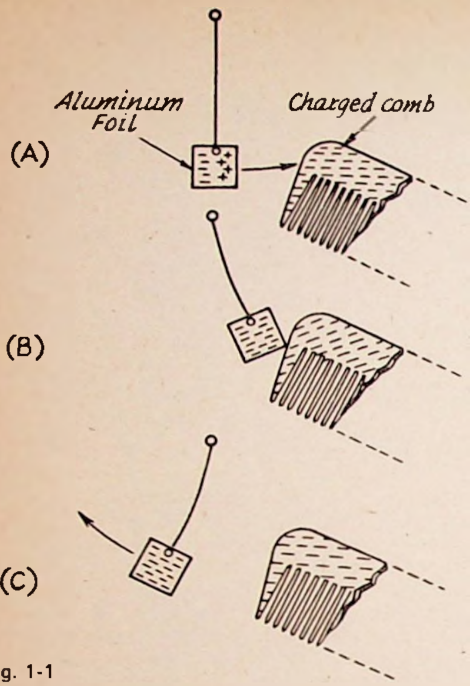


Fig. 1-1

a finger (B) the force of repulsion will cause the induced negative charge on the foil to run off to the body (and eventually to ground), leaving the foil with a positive charge. The foil is again "permanently" charged (at least until electrons leak back onto the foil through the thread) and will be attracted whenever the negatively charged comb is brought near but not allowed to make contact (C).

ELECTRIC FIELD: The action demonstrated by the comb-foil experiment can be explained by assuming that the forces are exerted along lines of force between two electric charges, as shown in Fig. 1-3(A). If a body of negligible dimensions and having a weak positive charge is placed at the spot indicated by the small circle, it will move along the associated line of force (shown dashed) to travel to the negative charge. A similar charge placed at the point shown by the solid dot would move along the line indicated.

Thus there is a field of force, the electric field, the direction in which the force is acting at each point in the field being determined by the direction of the individual line passing through that point. The field intensity, which depends on the quantity of electricity (the unit of which is called the coulomb) in the charges responsible for the field and on the distance between them, is represented by the *number* of lines of force per unit area in a surface perpendicular to the lines. (This is simply a convention, however; the actual field is continuous at every point, and should not be thought to be like a bundle of discrete wires laid side-by-side.) The field is electrostatic if it is not moving or changing in intensity. The number of lines per unit area is called the flux density of the field.

Lines of force acting in similar directions repel each other, just as like charges repel each other. This accounts for the spreading out of the field. The field intensity decreases in the outer parts of the field because lines of force there represent a greater distance between the charges.

As suggested by Fig. 1-3(A), every line of force leaving a charged body has to terminate on an opposite charge. If there is ostensibly only one such charged body, any large object in which electrons are free to move can supply the required charge, by induction. The earth itself is such an object, and serves to terminate the lines from the isolated charge shown in Fig. 1-3(B).

CONDUCTORS and INSULATORS: If electrons in a substance are free to move about, the substance is a conductor. When there is an electric field acting on them the electrons will move in the conductor according to the rules given above.

Metals are "good" conductors, in that they have many free electrons moving about at random in them at any given time. (However, each one has become detached from one of the atoms in the metal, so there are just as many positive ions as negative ions; hence there is no net charge.) Under the influence of an electric field the free electrons move in the direction determined by the lines of

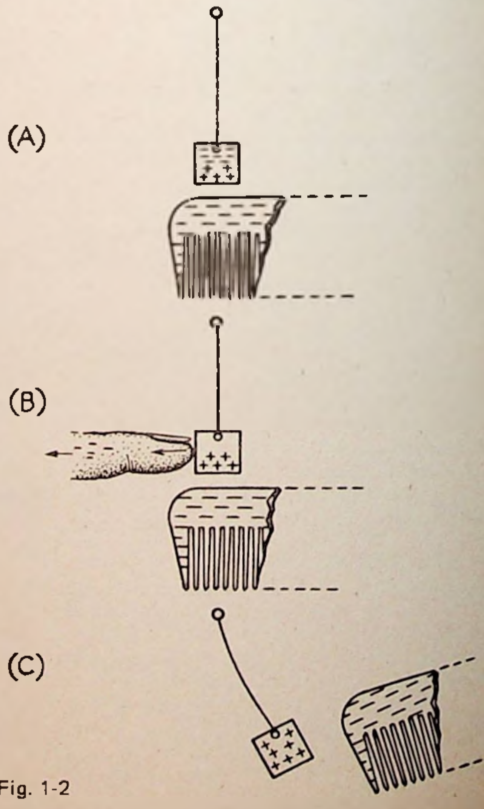


Fig. 1-2

force; this concerted movement constitutes an electric current.

If the electrons in a substance are firmly bound to their atoms and are not free to move, the substance is an insulator or dielectric. When an insulator is in an electric field there is a dielectric stress in it tending to pull electrons away from their atoms, but since the electrons are not freed there is no electric current.

The comb in Figs. 1-1 and 1-2 is an insulator, which is why it could hold a charge at one end; the electrons accumulated there could not move.

POTENTIAL and emf: The potential of a charge is defined as the work that would be done in bringing a "unit" charge of the same polarity from an infinite distance to the point where the charge whose potential is being measured is located. Potential is thus a measure of the electrical energy in the charge. The practical unit of electric potential is the volt.

The force acting between two charges is inversely proportional to the square of the distance between them, and so is relatively small if the distance is appreciable.

In electric circuits, the difference of potential between two points in the circuit is of primary interest, because this difference is the net electromotive force (emf) effective in moving electrons between the two points. Thus if one point in a circuit has a potential of 100 volts and another point a potential of 15 volts, the difference or emf is 85 volts between the two points. If one point is positive and the other negative with respect to some reference point in the circuit, the *algebraic* difference is the emf. That is, if the potential at the second point in the above example was −15 volts (a negative charge) the emf would be 100 − (−15), or 115 volts.

In any case, the free electrons in a conductor move toward the point of higher positive potential, obeying the rule of attraction and repulsion.

CAPACITANCE: Quantity of charge is proportional to potential; that is,

$$Q = CE,$$

where Q is the quantity of charge in coulombs, E is the potential in volts, and C is a constant called the capacitance of the charged object. The unit of capacitance is the farad. Capacitance increases with the size of the object, and also increases if the charge is distributed between two conductors by induction, if the conductors are insulated from each other. (This is because for the same total quantity of electricity the potential, E, is lowered by the presence of the nearby induced charge of opposite polarity.)

In a capacitor the charged conductors take the form of metallic plates or sheets placed close together and parallel, as in Fig. 1-4. Capacitance increases with the areas of the facing plates and decreases with increasing distance between them. The capacitance also can be increased by substituting a different substance for the air between the plates, the increase in capacitance depending on

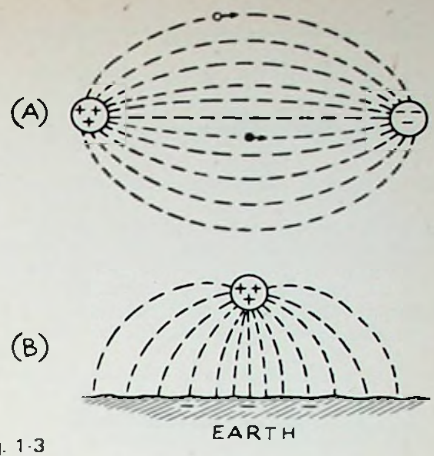


Fig. 1-3

the dielectric constant of the particular material used. This is summed up by the equation

$$C = \frac{kA}{d} K,$$

where C is the capacitance, k is the dielectric constant, A is the facing-plate area, d is the distance between plates, and K is a factor depending on the units in which C, A, and d are expressed. For C in microfarads, A in square inches, and d in inches, K is 0.224.

The area, A, in this equation is the area of *one* side of *one* plate if the capacitor consists of two plates that are alike. If more than two plates are used, alternate plates are connected together so that, effectively, there are still only two conductors, but of larger area. In such case the formula above should be multiplied by (n − 1), where n is the total number of plates in the capacitor. Stacking the plates with dielectric between all of them makes use of both sides of the plates, whereas the two-plate capacitor "wastes" the plate areas which do not face each other. This is also true of the outsides of the end plates in a multiplate capacitor.

DIELECTRICS: The dielectric constant is the ratio of the electric flux density set up in a material, for a given potential, to the flux density set up by the same potential in air, all other factors such as plate area, spacing, and so on being the same in both cases. The dielectric constant varies with different materials. The value 1 is assigned to vacuum (or air; the difference is extremely small). The dielectric constants of other materials are

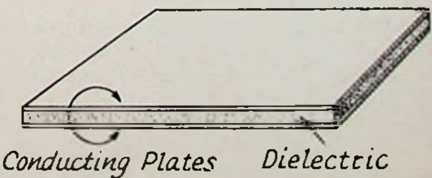
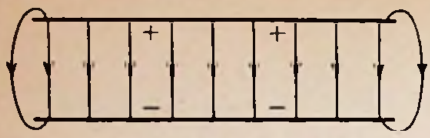


Fig. 1-4



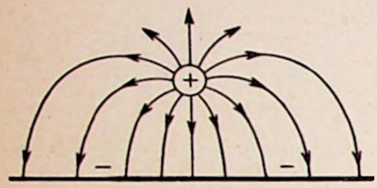


Fig. 1-5

established by substituting them for air in a capacitor of given dimensions whose capacitance with air dielectric is known. The measured capacitance with the new dielectric gives the capacitance ratio and dielectric constant.

In practical capacitors various dielectrics are used, depending on requirements. Solid dielectrics such as paper, glass, mica, and plastics have dielectric constants ranging from slightly over 1 to 5 or so. Certain ceramic materials have values of k as high as 1000 or more.

As the formula indicates, a capacitor with closely spaced plates (thin dielectric) will have greater capacitance than one with wide spacing (thick dielectric). However, the voltage that the capacitor will withstand, without breakdown of the dielectric because of the stress mentioned earlier, increases with the thickness of the dielectric. Thin insulating films on metal, with a conducting liquid or paste forming the facing "plate" (e.g., the **electrolytic** capacitor) result in high capacitance in a small space, but the safe voltage ranges from only a few volts to a few hundred, depending on the film thickness.

DISTRIBUTION OF CHARGE ON CONDUCTORS: Free electrons in an uncharged conductor are continually in random motion, moving about in such a way that the average potential is zero at every point. Because they have like charges the electrons repel each other, and in their effort to get apart they tend to distribute themselves on or near the surface of the conductor if no external field is acting on them.

When the conductor is charged (either an excess or deficiency of electrons) the charge similarly distributes itself over the surface. If the surface is smooth (flat or spherical, for example) the amount of charge per unit area will be the same at every point, but in general the actual charge distribution will depend on the nature of the surface. Charge tends to be more concentrated at protuberances, and particularly at sharp points or at edges.

In any case, the electric field lines always will leave the surface perpendicularly, because of the repulsion between similarly directed lines of force. This is illustrated by the two cases shown in Fig. 1-5. In (A) the field between the two conducting plates, seen edge-on in this view, is much stronger than the field between the outsides, although there may be appreciable "fringing" (indicated by the curved field lines) at the edges. In the wire shown in cross section above a flat plate (B) the field is more concentrated where the paths for the lines are shortest.

SHIELDING: Fig. 1-6 shows, in cross section, a charged metal sphere inside a completely closed insulated metal box. The positive charge on the sphere induces an equal negative charge on the inside surface of the box, and since the box itself is not charged an equal positive (opposite) charge must appear on the *outside* surface to keep the box as a whole electrically neutral. In effect, the electric flux lines extend straight through the box to terminate on some external body, such as the earth, which can supply the negative charge required for terminating them.

If the box is now connected to earth, as indicated by the arrowhead connection, the termination is in effect moved up to the box. Although the lines from the positive charge inside still terminate on the box, there is now no evidence of the charge outside because it flows off to earth. Thus the field of the charge on the sphere is confined inside the box so long as the box is grounded.

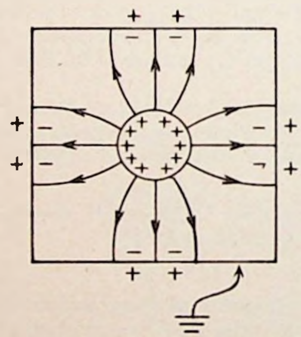


Fig. 1-6

By similar reasoning, a field external to the box cannot reach *inside* it, because the ground connection will not permit the box to be charged.

It is often necessary to confine electric fields in radio work, so such shields are widely used around radio-frequency circuits. The shape need not be a simple geometrical one such as is shown in Fig. 1-6; the actual shape does not matter so long as the shield completely surrounds the object to be shielded, has good conductivity, and is connected to earth or to an object that is sufficiently large, and has enough conductivity, to supply the required amount of terminating charge without appreciable change in its own potential.

Questions and Problems

1-1) What is meant by the electric field, and how is its strength described?

1-2) Define capacitance.

1-3) In what way would you expect the following factors to affect the capacitance of a capacitor?
- a) area of plates;
- b) separation between plates;
- c) dielectric material between plates;
- d) number of plates, when the capacitor consists of a set of interleaved plates with alternate ones connected together.

1-4) What is the nature of the force between two charges if
- a) both are positive;
- b) one is positive and one is negative;
- c) both are negative.

1-5) What is the meaning of electromotive force?

1-6) Name five conductors and five insulators.

1-7) What is the fundamental particle of electricity?

1-8) What is the nature of positive and negative electric charges?

1-9) Name the units for each of the following, giving a suitable definition in each case:
- a) quantity of electricity;
- b) capacitance;
- c) electromotive force.

1-10) Explain (a) how an insulated conductor can be charged by contact with a charged body; (b) how such a conductor can be charged by induction.

1-11) Is capacitance necessarily associated only with a capacitor?

1-12) What is the relationship between the polarity of the source of charge and that of a charged object when the latter has been charged by the former by induction? by direct contact?

1-13) If a capacitance of 4 microfarads is charged to a potential of 200 volts, what is the quantity of charge?

1-14) If the quantity of charge found in Q. 1-13 is placed on a 1-microfarad capacitor, what is the potential between the capacitor plates?

1-15) What is the difference between an ion and an electron?

1-16) What is the basic difference between a conductor and an insulator?

1-17) What is meant by saying that a certain insulating material has a dielectric constant of 2.5?

1-18) Why must a shield for the electric field be grounded?

Direct current; sources of emf; current and conductors; resistance; temperature coefficient of resistance; the circular mil; Ohm's Law; power; electrical energy; resistor construction.

Direct Current and Resistances

AN electric circuit is formed when a path is provided for current to flow. If the electromotive force applied to an ordinary circuit is constant the current also is constant, and if the polarity of the emf is fixed the current flows in only one direction through the circuit. A constant, unidirectional current is a direct current.

DIRECT CURRENT: Direct-current circuits must provide a continuous conducting path with no gaps of any kind that would prevent the passage of electrons. This is the type of circuit considered in this and the following chapter.

Strictly speaking, a continuous conducting path is required only for steady-state conditions. If the path is not completely conducting, current can flow – that is, electrons can move – at the instant the emf is applied and for a short time thereafter. (Induction, as described in Chapter 1, plays a part in this.) This short period during which a non-constant current flows is called the transient period. The steady state exists after the transient period is over – that is, when there is no longer a changing current in the circuit.

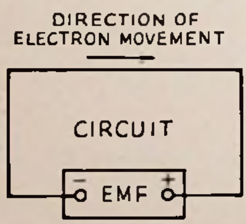


Fig. 2-1

SOURCES OF EMF: The electromotive force required for causing a direct current to flow through a circuit may be developed in a number of ways. The chemical cell is one of the oldest devices for this purpose, familiar in the form of either the dry battery or the storage battery. A primary cell, the common form used in dry batteries, is so called because the chemicals in it are gradually used up in producing electrical energy, and the cell eventually becomes useless. The storage battery or secondary cell can be recharged by forcing a direct current through it having the opposite polarity to the current normally produced by the cell. The chemical process in this case is reversible, the energy supplied to the battery during charging then being available to be taken from the battery in the form of direct current.

Mechanical energy can be turned into electrical energy by means of the electric generator, which uses electro-magnetism to produce the emf. In other cases, a converter may be used. The ac-to-dc converter changes alternating current into direct current having the desired emf. The dc-to-dc converter is used for changing the output of, for example, a storage battery into some other value of emf that may be needed for a particular purpose.

CURRENT and CONDUCTORS: Current flow is the movement of electrons or electric charge through a circuit. In general,

$$I = \frac{Q}{t}$$

where I is the current represented by a total amount or quantity of charge passing a given point in the circuit in time t. In practical units Q is the amount of charge in coulombs, I is the current in amperes, and t is the time in seconds. Thus one ampere is equal to one coulomb per second.

Because electrons are negative charges they are always attracted toward the point in the circuit where the emf has its highest *positive* polarity, and are repelled from points of more negative polarity. Thus the electron flow is from negative to positive, as indicated in Fig. 2-1.

The mass of an electron is unimaginably small. Thus for practical purposes the electron can be considered to have negligible inertia, and therefore moves instantly when it comes under the influence of the electric field from the emf. However, an individual electron does not move very far through a circuit before colliding with an atom and being absorbed. In this process it either dislodges an electron already in the atom, thus making the dislodged electron free to move, or else it drops into a vacant spot in an atom which has already lost an electron. As stated in Chapter 1, many of the atoms in conductors are missing an electron, and these free electrons roam around inside the material.

RESISTANCE: The collisions of electrons with atoms account for an inherent property of every circuit – and every electrical component – called resistance. Resistance can be defined as opposition to current flow, or as a property that limits the amount of current that can flow when a given emf is applied. Its value depends on the material through which the current flows, and is directly proportional to the *length* of the current path and inversely proportional to the *area* of the path.

For each conducting material, this property can be expressed in terms of its resistivity. Resistivity is defined as the resistance of a unit cube of the material. If resistivity is known, the actual resistance of a specific geometrical shape of the material can be calculated. However, as a practical convenience such calculations are usually made in terms of relative resistance, which is the ratio of the actual resistance of a substance of given size and shape to the resistance of an identical size and shape of annealed copper. The copper is assigned the value 1. Iron, for example, has a relative resistance of approximately 5.6, meaning that if two conductors have exactly the same size and shape (and are at the same temperature) one made of iron will have 5.6 times the resistance of one made of copper.

A circuit component whose function is to provide a desired amount of resistance is called a resistor.

TEMPERATURE COEFFICIENT OF RESISTANCE: Resistance also depends on the temperature of the material. In most metals temperature effects are small enough to be neglected if the temperature does not vary more than a few deg C. However, there are many cases where temperature changes in a resistor are considerable, and temperature must be taken into account if the resistance is to be known accurately.

In most metals the resistance increases when the temperature is increased (positive temperature coefficient) but in some materials, notably carbon, the reverse occurs (negative temperature coefficient).

The resistance of a conductor at a given temperature is

$$R = R_o(1 + at),$$

where R is the resistance to be found, R_o is the resistance of the same conductor at 0° C, a is the temperature coefficient of resistance (a negative sign is used instead of the plus sign if the coefficient is negative), and t is the actual temperature in deg C. For copper and other metals (not including special alloys which produce desired temperature/resistance relationships) the coefficient a can be taken to be about 0.0042. Fig. 2-2 shows the relative change in resistance over a representative temperature range.

If the resistance at 0° C is not known but is known for some other temperature, the formula

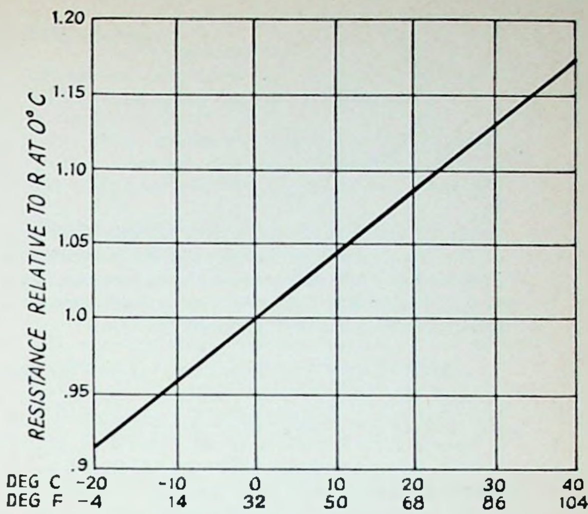


Fig. 2-2

can be transposed to find the resistance at 0°:

$$R_o = \frac{R}{1 + at}.$$

For example, if a conductor has a resistance of 10 ohms at 25° C, its resistance at 0° C is

$$R_o = \frac{R}{1 + (a \times 25)} = \frac{10}{1 + (0.0042 \times 25)}$$

$$= \frac{10}{1.105} = 9.05 \text{ ohms.}$$

If the resistance at 60° C is wanted, it is

$$R = 9.05\ [1 + (0.0042 \times 60)] = 9.05 \times 1.252$$
$$= 11.35 \text{ ohms.}$$

Certain alloys have very small temperature coefficients, so the resistance changes very little with large changes in temperature. The resistivities of these alloys are generally high, so they are used in making resistors for circuits that require fixed values of resistance that will not change appreciably with temperature.

THE CIRCULAR MIL: If a conductor is a round wire, as is often the case, its cross-sectional area is πr^2, where r is the wire radius. Since r is equal to d/2, where d is the wire diameter, the area is also equal to $\pi d^2/4$. For all round wires the cross-sectional area is proportional to d^2, therefore.

A round wire having a diameter of 0.001 inch is said to have a diameter of 1 mil. A circular mil, Fig. 2-3, is defined as the cross-sectional area of such a wire, so if the diameter of a wire of any size is expressed in mils, its area in circular mils is equal to its diameter in mils squared. This gives a convenient way of comparing the resistances, per unit length, of round wires of different diameters.

Wire tables give the circular-mil areas of standard wire sizes, along with the resistances per

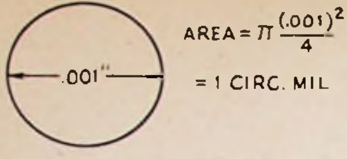


Fig. 2-3

thousand feet (usually for a temperature of 25° C). In the American Wire Gauge (B & S) the wire area doubles for each decrease of three numbers; e.g., No. 23 wire has half the circular-mil area, and therefore twice the resistance, of No. 20.

OHM'S LAW: The amplitude or size of the current that flows in a circuit is directly proportional to the amplitude of the applied emf, and is inversely proportional to the resistance of the circuit. This relationship, known as Ohm's Law, is expressed symbolically as

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

where I is the current, E is the emf, and R is the resistance. In practical units, I is measured in amperes, E in volts, and R in ohms. The transposed forms of the law,

$$E = IR$$

and

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

are equally useful. If any two of the three quantities, I, E, and R, are known the third can be found immediately from one of the above formulas. Fig. 2-4 shows the setup for measuring the current through a resistance and the voltage across it.

POWER: In a direct-current circuit the power dissipated – that is, either lost as heat or used for some desirable form of work – is equal to the applied emf multiplied by the current that it causes:

$$P = EI.$$

If E is in volts and I is in amperes the power, P, is in watts.

By substituting the expression for I from Ohm's Law, it follows that

$$P = \frac{E^2}{R}$$

and by substituting for E,

$$P = I^2R$$

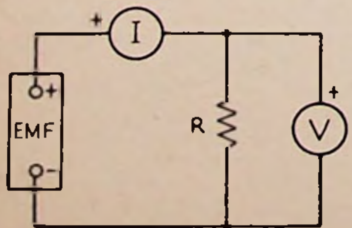


Fig. 2-4

In a resistor, the power represents the rate at which electrical energy is used in heating the resistance. The heat energy is that released by the electronic collisions mentioned above.

ELECTRICAL ENERGY: Potential energy is the *ability* to do a certain amount of work. The electric charge stored on a capacitor, for example, represents potential energy. When energy is actually being used to do work, it is called kinetic energy.

The distinction between energy and power is that power is the *rate* at which energy is expended. Expressed another way, the work done (energy expended) is equal to the power multiplied by the time it is being used. That is,

$$\text{Work} = Pt.$$

Energy or work done is measured in various units. Common ones are the watt-hour (watts × hours) and the kilowatt-hour (kilowatts × hours). For small amounts of energy the watt-second is frequently used.

RESISTOR CONSTRUCTION: Power dissipated in a resistor causes its temperature to rise, and for a given physical size the temperature becomes higher as the power dissipation becomes larger. As there is an upper limit of temperature which a given material can stand without damage, resistors must be made larger, for a given material, as the amount of power to be dissipated becomes greater. Also, the physical design must be such that the heat can be lost, either by convection or radiation, at a suitable rate for the size of the resistor.

Resistors which must dissipate relatively large amounts of power – 10 watts or more in communications circuits – almost always consist of a wire wound on a ceramic tube, since ceramics can stand high temperatures. The material, length, and diameter of the wire are chosen to give the desired values of resistance, as influenced by the required power dissipation. The wire is an alloy having high resistivity, low temperature coefficient, and ability to operate at high temperatures without damage.

Low-power resistors – from about 1/4 to 2 watts – for radio purposes are usually made of a carbon compound (composition resistors) and are as small in size as the power to be dissipated will permit. In many of their applications the temperature coefficient of resistance is not a critical factor. Such resistors are usually in the form of rods with wire leads taken off at the ends.

Adjustable resistors, including those in which the resistance can be continuously varied, can use either carbon or wire resistance elements. Continuously variable resistors in which the resistance is changed through a rotatable shaft are called potentiometers. For smoothest control they use carbon resistance elements, but this type is limited to a dissipation of a few watts. For higher power dissipations the wire-wound type is used.

Values of resistance used in radio circuits run from less than one ohm to several megohms, depending on the particular application.

Questions and Problems

2-1) How does conduction take place in metals?

2-2) What factors determine the resistance of a conductor?

2-3) How much electrical energy is used if a circuit carries a current of 75 mA for 10 minutes, the applied emf being 500 volts?

2-4) If 100 feet of copper wire having a diameter of 0.01 inch has a resistance of 5 ohms, what is the resistance of 200 feet of wire of the same material having a diameter of 0.02 inch?

2-5) Write the three forms of Ohm's Law.

2-6) Name a unit of electrical energy.

2-7) What factors determine the resistance of a conductor?

2-8) Write the formulas for power when any two of the three quantities, voltage, current, and resistance, are known.

2-9) In ordinary current flow does every electron go through the entire circuit?

2-10) What is the circular-mil area of a round wire having a diameter of 0.05 inch?

2-11) What is the difference between a primary cell and a secondary cell?

2-12 Explain the difference between resistivity and relative resistance.

2-13) What is the definition of the practical unit of electric current?

2-14) How would the resistance of the wire in Q. 2-10 compare, per foot, with the resistance of one having a diameter of 0.025 inch?

Questions concerning Exp. 1
(See Experiments Section):

2-15) In Exp. 1, Procedure (A), what is the approximate resistance of the milliammeter on the 0-50 mA range? On the 0-500 mA range? What are the corresponding meter resistances in *your* measurements, if you performed the experiment?

2-16) Calculate the power dissipated in each of the resistors used in Procedure (B). What is the power lost in each of the resistors used in *your* measurements?

2-17) Verify that the actual resistance of the "150-ohm" resistor used in Procedure (A) is 143 ohms if the tabulated voltage and current measurements are assumed to be accurate. Using this figure, what is the power dissipated in the resistor at each value of current?

Resistances in series; voltage drop; resistances in parallel; conductance; voltage dividers; voltage-divider design; voltage regulation; regulation vs. divider current; internal resistance of power supplies.

Resistance Combinations

IN MOST CIRCUITS there is more than one resistance to be considered in calculating the current that will flow when a given emf is applied to the circuit as a whole. These resistances may be combined in different ways, depending on how they are actually arranged in the circuit.

RESISTANCES IN SERIES: When two or more resistances are connected end-to-end with the emf applied to the ends of the string, as shown in Fig. 3-1, the resistances are said to be in series. The same current must flow through all, since there is no point where any current can be diverted into another circuit. The voltage measured between the terminals of any resistance in the string will be equal to the current multiplied by the value of that resistance.

Like the resistors, these voltages also are in series. Added together they must equal the applied emf:

$$E = E_1 + E_2 + E_3 = IR_1 + IR_2 + IR_3$$
$$= I(R_1 + R_2 + R_3)$$

and $$\frac{E}{I} = R_1 + R_2 + R_3 = R.$$

where R is the total resistance of the circuit. The equation can be extended to include any number of resistances.

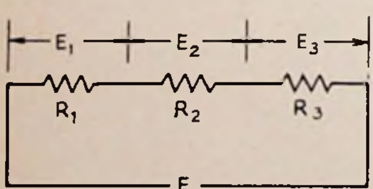


Fig. 3-1

VOLTAGE DROP: The voltages E_1, E_2, E_3, and so on, are called voltage drops. The voltage drop across any resistance in the string can have any desired value (less than the supply voltage, of course) if the value of the resistance is properly chosen in relation to the values of the other resistances.

A practical case is that of the voltage-dropping resistor, which is chosen so that the proper voltage is applied to some device, called the load, that takes a definite value of current at a definite value of voltage. Thus the load, too, has the characteristics of a resistance, its effective resistance being equal to the voltage applied to it divided by the current flowing through it.

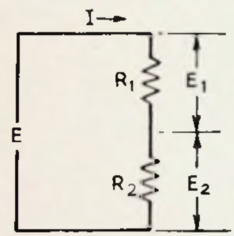


Fig. 3-2

In Fig. 3-2 a load is represented by R_2, which requires an operating voltage E_2 lower than the supply voltage E. If the required current is I, the dropping resistor R_1 must "lose" $E - E_2$ volts (that is, E_1) at the same current. The resistance required for R_1 is therefore

$$R_1 = \frac{E_1}{I}$$

RESISTANCES IN PARALLEL: The parallel connection of resistances is shown in Fig.3-3. In this case every resistance has the same applied voltage, E, but each carries a current that is determined by its resistance (Ohm's Law). Since all of the current must come from the source of emf, the total current is

$$I = I_1 + I_2 + I_3$$

and $$\frac{E}{I} = \frac{E}{I_1 + I_2 + I_3} = R.$$

Substituting E/R_1 for I_1, E/R_2 for I_2, and so on, leads to

$$R = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}}$$

The equation can be extended indefinitely.

For only two resistors in parallel the equation simplifies to

$$R = \frac{R_1 R_2}{R_1 + R_2}$$

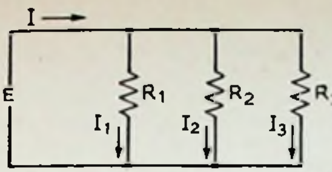


Fig. 3-3

CONDUCTANCE: It is sometimes simpler to handle circuits having resistances in parallel in terms of conductance rather than resistance. Conductance has the symbol G and is equal to the reciprocal of resistance – that is,

$$G = \frac{1}{R}$$

The unit of conductance, called the mho, is the reciprocal of one ohm.

Substituting G for 1/R in the parallel-resistance equation gives

$$R = \frac{1}{G_1 + G_2 + G_3}$$

or

$$\frac{1}{R} = G_1 + G_2 + G_3 = G.$$

Thus the conductance of the entire circuit is equal to the sum of the individual conductances. Any number of conductances can be in parallel.

VOLTAGE DIVIDERS: In Fig. 3-4 four resistors are in series across a source of emf, E. Since the voltage drops also are in series they add progressively, starting from one end such as *A* and measuring between *A* and each tap point. At the top end the total voltage is of course equal to E, the applied emf.

Such a system is called a voltage divider. Any desired voltage may be obtained between *A* and a tap point by appropriate choice of resistance values.

The voltages at the taps in Fig. 3-4 are open-circuit voltages. The open-circuit voltage will be available only if the load has infinite resistance and thus will not require any current. If current flows in the load the circuit is no longer simple, because the current from the source does not all flow through the four resistors; some of it leaves at one or more taps and flows externally through the load or loads and then back to the source through another tap or (more usually) through point *A*.

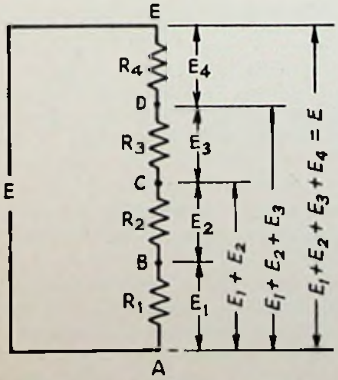


Fig. 3-4

A load on a voltage divider is equivalent to a resistance added directly between taps or between a tap and the common point (*A* in Fig. 3-4). This is a series-parallel case, since a combination of both types of connection exists in the circuit as a whole.

VOLTAGE-DIVIDER DESIGN: Designing a voltage divider is a straightforward step-by-step procedure. The method can be illustrated by Fig. 3-5, in which a voltage divider $R_1R_2R_3$ has a load R_4 connected to tap *B* and a load R_5 connected to tap *C*. The first step is to draw the circuit and label the resistances, voltages, and currents for purposes of calculation. Since E_4, E_5, and I_4, I_5 are determined by the loads, they are known quantities. E also is known. Begin by choosing a value

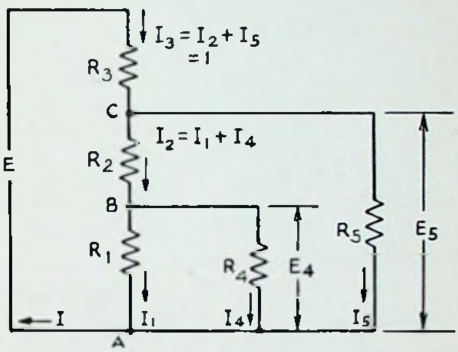


Fig. 3-5

for R_1 (the considerations determining this choice are discussed below) and calculate $I_1 (I_1 = E_4/R_1)$. I_2, the current through R_2, now is known $(I_2 = I_1 + I_4)$ and the voltage across it is the difference between E_5 and E_4. So

$$R_2 = \frac{E_5 - E_4}{I_1 + I_4} = \frac{E_5 - E_4}{I_2}$$

The current I_3 is the sum of I_2 and I_5 and the voltage across it is $E - E_5$. Hence

$$R_3 = \frac{E - E_5}{I_3}$$

And the design is complete. More complex arrangements can be solved by a similar procedure.

VOLTAGE REGULATION: The voltages at the loads will be constant only if the loads themselves have constant resistance under all operating conditions. This is not always the case, since the current required by a device may change. For example, the plate current of a vacuum tube or the collector current of a transistor may be varied by the gain control in a receiver. As current variations result in a change in the effective resistance of the device, the equivalent circuit also will be different because new values of load resistance are present. Changes in load currents cause the voltages at the tap points to vary from the figures that the divider was designed for.

The term for this change in voltage with load current is voltage regulation. The regulation of any source of power (a voltage divider is part of the source, in that it is a circuit associated with delivery of power from the source) is defined as the rise in voltage when the load is removed. It is frequently expressed as a percentage of the voltage when the load current is the rated value. That is,

$$\% \text{ Regulation} = \frac{E_o - E_L}{E_L} \times 100$$

where E_L is the voltage at full load and E_o is the open-circuit voltage.

REGULATION vs. DIVIDER CURRENT: In the design of a voltage divider it is always desirable to hold the voltage regulation to a minimum, consistent with other requirements. For good regulation the resistances *above* the taps must be low in comparison with the load resistances so that changes in current through them will result in only comparatively small changes in the output voltage of the divider. However, these resistances cannot be low unless the resistance of R_1 also is low. This means that the current taken by the divider alone must be large compared with the load currents taken from the taps.

The choice of R_1 has to be based on a compromise between good voltage regulation and having the divider take excessive power from the source. This compromise depends on balancing the importance of having good regulation against the importance of saving power, and has to be decided in each individual case, based on trial calculations.

If regulation is not a factor, as is the case when the loads will be constant, R_1 can be a high resistance or even can be omitted.

INTERNAL RESISTANCE OF POWER SUPPLIES: Voltage regulation is caused by the voltage drop inside a power source when current is being delivered to a load. In a chemical cell or battery this can be ascribed to an internal resistance which is equal to the voltage regulation in volts divided by the load current. The internal resistance of a battery varies with age and use, becoming greater the longer the service.

In other types of dc power supplies there are reactive as well as resistive effects. These will be covered later. The important point is that *every* power source has an internal voltage drop when it is supplying current to a load. Although this drop may be very small with some types of supplies, it often must be considered in circuit design. Also, while it is possible to assume that the supply voltage is constant and is in series with some value of internal resistance, this usually can be done with accuracy only at a specific value of load current, or over only a small range of load currents.

Questions and Problems

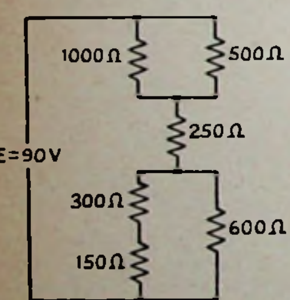


Fig. 3-6

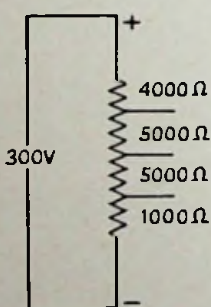


Fig. 3-7

3-1) Three resistances, 5, 14, and 22 ohms, are connected in parallel. What is the resulting resistance? If 6 volts are applied to the combination, what is the total current, the current through each resistor, and the power dissipated in each?

3-2) If a current of 350 microamperes flows through a circuit with an applied voltage of 40, what is the resistance of the circuit?

3-3) A resistance of 50,000 ohms is connected in parallel with one of 25,000 ohms. What is the resultant resistance?

3-4) In Fig. 3-6, find the current through each resistor and the voltage across it.

3-5) If three resistors, 10,000, 40,000, and 12,000 ohms, are available, how can they be connected to give a total resistance of 20,000 ohms?

3-6) If the power consumed in a 50,000-ohm resistor is 2 watts, what is the applied voltage? What is the current through the resistor?

3-7) What are the voltages between the negative terminal and the tap points in Fig. 3-7?

3-8) How may a source of power be represented for determining its output voltage at various load currents?

3-9) What is meant by load resistance?

3-10) A voltage divider is to supply a load current which may range between 0 and 20 mA. The source voltage is constant at 90 V, and the voltage at the load must not drop below 50 at any time. Design a voltage divider to meet these conditions.

3-11) Using your design for Q. 3-10, what will be the variation in load voltage when the load current varies throughout the specified range?

3-12) What is the voltage regulation of the divider in Q. 3-10? If it is more than 30%, or if the total current taken from the 90-volt source with *no* load on the divider is more than 70% of the total current at full load, redesign the divider.

3-13) What is the power lost in the divider resistors, at full load, in your final design of the divider in Q. 3-10? At no load?

Questions concerning Exp. 2

(See Experiments Section)

3-14) Using the data of Exp. 2, Procedure (A), what is the actual power loss in each resistor in each of the four cases if the rated resistance values are exact?

3-15) Calculate the internal resistance of the power supply from the data you obtain from performing Procedure (B).

Chapter 4

Magnetic field; temporary and permanent magnetism; electromagnetism; magnetizing force; permeability; saturation; reluctance; hysteresis; induction; inductance; mutual inductance.

Electromagnetism and Inductance

A MAGNET IS AN object, usually of iron or one of its various compounds or alloys, that exhibits the phenomenon of magnetism. Very few nonferrous materials have this property, and therefore are unaffected by the presence of a magnet.

MAGNETIC FIELD: A magnetic field, analogous to the electric field, surrounds a magnet, and magnetic force will be exerted on magnetic materials that are in the field.

Magnets, like electric charges, exhibit polarity. Thus the magnetic force, like electric force, will be either attraction (unlike poles) or repulsion (like poles). Magnetic poles are designated north-seeking and south-seeking instead of positive and negative. Both types always exist simultaneously in the same magnet; there cannot be an isolated north pole or an isolated south pole.

A line of force which leaves a north pole returns externally to the south pole *in the same magnet*, thus magnetic lines of force are always closed upon themselves. The differences between the electric and magnetic fields are illustrated by Fig. 4-1.

In space, magnetic force, like electric force, decreases with the square of the distance from the source. The field intensity or flux density is expressed by the number of lines of force per unit of cross-sectional area perpendicular to the lines of force.

TEMPORARY and PERMANENT MAGNETISM: All magnetic materials become magnetized when subjected to a magnetic field, and thereafter exhibit magnetism themselves. However, the magnetism is not retained indefinitely. The tendency to retain magnetism is called retentivity. It varies greatly with the magnetic material, and magnetized objects may be either temporary magnets or permanent magnets, depending on whether the magnetism disappears in a short time such as a fraction of a second or whether it continues to exist for long periods – many years, for example. Soft iron and the type of steel used in electrical machines have little retentivity, but other materials – Alnico being an example – have very great retentivity.

ELECTROMAGNETISM: Moving electrons are always surrounded by closed magnetic lines of force lying in a plane perpendicular to their motion. That is, an electric current generates a magnetic field which surrounds the conductor carrying the current and is at right angles to the direction of current flow. This is shown in Fig. 4-2. The field disappears when the current is shut off, because air has no magnetic retentivity.

If the conductor is formed into a coil, the field (in the coil cross section) resembles that shown in the lower drawing of Fig. 4-2. The upper row of turns in cross section is shown by the solid circles and the lower row by open circles to indicate that the direction of current flow is *into* the paper in one row and *out of* the paper in the other. The coil is an electromagnet, and has a much stronger magnetic field than a simple conductor because the flux lines surrounding the turns nearly all add together to increase the flux density. Only those

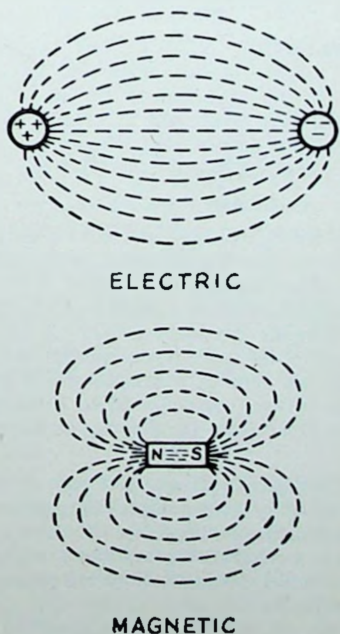
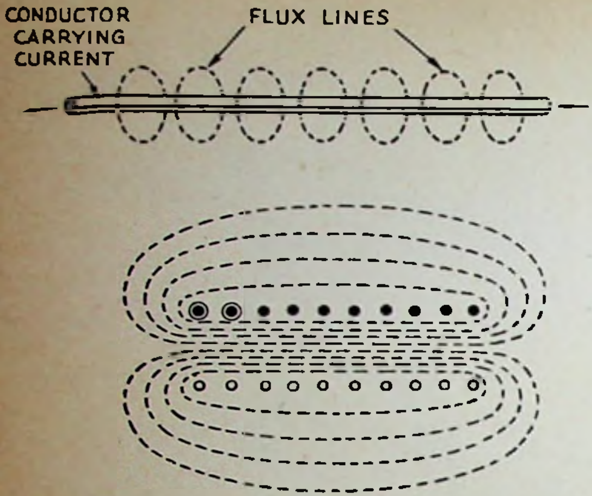


Fig. 4-1



CROSS SECTION OF COIL WITH FLUX LINES

Fig. 4-2

flux lines that are very close to the wire do not add to those of at least the adjacent turn. Two such lines are indicated by the circles surrounding the two left-hand turns in the upper row; the case would be similar for the remainder of the turns.

The strength of the magnetic field set up by an electromagnet depends on the amplitude of the current, the number of turns in the coil, the shape of the coil, the magnetic properties of the material surrounding the coil, and the length of the magnetic path. Nonferrous materials are in general the same as air in their magnetic properties.

MAGNETIZING FORCE: The magnetizing force or magnetomotive force in an electromagnetic circuit, analogous to emf in an electric circuit, is usually expressed in ampere-turns per unit length. One ampere-turn is equal to one turn carrying a current of one ampere. That is,

$$\text{Ampere turns} = NI,$$

where N is the number of turns and I is the current in amperes, so that the ampere-turn product gives the magnetizing force per unit length.

The magnetic flux in a path is proportional to NI and inversely proportional to the reluctance (see below) of the path. Consequently if there is more than one type of substance in the path the reluctances of each must be added to obtain the total reluctance before the flux can be determined.

PERMEABILITY: If an iron core is inserted into the coil of an electromagnet the magnetic flux density is greatly increased for the same current. The increase is greatest when, among other things, the path for the magnetic lines is completely in magnetic material.

The ratio of flux density produced in such a closed magnetic path to that produced in air by the same coil and current is called the permeability of the core material. Air is assigned a permeability of 1. Maximum values of permeability in some materials may be as high as 1,000,000. The "transformer iron" (silicon steel) used in power and audio transformers has a maximum permeability of 5000 or more.

SATURATION: As the magnetizing force is increased, the iron in a magnetic circuit tends to become magnetically saturated – that is, a point is reached where a further increase in magnetizing force will not cause a proportional increase in field strength. See Fig. 4-3. The saturation point varies with different magnetic materials. (Air and nonmagnetic substances do not saturate.) Because of saturation, permeability goes through a maximum value at some point below the level of magnetizing force at which saturation effects begin to become noticeable. The permeability is greatest at the point on the curve in Fig. 4-3 where the slope is maximum.

RELUCTANCE: Magnetic field lines are closed upon themselves, as stated earlier, and a path for them therefore has some resemblance to an electric circuit, which also is closed upon itself before current can flow. It is possible, therefore, to draw a parallel between the continuous magnetic field and current flow in an electric circuit.

Just as materials used in conductors have resistivity, substances in a magnetic path have the property of reluctivity. Reluctivity is the reciprocal of permeability, and since the permeability of all nonmagnetic materials is 1, the reluctivity of such materials also is 1. From the known reluctivity of all parts of a given magnetic circuit it is possible to determine the reluctance of the path. Reluctance is the magnetic equivalent of resistance in the electric circuit, and, like resistance, it increases with the length of the path and decreases with the cross-sectional area, along with depending on the material or materials used.

Reluctances combine in series and parallel in the same way as resistances, and similar formulas hold.

Reluctance is important in calculating the performance of magnetic circuits comprised chiefly of magnetic materials, and is seldom needed in calculations involving only air paths for the magnetic flux, as in air-core coils used at radio frequencies. Its value, for a given path, will vary if the permeability varies with flux density, as is the case in magnetic materials. Resistance, on the other

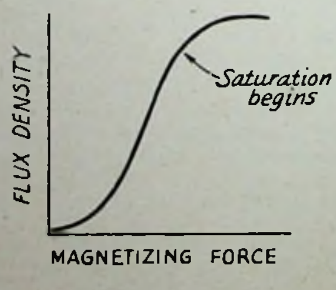


Fig. 4-3

hand, is independent of the amount of current flowing through it.

HYSTERESIS: In electromagnetic circuits operating on ac the direction of the magnetizing force is continually being altered. If the resulting magnetization of a core is plotted as in Fig. 4-4, starting at *A* with no residual flux and no magnetizing force and going through successive alternations, a hysteresis loop results. At the maximum force applied to the right the flux reaches point *B*; then as the force decreases (still applied in the same direction) the curve follows the downward arrow. At zero force the flux has not returned to zero but has some value *D*. This is the result of the retentivity of the magnetic material. The magnetizing force reverses after reaching zero, and must demagnetize the core before zero flux, *E*, is reached, after which the maximum *reverse* magnetization is reached at *C*. The curve can be followed through the complete cycle from there, and it is seen that the flux curve is actually a loop and not a straight line.

The hysteresis loop is a measure of power lost in the core in overcoming the effect of magnetic retentivity. This power causes heating of the core. An air core has no hysteresis loss.

INDUCTION: A flux line around a conductor and the current in the conductor are said to link with each other, since both the flux and current have closed paths (a closed path is necessary for current flow). The number of flux linkages in a coil carrying current is equal to the number of turns in the coil multiplied by the total number of flux lines which the current sets up in the core. (If the core is not made of magnetic material the coil is said to have an air core).

Any *change* in the flux linkages causes a voltage to be induced in the conductor or coil. The polarity of the induced voltage is always such as to *oppose* the current change that generated it. This is called **Lenz's Law**. The change in linkages may be the result of a change in the amplitude of the current (which in turn causes a change in flux), or the result of moving the conductor through an external magnetic field in such a way that the field lines are cut by the moving conductor. (This is the principle used in electrical machinery.)

The induced voltage is proportional to the *rate of change* of flux linkages. If there is no change in the linkages there is no induced voltage, even though a magnetic field may be present.

INDUCTANCE: The inductance of a coil (or any conductor, of whatever shape) is defined as the number of linkages per unit of current. The practical unit of inductance is the henry. Since linkages increase with the number of turns in a coil (except in those cases, indicated in Fig. 4-2, where a flux line links only with the current that causes it), it follows that inductance also increases with the number of turns. It follows, too, that putting an iron core in a coil increases the inductance, because the flux is increased by the iron for the same current. Finally, the definition indicates that

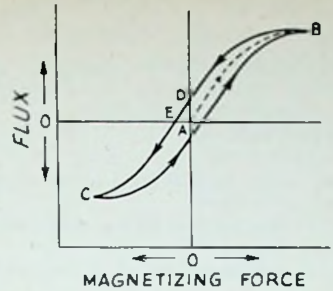


Fig. 4-4

for a coil of inductance L the self-induced voltage is equal to L multiplied by the rate of change of current through the coil.

When a coil is carrying a steady direct current there is no induced voltage, because there is no change in such a current (however, a voltage is induced when the current is started or stopped, because at those instants it *is* changing).

By Lenz's Law the polarity of the induced voltage is such as to cause a current which opposes the change in the original current. Thus if the current through the coil is increasing (applied voltage is increasing) the induced voltage opposes the applied voltage and causes a current which subtracts from the original current; if the current is decreasing, the induced current adds. In other words, the effect of self-induction is always to oppose *any* change in current, or any change in applied voltage.

MUTUAL INDUCTANCE: Two coils in the same magnetic field are said to be magnetically coupled. If the field changes, voltages are induced in both, in proportion to the flux linkages in each.

In general, not all of the magnetic flux set up by a coil carrying current will link with the turns of a second coil coupled to it. The ratio of the flux linking the second coil to the flux set up by the first is called the coefficient of coupling. This factor (usually designated k) is always less than 1. In practice, it may be quite close to 1 for two coils wound on the same closed magnetic core, but will be considerably less for air-core coils, depending on their construction, separation, and orientation with respect to each other.

Coupled coils are said to have mutual inductance, which is defined as the voltage induced in the second coil by the rate of current change in the first. The mutual inductance, M, of two coils, L_1 and L_2, is

$$M = k\sqrt{L_1 L_2}$$

where k is the coefficient of coupling and L_1 and L_2 are the respective inductances of the two coils. The unit of mutual inductance is the henry – the same unit that is used for self-inductance.

The coil carrying the original current is called the primary coil and the other the secondary coil. Either of the two coupled coils can be used as the primary or secondary, depending on circuit requirements.

Questions and Problems

4-1) Describe self-induction.

4-2) What is the nature of the force between
- a) a north pole and south pole;
- b) two north poles;
- c) two south poles.

4-3) Under what conditions can a voltage be induced in a conductor?

4-4) Is inductance necessarily associated only with wire wound in a coil?

4-5) How is the intensity of a magnetic field described?

4-6) Name the unit of inductance, and give its definition.

4-7) Explain what is meant by the term permeability.

4-8) What factors determine the inductance of coil?

4-9) What is the direction of flow of an induced current compared to the direction of flow of the current causing the induction?

4-10) When the current through a coil is broken, is the induced voltage larger or smaller than the voltage induced when the current is started? Why?

4-11) Upon what factors does the strength of the magnetic field set up about an electromagnet depend?

4-12) What is hysteresis, and why does it occur?

4-13) What is the definition of mutual inductance? Upon what does its value depend?

4-14) Two coils each having an inductance of 1 millihenry are coupled together. Measurement shows that the mutual inductance is 400 microhenrys. What is the coefficient of coupling?

4-15) What would you expect the effect of magnetic saturation to be on the inductance of an iron-core coil?

4-16) What is reluctance, and how does it compare with the analogous quantity in an electric circuit?

4-17) Why is the inductance of a wire increased by winding it into a coil?

4-18) Two coils having identical overall dimensions are wound on identical iron cores. If one has 500 turns and carries a current of 0.1 ampere while the other has 50 turns and carries 2 amperes, what is the relative magnetic force in the two cases?

4-19) Does an electron at rest have a magnetic field?

4-20) How does hysteresis affect a magnetic circuit?

Chapter 5

Energy storage in capacitance; energy in the magnetic field; capacitances in parallel; capacitances in series; inductances in series; inductances in parallel; time constant; rise and decay time.

Energy in Electric and Magnetic Fields

IN CHAPTER 1 it was pointed out that the definition of potential involves the ability of the electric field to do work. By analogy, the same thing is true of the magnetic field. Both types of field represent potential (ready to be used) energy.

ENERGY STORAGE IN CAPACITANCE: An electric charge placed on a capacitor represents stored energy – energy which came from the work done in separating electrons from the atoms with which they had been associated. This energy is stored in the electric field in the dielectric between the plates, whether that dielectric is air or some other material. Since the quantity of charge depends on the capacitance, for a given difference of potential or voltage between the plates, the amount of energy stored likewise depends on the capacitance.

If two capacitors are charged to the same voltage, the one with the larger capacitance will have the greater amount of stored energy. Another consequence of the dependence of charge on capacitance is that if two capacitors have the same amount of stored energy, the one with the lower capacitance must have the higher voltage between its plates.

If energy is stored in a capacitor and then left in storage (capacitor disconnected from the circuit or so situated that the voltage across it is unchanging) the energy remains potential. It is only when the voltage across the capacitor is *changing* that energy is either being stored in or being withdrawn from the capacitive reservoir. If the voltage across the capacitor is increasing, energy is being stored; if decreasing, it is being released from the capacitor.

ENERGY IN THE MAGNETIC FIELD: A similar situation exists in the case of inductance. Here the energy is stored in the magnetic field which is set up around the inductor by the current flowing through it. A steady direct current flowing through inductance will set up a steady, unchanging magnetic field, and in such a field the energy is *potential*. When the current is changing, energy is being either added to the field or taken from it; increasing current means more energy being stored, and decreasing current means that energy is being returned to the circuit.

While a steady direct current is required for "quiet" conditions in an electromagnet, no *voltage* is needed to maintain the field. In contrast, the electric field is "quiet" when there is a steady voltage but no current. No electrical power is needed for maintaining a steady field of either type, because if either the voltage or current is zero the power (current × voltage) likewise is zero. This is true only if the inductance and capacitance are "pure" – that is, have no resistance to dissipate power. A capacitor with very low leakage approaches this ideal condition, but an inductor never has zero resistance. However, any power wasted in the capacitor or inductor is not used in the capacitance or inductance as such, but in heating the resistance.

CAPACITANCES IN PARALLEL: If two capacitors are connected in parallel, as in Fig. 5-1(A), and charged from the same source, the voltage between plates is obviously the same in both. The charge on each is therefore in proportion to its capacitance, and the *total* charge is the sum of the two. This is equivalent to saying that the total capacitance of two capacitances in parallel is the sum of the individual capacitances. The statement is true for any number of capacitors in parallel.

CAPACITANCES IN SERIES: If the two capacitances are connected in series and to a source of

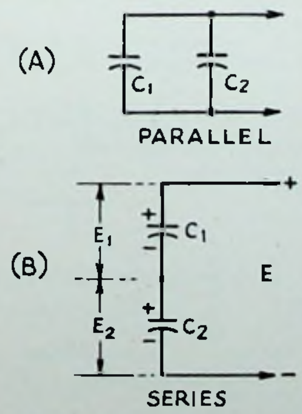


Fig. 5-1

voltage having the polarity indicated in Fig. 5-1(B), electrons are drawn from the top plate of C_1, leaving it with a positive charge. An *equal number* of electrons is attracted to the bottom plate by induction; these are drawn from the top plate of C_2. In turn, the resultant positive charge on C_2's top plate draws an equal number of electrons from the source into the bottom plate. Thus all plates have either an excess or deficiency of electrons, the quantity being the same in each case. This is also the quantity of charge taken from the source.

The voltage across each capacitor is equal to the quantity of charge divided by the capacitance. This is true also of the *total* capacitance C, and the source voltage, E. Then since Q is the same in all cases and $E_1 + E_2$ must equal E,

$$\frac{Q}{C} = \frac{Q}{C_1} + \frac{Q}{C_2} \quad \text{or} \quad \frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2}$$

Thus

$$C = \frac{1}{\frac{1}{C_1} + \frac{1}{C_2}}$$

which is similar to the expression for resistances in parallel, and can be extended in the same way for any number of capacitances in series.

INDUCTANCES IN SERIES: If the two inductances in Fig. 5-2(A) are not magnetically coupled in any way (no mutual inductance) each sets up a number of flux linkages in proportion to its own inductance when the current, I, flows (Chapter 4). Since the same current flows through both coils the total number of linkages, and therefore the total inductance, is the sum. The total inductance of any number of inductances (having no mutual coupling) in series is the sum of their individual inductances.

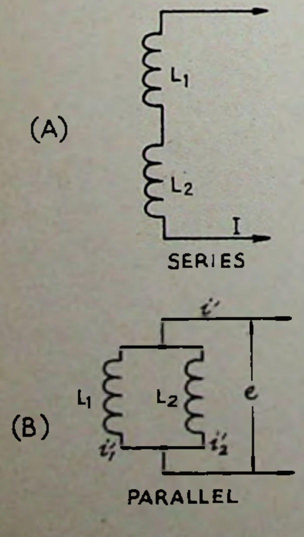


Fig. 5-2

INDUCTANCES IN PARALLEL: When inductances are connected in parallel, Fig. 5-2(B), the relationships are similar to those for resistances in parallel. (The inductances are assumed to have no mutual coupling.) In this case it is necessary to deal with the *instantaneous* induced voltage, *e*, and *rate of change* of current, here labelled i', instead of the steady voltages and currents used in analyzing the parallel-resistance circuit. Using similar reasoning (see Chapter 3) leads to

$$L = \frac{1}{\frac{1}{L_1} + \frac{1}{L_2}}$$

and so on for any number of inductances in parallel.

TIME CONSTANT: When capacitance or inductance are in a circuit which also contains resistance, a finite time is required to build up the field in which energy is stored. Also, a finite time is required for the energy to be withdrawn. This time-dependence is expressed by the time constant of a capacitance-resistance or inductance-resistance circuit.

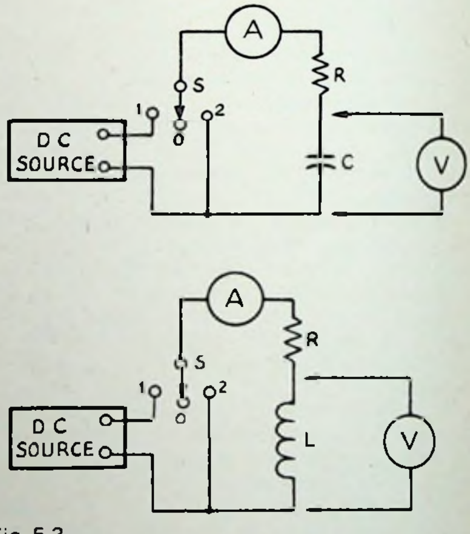


Fig. 5-3

Referring to Fig. 5-3, if the switch S is moved from the open position (0) to position 1, current begins to flow into the capacitor C through resistance R. At the instant of closing S the maximum current that can flow is equal to the source voltage divided by R and the voltage measured by the voltmeter is zero. The voltage then rises at a rate determined by the rate at which charge can be placed on C through R, until eventually V will read the source voltage. (Theoretically, the capacitor voltage never quite reaches the source voltage; practically, the difference between the two becomes unmeasurably small after a time.) If S is moved to 0 again, the charge remains on the capacitor indefinitely, depending on the leakage (gradual loss of charge through a path for current flow either in the capacitor itself or externally). The time in seconds required for the voltage across C to reach 63% of the final value during charging is the time constant of the CR circuit. If S is now moved to position 2, the capacitor will discharge through R, and the same amount of time will be required for the voltage across C to decrease to 37% of its fully charged value. These figures are based on the fact that the voltage increases and decreases logarithmically.

In the inductance-resistance circuit, Fig. 5-3, on moving S to position 1 from 0 the current begins to rise from zero, being held back at first by the opposing self-induced voltage generated in L by the change in current. As time goes on, the rate of change becomes smaller and the current finally is determined solely by R and the source voltage. (Again, in theory the current never quite reaches this value, but in practice the difference is too small to be measured after a short time.) The voltage measured across L thus starts at the source voltage and decreases to zero. The time constant of such a circuit is defined for current in the LR circuit in the same way as for voltage in the CR case, and is equal to L/R.

If S could be moved from 1 to 2 *instantaneously*, the current would at first equal the voltage induced in L divided by the circuit resistance R (the induced voltage is the result of cutting off the steady current). It would then gradually decrease to zero, reaching 37% of its initial value in the time constant of the circuit.

RISE and DECAY TIME: The way in which voltage and current rise or decay in CR and LR circuits is shown in Fig. 5-4. As applied to the capacitive circuit in Fig. 5-3, the "rise" curve shows how, on charging, the voltage V increases with time, and the "decay" curve shows how the charging current A decreases with time. On *dis-*

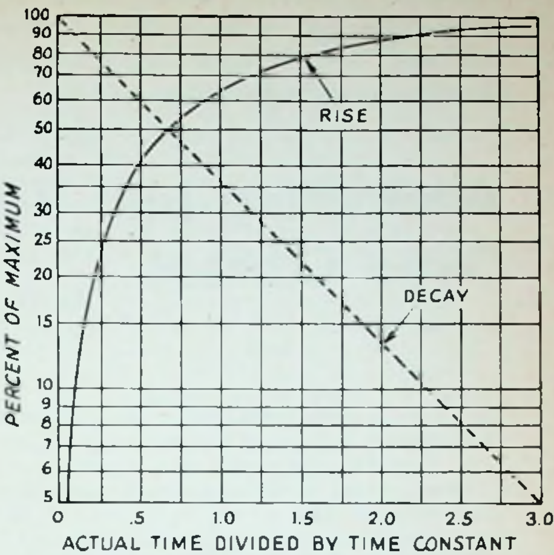


Fig. 5-4

charge the "decay" curve shows both current and voltage.

Similarly, in the LR circuit the "rise" curve shows how the current builds up, and the "decay" curve gives the voltage change across the inductor when the field is being created.

These curves are plotted against the number of time constants of elapsed time – that is, *actual* time divided by the time constant. Thus if a circuit has a time constant of 0.25 second, the voltage or current would decay to 5% of the maximum value in 3 × 0.25 = 0.75 second. Actual rise times are calculated in the same way, using the "rise" curve.

Questions and Problems

5-1) An inductance of 10 henrys is connected in series with one of 15 henrys. What is the total inductance if the fields of the two inductances do not interact?

5-2) If the two inductances of Q. 5-1 have resistances of 100 ohms and 150 ohms, respectively, and a current of 0.1 ampere is flowing through them, how much power is required for maintaining the magnetic field of each?

5-3) What is the total inductance of the two inductances of Q. 5-1 if they are connected in parallel?

5-4) What is the time constant of an LR circuit consisting of 6 henrys and 400 ohms? How long would it take the current to rise to 95% of its final value after application of voltage?

5-5) If two 8-μF capacitors are connected in series, what is the total capacitance?

5-6) Define time constant. What is the time constant of a circuit having only a resistance of 500 ohms?

5-7) What is the time constant of a circuit consisting of a 4-μF capacitor in series with 150,000 ohms?

5-8) A 10-μF capacitor is charged to 325 volts. If the voltage is to decrease to 25 volts in 1 second after cutting off the supply voltage, how much resistance must be placed in parallel with the capacitor?

Questions concerning Exp. 4

(See Experiments Section)

5-9) Is the agreement good between the experimental curve of Fig. E4-2 and the theoretical curve of Fig. 5-4? (Hint: Convert the times given in the tabulated data to fractional time constants.)

5-10) If you performed Exp. 4, calculate the value of the capacitor you used, by the method described in the comments on the experiment.

Varying direct current; average value; period and frequency; alternating current units; power in combined ac and dc; phase; angular measure; the sine wave; complex waveforms, harmonics; ac in resistance; linear and nonlinear circuits.

Alternating and Varying Direct Current

AN ALTERNATING CURRENT is often defined as one whose direction of flow reverses periodically. This is descriptive of a large proportion of the currents used in power and communications work. However, there are also many cases where the current does not reverse direction, but nevertheless has one of the distinguishing characteristics of ac in that energy can be transferred by induction, without the unbroken conducting path that is required for direct-current flow.

VARYING DIRECT CURRENT: The essential factor in induction is *change*. For example, if there is even a momentary change in a direct current, as when the current is started or stopped or its amplitude shifts with a change in load, the *variations* cause changes in the electric or magnetic fields, and these changes cause induced voltages and currents.

A great many of the signal currents in communication are periodic, or quasi-periodic, and have the form of variations in direct current without actual reversal of polarity. They occur in amplifying devices such as vacuum tubes and transistors. Such variations constitute an alternating current superimposed on the direct current, and by using induction the ac can be separated from the dc.

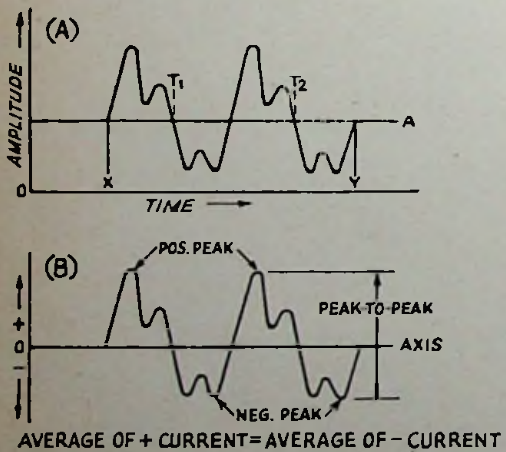


Fig. 6-1

Fig. 6-1(A) shows a direct current normally having the value A on which a repetitive variation is superimposed during the time X-Y. The result of separating the ac from the dc is shown in Fig. 6-1(B), where the dc component has disappeared and the current variations now swing back and forth, positive and negative, about an ac axis or zero axis. Transient effects which will occur when the ac begins and ends have been ignored in this picture, as they depend on the type of coupling and the circuit; Fig. 6-1(B) represents a steady-state condition such as would occur throughout all but the beginning and ending stages of a long series of repetitive variations.

AVERAGE VALUE: The average value of any current is the average of all the values that the current goes through during some specified length of time. The selection of a suitable time interval depends on what is being measured. In Fig. 6-1(A) the average value is A. This is the current that would be read by a dc meter.

When current variations are transferred by induction the average induced-current flow in one direction cannot be greater than in the other. This is because the excess in one direction would constitute a direct current in that direction, and direct current cannot be transferred by induction. Consequently, the overall average value of an alternating current *has to be zero*, which means that the ac axis is compelled to set itself so that the required equality of average currents in both directions is maintained. However, the amplitude variations in the two polarities may be quite different from each other even though the averages must be equal. Fig. 6-1(B) illustrates this point.

PERIOD and FREQUENCY: A graph of the current (or voltage) amplitude variations with respect to time shows the waveform of the alternating current. If the waveform repeats itself regularly, taking on identical shape in each successive equal-time interval, the waveform is periodic.

Each repetition is called a cycle. The period is the length of time occupied by one cycle. The beginning and ending points of one period or cycle are marked by T_1 and T_2 in Fig. 6-1 (A).

ALTERNATING-CURRENT UNITS: The fact that an alternating current or voltage is continually varying in amplitude and polarity makes it necessary to adopt a system of units that will accommodate these variations.

The instantaneous value of the voltage or current is the amplitude that occurs – in transit, as it were – at any given instant. A system that is of considerable value in communications work is that of specifying the *maximum* instantaneous values attained by the waveform, since these frequently determine the operating limits of amplifying circuits. There are two such maxima, the **positive peak** and the **negative peak**, as shown in Fig. 6-1(B). The larger of the two, which is generally of more interest, is simply called the **peak value** of the wave. The absolute sum (polarity ignored) of the positive and negative peaks is called the **peak-to-peak value** and represents the total voltage or current excursion during the cycle.

As explained earlier, the average of the instantaneous values of voltage or current is always zero in an ac waveform. However, it is important to have a definite number for measuring some property which two currents (or voltages) have in common so that different currents can be compared on an understood basis. A solution is found by using *power* as the reference.

Power in an alternating waveform is obviously varying throughout the cycle. The **instantaneous power** is found by multiplying the instantaneous voltage by the current at that same instant, and thus is defined in the same way as in a dc circuit. However, the average power – defined as the average of all the values that the power goes through during a complete cycle – is constant from cycle to cycle if the waveform is periodic. Furthermore, if the power is dissipated – that is, is used up in heat or doing some other form of work – the average during a cycle is not zero; the energy used in doing work cannot be created out of nothing.

An ac ampere is therefore defined as one that will result in the same heating of a given resistance as a dc ampere; similarly for the ac volt. These are known as **effective** or **rms** values of the ampere and volt in ac circuits. (The abbreviation stands for "root-mean-square," the mathematical process of deriving the average power from the instantaneous-power values during the cycle.)

Dissipated power is sometimes called "real" power to distinguish it from "imaginary" power, which is not really imaginary but is power used in storing energy in an electric or magnetic field and is later returned to the circuit as outlined in Chapter 5. A better name for it is **reactive power.**

POWER IN COMBINED ac and dc: A dc meter indicates the average of the current through it taken over an appreciable fraction of a second, and if the current varies at a rate of more than a few cycles per second the pointer cannot follow. And since the average of an alternating current is zero, the meter reads zero on ac.

Hence when such an instrument is in a circuit carrying both ac and dc, as in amplifiers, it ignores the ac and reads only the dc. In such circuits, therefore, the total power is not that ostensibly given by multiplying the dc voltage by the direct component of current. The power is actually greater when the alternating component is present, and is equal to the ac power plus the dc power.

PHASE: An important factor in ac circuit operation is the time at which the instantaneous current or voltage goes through some specified point in the waveform. This is called the **phase** of the wave. If there are two waves the **phase difference** between the two can be expressed either as a time difference or, if the waveform is periodic, as the appropriate fraction of either the period or cycle. The cycle is generally the preferred reference.

ANGULAR MEASURE: A very useful method of specifying an exact point in a periodic waveform, particularly with the sine wave discussed below, results from expressing the time at which it occurs, measured from the beginning of the period or cycle, as an angle (phase angle) related to the total angle of a circle. That is, the cycle is assumed to be equivalent to 360 degrees, and the fractional cycle is directly related to the equivalent fraction of 360 degrees. For example, a quarter cycle is 90 degrees, a half cycle is 180 degrees, 5/8 cycle is 225 degrees, and so on. The beginning of the cycle (or period) usually is assumed to be the point at which the wave goes through zero moving in the positive direction.

THE SINE WAVE: Periodic waveforms can have any shape. An extremely important one is the sine wave, shown graphically in Fig. 6-2. It is so called because the amplitude at any point is proportional to the sine of the angle which measures the position of that point in the cycle. The effective or rms value of a sine wave is its maximum or peak value divided by $\sqrt{2}$ – that is, the peak value multiplied by 0.707.

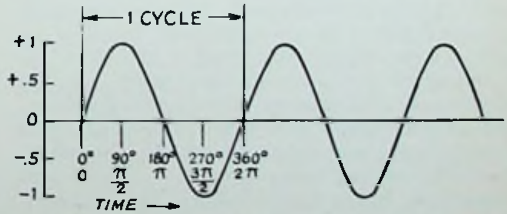


Fig. 6-2

Fig. 6-2 also shows the angular measure of the cycle in both degrees and radians. There are 2π radians (the radian is the fundamental unit of angle) in a circle, and the quantity $2\pi f$, radians per cycle multiplied by cycles per second, enters into

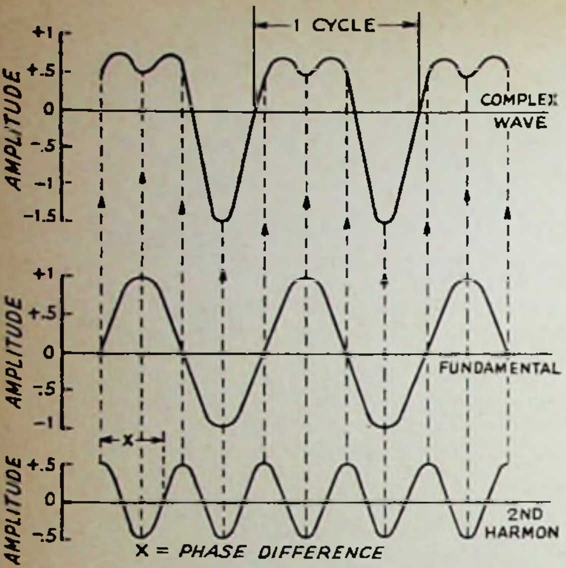


Fig. 6-3

many ac circuit calculations. It is often called the radianfrequency of the ac wave.

The importance of the sine wave stems from the fact that it is the simplest of all possible periodic waveforms.

COMPLEX WAVEFORMS, HARMONICS: All waveforms other than the sine wave are called complex waveforms. Any periodic complex waveform can be synthesized from a series of sine waves having different amplitudes and phases, but all having frequencies that are simple multiples of a lowest frequency called the fundamental frequency. That these multiples must be integers can be seen from the fact that, for the wave to repeat itself regularly, any components it may have must fit exactly into the same time interval as the period of the fundamental. In any other case the zero-axis crossings marking the beginning of the cycle of the complex wave would not be equally spaced in time. The time interval therefore has to be divided so that this will occur; i.e., into two, three, four, etc., sine-wave periods.

The frequencies above the fundamental are called harmonic frequencies, and are numbered (second, third, fifth harmonic, and so on) according to the multiple that applies. If desired, the complex wave can be broken down into its fundamental and harmonics by electrical circuits, and each extracted separately. An example of a complex waveform and its breadkown is shown in Fig. 6-3. At any given instant – several such instants are shown by the vertical dashed lines – the instantaneous amplitude of the complex wave is the algebraic sum of the amplitudes of its components at that instant, taking polarity into account.

A complex waveform that is *not* periodic cannot be synthesized or broken down by this method. Such waveforms are transients (the rise and decay in dc circuits having time constants are examples) and will in general contain a great many "instantaneous" frequencies (a rather ambiguous term), meaning that the circuit must be able to operate over a wide band of frequencies without undue discrimination against any. The frequency band required depends on the waveform and how closely it must be maintained in transmission through a circuit.

The relationship between the maximum or peak value and the rms value in a complex waveform varies with the waveform. The ratio peak/rms can have a wide range of values.

AC IN RESISTANCE: There is no energy storage associated with resistance (all the energy is dissipated) so power in a resistance is always "real." The current at every instant is exactly proportional to the voltage; the current and voltage are in phase. This is true of any waveform, so in a purely resistive circuit Ohm's Law applies to instantaneous values without modification. However, these values are continually varying, and if the average power over a cycle is wanted the rms values of voltage and current must be used; as stated earlier, these are not the same in complex waveforms as in sine waves.

It is often necessary to know the peak or maximum instantaneous power (P_m) in a communication circuit, together with the average power. In the case of the sine wave there is a simple relationship between the two. Since power at any instant is i^2R, where i is the instantaneous current, the peak power is i_m^2R, i_m being the peak or maximum current. Also, $i_m = \sqrt{2}I$, where I is the rms current. Then the ratio of peak to average power is

$$\frac{P_m}{P} = \frac{i_m^2 R}{I^2 R} = \frac{(\sqrt{2}I)^2 R}{I^2 R} = \frac{2I^2 R}{I^2 R} = 2$$

Thus the peak power in a sine wave is twice the average power. In general, however, this ratio will depend on the waveform.

LINEAR and NONLINEAR CIRCUITS: A circuit is linear when the current is exactly proportional to the applied voltage for all possible values of voltage, including both polarities. In such a

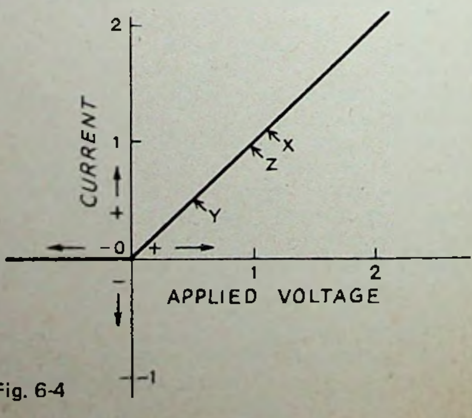


Fig. 6-4

circuit a plot of current vs. voltage results in a straight line. If there is anywhere a departure from this straight-line relationship the circuit is nonlinear in that region. Simple resistive circuits of the type previously considered are linear.

However, in communication work most circuits in which ac is combined with dc are not strictly linear. A principal cause of nonlinearity is the fact that amplifying devices will allow electrons to move through them in one direction only. Thus direct current of only one polarity can flow.

In Fig. 6-4 the heavy line is a plot of voltage vs. current in a hypothetical circuit. When the applied voltage is positive the current amplitude follows the voltage amplitude exactly. However, with a reversal of voltage polarity (voltage is increasingly negative to the left, and current of negative polarity increases downward from zero) the current becomes zero for *any* value of applied negative voltage. That is, the circuit will allow conduction only in one direction. Hence for a voltage *alternating* in polarity around the zero point the response of the circuit is extremely nonlinear. The current in such a circuit will not have the same waveshape as the ac voltage, since current can flow only during the positive half cycle.

On the other hand, if the ac axis is at a point such as *X* on the curve, the circuit will be linear for all values of ac voltage that do not "swing" the total instantaneous applied voltage (algebraic sum of positive dc and negative ac) into the negative region of the graph. There is thus a linear region of limited extent, confined to the (in this case) positive region of the graph. The resultant effect on the response of the circuit is shown in Fig. 6-5(A), where the negative peak of the ac wave is less than the positive value of applied dc voltage.

Fig. 6-5(B) shows a case where the negative swing of ac voltage *exceeds* the positive dc voltage. In this case the dc voltage is set at *Y*, Fig. 6-4, and when an ac voltage of the same amplitude as Fig. 6-5(A) is added, it swings the net instantaneous voltage negative – that is, into the region where the current is cut off from flowing. The fact that no current can flow during this cut-off interval would tend to decrease the average value of the direct current over the whole ac cycle. However, the current increases during the *entire* positive half of the ac cycle, and the increase in current more than counterbalances the decrease during the cut-off interval. The result is that the average value of direct current shifts upward, as shown by the line *Y'*. This shift in average value would be read by a dc meter.

The limiting case of linearity, where the sum of the positive dc and negative ac just touches zero, is shown in Fig. 6-5(C). Since the sine wave is

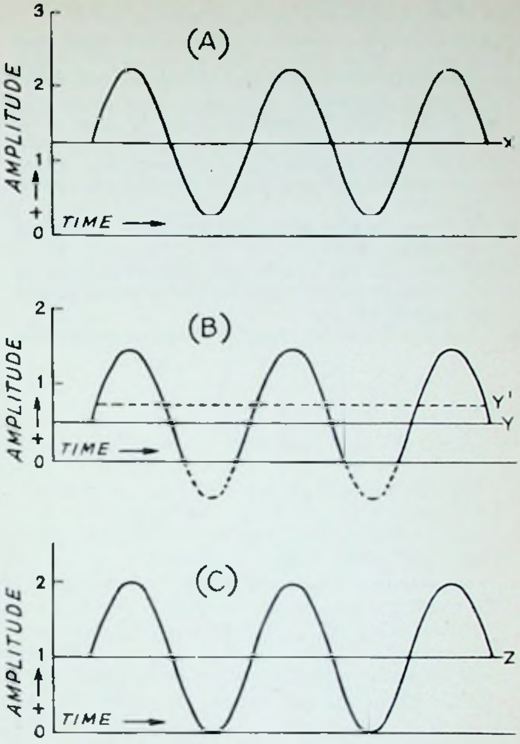


Fig. 6-5

symmetrical, the sum of the dc and the *positive* ac peak just reaches 2, on a scale such that the dc value is 1. *This is the largest sine wave that can be superimposed on the linear portion of a circuit of this type*, and represents an important limit of one method of operating an amplifying device. Any larger ac amplitude would result in cutting off part of the current, as in Fig. 6-5(B).

Several important facts are associated with this limiting case: 1) If the sine-wave amplitude exceeds the limit, there is a change in the average dc amplitude; below the limit there is not. This is a clue to detecting nonlinearity with a simple dc measurement. 2) Since the average power in a sine wave is equal to 1/2 the peak power (see earlier discussion), the ac power is equal to 1/2 the dc power (peak ac voltage swing is +1 and dc voltage is 1). 3) The sum of the ac and dc powers is therefore 1-1/2 times the dc power. 4) Since the peak combined ac and dc voltage is 2 units, the *peak* circuit power is 2^2 or 4 times the dc power. (On the negative peak it is of course zero.)

Questions and Problems

6-1) What is the average value of an ac waveform?

6-2) What is meant by a harmonic frequency? If the fundamental frequency is 850 kHz, what is the frequency of the 5th harmonic?

6-3) If a 3rd harmonic is 10.5 MHz, what is the fundamental?

6-4) If the rms value of a sine wave is 115 volts, what is the peak value?

6-5) How is rms related to peak-to-peak value in a complex waveform? In a sine wave?

6-6) How many degrees are there in a single cycle of a complex periodic waveform? In a sine wave?

6-7) How is angular measure related to time?

6-8) In a linear circuit having simultaneous ac and dc currents, what is the effect on the ac if the dc is doubled in value? Tripled? Reduced to zero? Reversed in direction?

6-9) Under what conditions can a sine wave superimposed on direct current have its amplitude changed without affecting the waveform, if the circuit has the voltage/current relationship shown in Fig. 6-4?

6-10) What is the difference between dissipated and reactive power?

6-11) What is the largest sine wave that can be fitted to the linear portion of a circuit when dc and ac are combined?

6-12) What is the relationship between period and frequency?

6-13) What is the relationship between peak and average power in a sine wave?

6-14) When ac and dc are simultaneously present in a circuit, what is the total power?

6-15) If ac is added to dc in a linear circuit what is the effect on the dc meter readings?

Questions concerning Exp. 5

(See Experiments Section):

6-16) In Exp. E-5, Procedure (B), why is there a change in the ac voltmeter reading when the direct current is reduced in value? Does an electronic voltmeter read rms in this case?

Chapter 7

Rate of change in sine waves; inductive reactance; capacitive reactance; nature of reactance; reactances of the same kind in series and parallel; unlike reactances in series; unlike reactances in parallel; more than two reactances; resonance; *L/C* ratio.

Reactance

CHANGE, and particularly the *rate* of change, is all-important in phenomena involving magnetic and electric fields. The instant a field or its relationship to the circuit undergoes a change there is an induced emf or change in current, or both. Conversely, a change in either voltage or current in a circuit causes a change in the field.

RATE OF CHANGE IN SINE WAVES: Since the sine wave is so basic to ac, its rate of change is of first importance. Fig. 7-1 shows the relationship between instantaneous amplitude (heavy) and the rate of change (light) of that amplitude. The rate-of-change curve has the same sine-wave shape as the amplitude curve, but is displaced one-quarter cycle or 90 deg to the left (one quarter cycle earlier in time).

Both the instantaneous amplitude and its rate of change are functions of the angle. In turn, the angular rate of change is proportional to the frequency. In fundamental units, the rate at which the angle changes is $2\pi f$, the total angle of one cycle, 2π, multiplied by the frequency, f. This is called the angular velocity.

INDUCTIVE REACTANCE: The instantaneous amplitude of the induced voltage is proportional to the rate of change of magnetic flux, and therefore is proportional to the rate of change of current. As explained in Chapter 4, the induced voltage opposes the change in current, and therefore opposes any voltage that may be *applied* to the inductance to cause current to flow. Mathematically, this is equivalent to giving the induced voltage a negative sign to indicate the opposition, and means, physically, that the polarity of the rate-of-change curve in Fig. 7-1 must be reversed. This leads to the representation in Fig. 7-2. The induced voltage is said to lag the current by 90 deg, and since it opposes – that is, is 180 deg out of phase with – the voltage applied to the inductance, the rate-of-change curve in Fig. 7-1 represents the *applied* voltage if the amplitude curve is taken to represent the current. The applied voltage is therefore 90 deg *ahead* of the current, and the current in the inductance is said to lag the applied voltage by 90 deg.

Because the rate of change of current increases directly with frequency, less current is required for a given induced voltage in an inductance of fixed value as the frequency is raised. Also, since the induced voltage is proportional to inductance for a given rate of change of current, less current is needed at a given frequency if the inductance is increased.

The effect of frequency and inductance on the ratio of voltage to current can be combined as follows:

$$\frac{E}{I} = 2\pi f L = X_L.$$

That is, the voltage across an inductance per unit of alternating current flowing through it is equal to the angular velocity multiplied by the inductance. If E and I are expressed in volts and amperes, respectively, and L is in henrys, the inductive reactance, X_L, is in ohms, the same unit that is used for resistance (note the similarity between the above expression and Ohm's Law for resistance).

Defined generally, an ohm is simply the ratio of one volt to one ampere. However, a reactive ohm differs from a resistive ohm in that in the former there is no energy dissipation; instead, the energy is stored in the field and eventually returned to the circuit without loss.

CAPACITIVE REACTANCE: As explained in Chapter 1, the charge on a capacitor is Q = CE; that is, the amount of charge is proportional to the

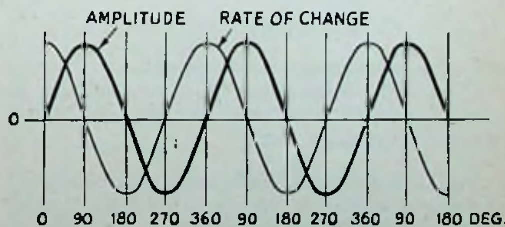


Fig. 7-1

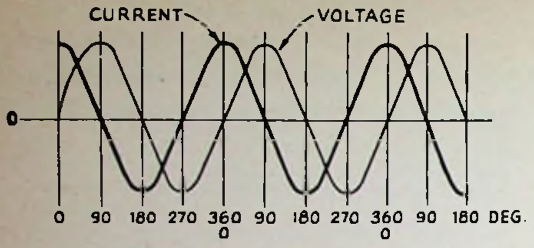


Fig. 7-2

capacitance and the voltage to which it is charged. Also, electric current is the rate at which charge is moved from one point to another in a circuit; i.e, $I = Q/t$. When an alternating voltage is applied to a capacitor the *rate* at which charge is placed on the capacitor (which is also the current flowing into the capacitor) is proportional to the *rate of change* of applied voltage.

If the amplitude curve (heavy) of Fig. 7-1 represents applied voltage, its rate of change is greatest when the voltage is passing through zero, and becomes zero at the instant the voltage reaches its maximum value. Thus the current flowing into or out of the capacitor is greatest when the voltage is passing through zero, and is zero when the voltage is at maximum. This is shown in Fig. 7-2. In the first quarter cycle the current has the same polarity as the voltage, because the capacitor is being charged at that time. In the next quarter cycle the capacitor discharges because the applied voltage is decreasing. The process then repeats with the *opposite* polarity, being back at the starting point at the end of one complete cycle. The current in a capacitor is said to lead the voltage by 90 degrees, since the current maximum is one-quarter cycle ahead of the voltage maximum, both being of the same polarity.

The rate of change of voltage in a sine wave increases directly with the frequency, so the current into the capacitor likewise increases directly with frequency for a given applied voltage. The current also increases directly with the capacitance. Thus the overall effect of increasing capacitance and frequency is to increase the current for a given applied voltage. Combining the two in terms of the ratio of current to voltage gives

$$\frac{I}{E} = 2\pi f C,$$

which inverted becomes

$$\frac{E}{I} = \frac{1}{2\pi f C} = X_C$$

where X_C is the capacitive reactance.

Capacitive reactance, like inductive reactance, is measured in ohms when fundamental units – volts, amperes, farads – are used.

NATURE OF REACTANCE: As stated above, reactance is associated entirely with the alternate storage and release of field energy, so no power is dissipated. In the first quarter cycle (0-90 degrees) in Figs. 7-1 and 7-2, the voltage and current have the same polarity and power is expended in putting energy *into* the field. In the next quarter cycle (90-180 degrees) the current and voltage have opposite polarities – "negative power" – and the same amount of energy is *returned* to the circuit. The process repeats each half cycle, but with alternating polarities.

Although there is no power dissipation, the effect of reactance on the *amplitude* of the current for a given applied voltage is similar to that of resistance. In a purely resistive circuit the current is

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

and in a purely reactive circuit the current is

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

for either inductive or capacitive reactance. Thus Ohm's Law applies to pure reactances.

REACTANCES OF THE SAME KIND IN SERIES and PARALLEL: With two restrictions, reactances in series and parallel combine by the same rules that govern resistances in series and parallel, respectively. That is, the total reactance of a number of reactances in series is the sum:

$$X = X_1 + X_2 + X_3 \ . \ . \ . \text{ etc.},$$

and the total reactance of a number in parallel is

$$X = \frac{1}{\frac{1}{X_1} + \frac{1}{X_2} + \frac{1}{X_3} \ . \ . \ . \ .} \text{ etc.}$$

The reasoning leading to these relationships is similar to that used with resistances in series and parallel, and the current and voltage relationships are similar.

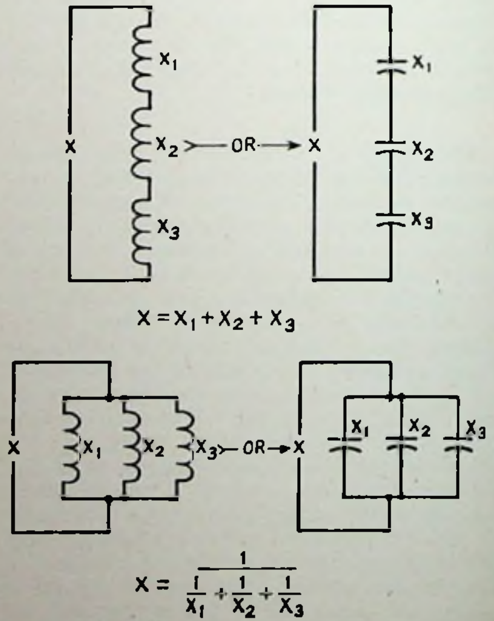


Fig. 7-3

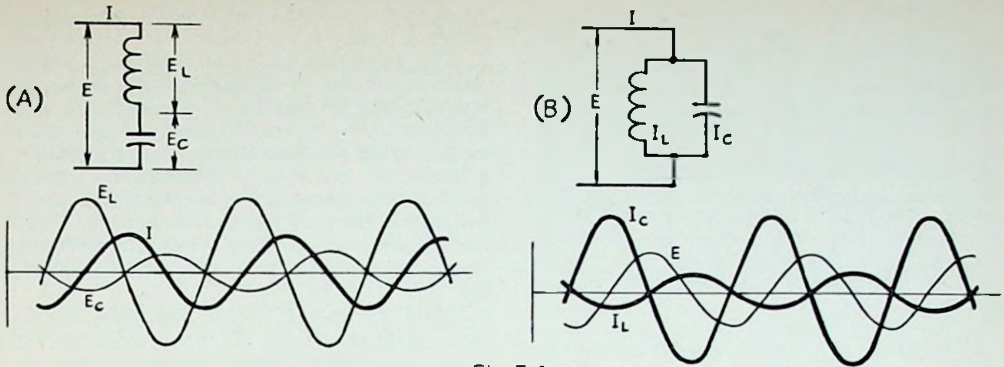


Fig. 7-4

The first restriction on the use of these formulas is that the fields of the reactances must not interact. If there are any induced voltages or currents in a series or parallel system, the relationships are no longer true.

The second restriction is that all reactances must be of the same kind – i.e., all inductive, or all capacitive, for the formulas to be used without modification. The permissible cases are shown in Fig. 7-3.

UNLIKE REACTANCES IN SERIES: In combining unlike reactances in series it is necessary to take into account that in an inductance the current lags 90 degrees behind the voltage (that is, the voltage leads the current by 90 degrees), and that in a capacitance the current leads the voltage (the voltage lags behind the current by 90 degrees). These relationships are shown in Fig. 7-4(A). Since the reactances are in series the current has to be the same in both, but the voltage E_L leads the current by 90 degrees and the voltage E_C lags the current by the same angle. E_L and E_C therefore always have opposite polarity, and the total voltage, E (not shown in the wave drawing), across the two is their *difference.*

Conventionally, the inductive reactance is called positive and the capacitive reactance negative, and the total voltage is the *algebraic* sum; i.e., $+ X_L - X_C$. In the case shown, the inductive reactance is larger (E_L is larger than E_C) and the sum is positive, indicating that the total reactance is inductive. If the converse had been the case the total reactance would have been negative and therefore capacitive. Because of the algebraic addition, the total voltage is always *smaller* than the voltage across the larger of two reactances in series. Since the total reactance is E/I, it, too, is smaller than the larger of the two reactances.

Except for the signs, the generalized rule for reactances in series is the same as the one for reactances that are all alike. That is, the total reactance of a number of reactances in series is the *algebraic* sum of the individual reactances.

UNLIKE REACTANCES IN PARALLEL: Fig. 7-4(B) shows the parallel case. As the voltage is common to both reactances, it is used as the reference. I_C, the current through the capacitance, leads the voltage by 90 degrees and I_L, the current through the inductance, lags by the same angle. The two *currents* are 180 degrees out of phase in this case, so the total current is their difference. Thus the total current is

$$I = I_L - I_C$$

and the total reactance is

$$X = \frac{E}{I} = \frac{E}{I_L - I_C}$$

In the case shown in the drawing, I_C is larger than I_L, so the resultant reactance is capacitive. The total reactance is *larger* than the larger of the two reactances, because the total current is smaller than the larger of the two individual currents.

The formula above is in terms of voltage and currents rather than individual reactances. An equation based on reactances only can be developed algebraically. For two unlike reactances the result is

$$X = \frac{X_C X_L}{X_C - X_L}$$

In the equation as given, both reactances are absolute values, i.e., signs are disregarded. If the solution has a negative sign (X_L larger than X_C) the resultant reactance is capacitive; if the solution

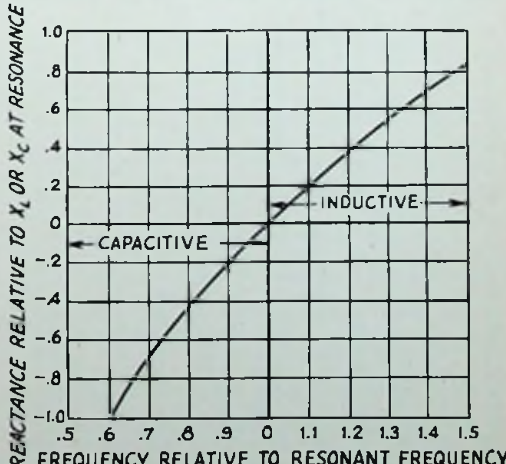


Fig. 7-5

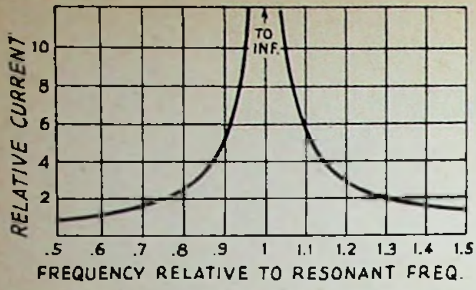


Fig. 7-6

has a positive sign (X_C larger than X_L) the resultant is inductive.

MORE THAN TWO REACTANCES: If there are more than two unlike reactances in series or parallel, their combined effect can be reduced to that of *two* unlike reactances, since there can be only two kinds of reactance.

In the series case, simply add all the inductive reactances and consider them as a single reactance of that value. The capacitive reactances should be similarly combined, after which the formula above for two unlike reactances in parallel can be used.

RESONANCE: A special case occurs when two unlike reactances both have the same numerical value. If they are in series the two voltages E_L and E_C are equal and their sum is zero. In this case the only limit on current flow is the resistance in the circuit. There is always some resistance in the reactances themselves, as well as in other circuit elements. However, if the internal resistances are very small compared to the reactances the voltages across the individual reactances can become quite large, even though the total voltage across them is very low.

When equal reactances are in parallel, the current is theoretically zero no matter what the applied voltage may be. Again this is true only if the reactances are pure – i.e., have no internal resistance. In practice, the inevitable resistance

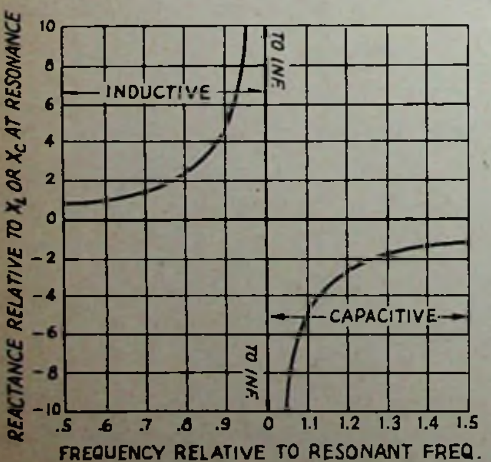


Fig. 7-7

prevents the current from actually becoming zero, although it may be very small.

The condition of equal reactances, either in series or parallel, is called resonance. With a given inductance and capacitance, it occurs at only one frequency, the resonant frequency. If the frequency is changed the reactances change with it, and as a result, the two voltages differ in the series case and the two currents differ in the parallel case. The behavior of the total reactance is shown in Fig. 7-5, for the series circuit. With fixed voltage of variable frequency applied, the current varies as shown in Fig. 7-6.

With the parallel circuit the total reactance varies as shown in Fig. 7-7 when the frequency is varied. The relative current flowing into and out of the parallel circuit (generally called the line current) varies as shown in Fig. 7-8, for constant voltage of variable frequency.

The curves in Figs. 7-5 to 7-8, inclusive, are not symmetrical about the resonant frequency, although for very *small* departures from resonance they are essentially so. Also, the current curve in Fig. 7-6 does not actually reach infinity at resonance; it could do so only in the theoretical case of reactances and circuits having no resistance. The reactance curves in Fig. 7-7 likewise do not reach infinity except theoretically; however, there is always a shift from inductive to capacitive reactance (or vice versa) right at the resonant frequency, when the applied frequency is varied.

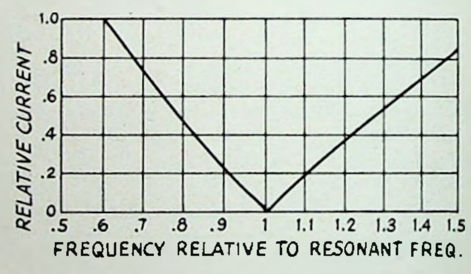


Fig. 7-8

L/C RATIO: Although there is only one resonant frequency associated with a given inductance and capacitance, a given resonant frequency can be obtained with, theoretically, an infinite number of such pairs. All that is necessary is that the reactances should be equal; this condition can be satisfied with a small inductance and large capacitance, large inductance and small capacitance, and values in between.

In practice, the range of usable values is limited because of stray effects. A practical inductor will also have "distributed" capacitance because of the voltage between turns (there is further discussion of this in Chapter 11) and a practical capacitor will have internal inductance because conductors have finite length. Within the usable range of values for a given frequency, the L/C ratio expresses in a general way the relative values of inductance and capacitance used for making a resonant circuit for a particular application. There is no exact definition associated with the term, as a large L/C ratio in one application may be a small ratio in another.

Questions and Problems

7-1) What is the formula for computing power dissipation in a reactance?

7-2) What is the reactance of a 15-henry inductor at 120 Hz?

7-3) What is the reactance of a 0.5-μF capacitor at 5000 Hz? At 100 Hz?

7-4) If two 1-μF capacitors are in parallel what is their combined reactance at 1000 Hz? If the capacitors are in series what is their combined reactance at the same frequency?

7-5) What is the total reactance of a 100-pF capacitor and 15-μH inductor in series at a frequency of 2000 kHz? At 6 MHz? At 11.5 MHz?

7-6) At what frequency is the LC combination in Q. 7-5 resonant?

7-7) If the inductor and capacitor of Q. 7-5 are connected in parallel instead of in series, what are the total reactances of the parallel circuit at the same frequencies?

7-8) A 4-μH inductor, 150-pF capacitor, 25-μH inductor, and 20-pF capacitor are connected in series. What is the total reactance of the circuit at 2650 kHz?

7-9) If the four components of Q. 7-8 are connected in parallel, what is the circuit reactance at 14 MHz?

7-10) What is the resonant frequency of the combination in Q. 7-8?

Impedance; admittance; more than two circuit elements in series; more than two circuit elements in parallel; series-parallel combinations; apparent power and power factor.

Impedance in Series and Parallel Circuits

WHEN A CIRCUIT contains both resistance and reactance the opposition to current flow is called impedance. It is a more general term than resistance or reactance, and is frequently used for either even when pure resistance or reactance actually is meant.

IMPEDANCE: Impedance is defined as the ratio of voltage to current, i.e.;

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

where Z is the impedance in ohms, E is the emf in volts, and I is the current in amperes.

In a circuit consisting of reactance and resistance in series such as is shown in Fig. 8-1, where an inductive reactance X is in series with a resistance R, the same current, I, flows through

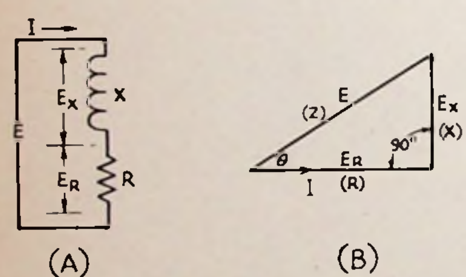


Fig. 8-1

both. However, the voltage E_X is 90 deg out of phase with the current, while in R the current and voltage E_R are in phase; thus E_X is 90 deg out of phase with E_R. If the amplitudes of these voltages are represented by straight lines drawn at right angles or quadrature as shown at (B), the combined voltage E is given by the length of the line joining their tips. The phase angle θ is something less than 90 deg, the exact value depending on the ratio of reactance to resistance.

From the well-known relationship between the lengths of the sides of a right triangle,

$$E^2 = E_R{}^2 + E_X{}^2$$

and dividing both sides by the current I leads to

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

Thus the sides of the triangle also represent R, X, and Z, as indicated parenthetically in Fig. 8-1(B). This is the impedance triangle of a series circuit.

By trigonometry, the phase angle θ can be found from

$$\tan\theta = \frac{X}{R}.$$

The same equations can be used for either capacitive or inductive reactances. It is necessary to remember, however, that with capacitive reactance the voltage lags behind the current, as contrasted with the inductive case.

ADMITTANCE: If a circuit has resistance and reactance in parallel, the voltage is common to both elements and is used as the reference in determining the impedance and phase angle. In this case it is convenient to find the circuit values in terms of admittance, which is the reciprocal of impedance. That is,

$$Y = \frac{1}{Z} = \frac{I}{E}$$

where Y is the admittance of the circuit. As shown in Fig. 8-2(B), the currents are combined to find the total current in the same way that voltages were combined in Fig. 8-1(B). Note that because the current through the inductance lags the voltage across it, I_X is drawn *downward*, since the voltage is the reference. In a comparable capacitive circuit the current would be drawn *upward*.

The equation for resultant current is

$$I = \sqrt{I^2{}_R + I^2{}_X}.$$

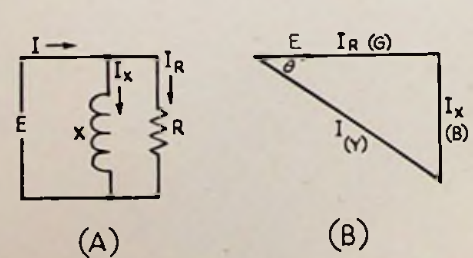


Fig. 8-2

If both sides of this equation are divided by E, the result is

$$Y = \sqrt{G^2 + B^2}$$

where G is the conductance or reciprocal of the resistance (see Chapter 3) and B is the reciprocal of the reactance and is called the susceptance. That is,

$$G = \frac{1}{R},$$

and

$$B = \frac{1}{X}.$$

All three quantities, Y, B, and G, are measured in mhos, or reciprocal ohms.

Once the admittance is found it can be converted into impedance by inversion:

$$Z = \frac{1}{Y}.$$

The phase angle of the parallel circuit, by analogy to the series circuit, is found from

$$\tan\theta = \frac{B}{G}.$$

Note that mathematically the change of reference from voltage to current in Figs. 8-1 and 8-2 leads to a reversal of the signs, so negative (capacitive) reactance becomes positive susceptance, and positive (inductive) reactance becomes negative susceptance. However, this distinction need cause no difficulty, because both quantities are squared in calculating the resultant admittance or impedance, and this makes them both positive. The important thing to remember, in using the equations, is that inductive and capacitive reactances have *opposite* signs, and so do inductive and capacitive susceptances.

MORE THAN TWO CIRCUIT ELEMENTS IN SERIES: If several resistances and several reactances are in series, the resistances can be added together to obtain a sum which is used as R in the diagrams and equations just discussed. Similarly, as described in Chapter 7, the reactances can be combined, if more than one is present, to obtain a total reactance which can then be used as X in the diagrams and equations. Reactances must of course be added algebraically, according to their signs.

This is illustrated in Fig. 8-3, where unlike reactances are shown in series with resistance. There could be more than one of each kind of reactance in the series string, in which case all the inductive reactances would first be added together to find the total inductive reactance, X_L, and all the capacitive reactances would be added together to find the total capacitive reactance, X_C. If there is more than one resistance in the circuit, including the series resistances of the reactive components, these too should be totalled, after which the procedure is the same as for Fig. 8-1.

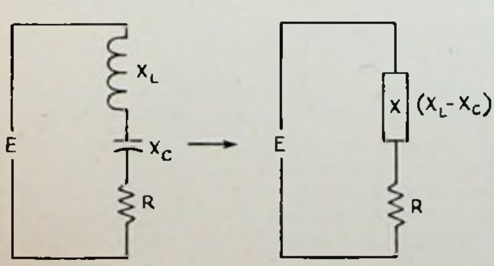


Fig. 8-3

A plain block is used to represent $X_L - X_C$, since the resultant reactance could be either inductive or capacitive.

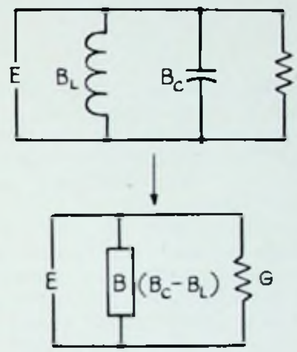


Fig. 8-4

MORE THAN TWO CIRCUIT ELEMENTS IN PARALLEL: In the parallel case, also, there may be several reactances (including both types) in parallel with one or more resistances. The procedure again is to combine all the *like* elements first. Find the net resistance of all the resistances in parallel, the net inductive reactance of all the inductive reactances in parallel, and the net capacitive reactance of all the capacitive reactances in parallel. Then convert the net reactances into susceptances, and the net resistance into conductance. This reduces the circuit to the simple form shown in Fig. 8-4. The net susceptance, B, is the difference between the capacitive and inductive susceptances, as in the diagram. The procedure for finding the total admittance (and from it the impedance) is then the same as for Fig. 8-2.

SERIES-PARALLEL COMBINATIONS: The process of reducing a complicated arrangement of circuit components to the simplest possible one in a step-by-step procedure is quite universal, but certain precautions must be observed in applying it. In particular, mixtures of series and parallel connections – such as might be formed by connecting Figs. 8-1 and 8-2 in series, or by connecting them in parallel, are too complex to be handled by the methods just described.

This is because the phase angles of the separate series and parallel subcircuits will, in general, be different, so their voltages and currents will not combine in the same way as they do in the circuits just considered. (That is, the phases will be neither 0, 90, nor 180 degrees, which are the only phase relationships entering into the solutions of the separate series and parallel circuits.) Solving such circuits requires using equivalent circuits, which convert resistance-reactance combinations from

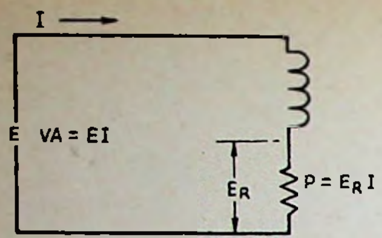


Fig. 8-5

series to parallel, and vice versa, so that the resultant circuit is either entirely series or entirely parallel. This subject is taken up in Chapter 9.

Many practical circuits have the *appearance* of simplicity, but on a closer look turn out to be rather complex. It must not be forgotten that reactances are seldom pure, but have losses that represent hidden resistance. Thus a capacitor may have leakage which is equivalent to a resistance in parallel with the capacitance, and while a circuit containint it may appear to be a simple series arrangement, it is actually a series-parallel circuit. The losses in an inductor are equivalent to a resistance in series with the inductance, so such an inductor in an apparently simple parallel circuit actually makes the circuit a series-parallel arrangement.

If such losses are very small compared with the energy in the field, these resistances can be ignored and the calculations will still be within the accuracy needed for most practical work. However, when losses are appreciable this is no longer the case. The methods of Chapter 9 have to be used in such cases if calculation is to agree reasonably well with measurement.

APPARENT POWER AND POWER FACTOR: When energy is supplied to a circuit containing both resistance and reactance, part is dissipated in the resistance and part is stored in the field of the reactance. As stated earlier, this field energy is handed back and forth between the circuit and the source. The rate at which the energy is supplied to the field is termed the reactive power. It is measured in volt-amperes, which are simply the result of multiplying the voltage by the current. Energy consumed in the resistance, however, is "real" power, and is measured in watts.

The volt-ampere is also used as a measure of the apparent power in a resistive/reactive circuit. The ratio of the resistive watts to the total volt-amperes (apparent power) is called the power factor of the circuit. Fig. 8-5 is a simple illustration using a series circuit. Since the power in the resistance is $E_R I$ while the total volt-ampere product is EI, the power factor is

$$\frac{E_R I}{EI} = \frac{E_R}{E}.$$

From trigonometry and Fig. 8-1(B),

$$\frac{E_R}{E} = \cos\theta.$$

Thus the power factor is equal to the cosine of the phase angle. This is true for a circuit of any type containing both resistance and reactance. The power actually dissipated in such a circuit is

$$P = EI\cos\theta.$$

Questions and Problems

8-1) How are impedance and admittance related? Susceptance and reactance?

8-2) A resistance of 400 ohms is in series with a reactance of 600 ohms. What is the impedance? What is the power factor of this circuit? Is the voltage leading or lagging the current?

8-3) A capacitive reactance of 1000 ohms is in series with an inductive reactance of 500 ohms and a resistance of 500 ohms. Does the current lead or lag the voltage? What is the impedance?

8-4) A 0.25-μF capacitor is in parallel with a 4700-ohm resistor. What is the impedance of the circuit at 250 Hz? At 1 kHz? At 40 Hz?

8-5) Compare volt-amperes and watts.

8-6) If the voltage applied to a circuit is doubled, what is the effect on the impedance? On the phase angle?

8-7) How much power is dissipated in a circuit consisting of a 1-henry inductance in parallel with a 200-ohm resistor if 50 volts at 60 Hz is applied? If 50 volts at 500 Hz is applied? What is the line current at both frequencies?

8-8) How much power is dissipated in a circuit consisting of a 100-ohm resistor in series with a 1-μF capacitor at 60 Hz if 100 volts is applied? If the same voltage at 1200 Hz is applied?

8-9) a 1-henry inductance, a 25-μF capacitance, and a 2-μF capacitance are in series with a 500-ohm resistor. What is the impedance of the circuit at 100 Hz?

8-10) What is the admittance of the circuit in Q. 8-7?

Question concerning Exp. 6
(See Experiments Section)

8-11) Perform the computations necessary for arriving at the values of impedances listed in Exp. 6.

Chapter 9

Phasors; uses of phasors; reactance and resistance in series; reactance and resistance in parallel; impedance and admittance diagrams; complex circuits; parallel and series equivalents; practical circuit components.

Applications of Phasors

WAVE DRAWINGS such as those shown in Chapter 7 (Fig. 7-4) do not lend themselves to easy interpretation or ready numerical solution. In solving circuit problems, quantities that can be indicated by meters – maximum or rms values, especially the latter – are of more interest than instantaneous values. This need can be met by phasor diagrams – without, however, sacrificing the availability of instantaneous values if it is necessary to know them.

PHASORS: The phasor method of representing a sine wave is shown in Fig. 9-1. The line *OA* is assumed to rotate around *O* the way the spoke of a wheel rotates around the hub. Rotation is counterclockwise at constant speed, and each complete rotation represents one cycle. The beginning of the cycle is taken as the position of *OA* when it passes through *OB* (labeled 0 or 360 deg on the diagram).

The length of the rotating line *OA* represents the maximum or peak value of a sine wave. Then since the position of the line at any instant represents the number of degrees that have elapsed since the beginning of the cycle as measured counterclockwise from *B*, the length of the line *AC* dropped perpendicularly from *A* to *OB* is proportional to the instantaneous value of the wave at that instant. This is because the length of *AC* is equal to *OA* multiplied by the sine of the angle θ. In other words, the length of *AC* follows the sine curve shown in Fig. 7-1, Chapter 7. The values are positive when *AC* is above the line *DOB* and negative when *AC* is below *DOB*. Thus when *OA* passes through *DO* or *OB*, at which instant *AC* is zero, the polarity reverses just as it does at the start of each half cycle in Fig. 7-1.

USES OF PHASORS: The usefulness of this method of representation lies in the fact that two or more voltages or currents, or both, of the same frequency can be shown in their proper phase and amplitude relationships by phasors having a common center of rotation and so positioned that the angle between them is their phase difference. Since this angle is constant for a given circuit condition, the *relative* positions of the phasors do not change during rotation. This being the case, phasors can be combined graphically to find their sum, with due regard for phase. This sum is also a phasor, rotating at the same speed. Further, since the ratio of maximum to rms amplitude is constant in a sine wave, the lengths of the phasors can be scaled to either rms or peak values, as convenient.

If instantaneous values are not of primary concern, the rotation can be ignored and the phasors considered solely on the basis of their relationship to one another. They can be placed in any convenient fixed position in this case. Customarily, one phasor is drawn horizontally and used as a reference for determining the relative positions of the others.

Figs. 9-2 and 9-3 show the application of phasors to series and parallel combinations of

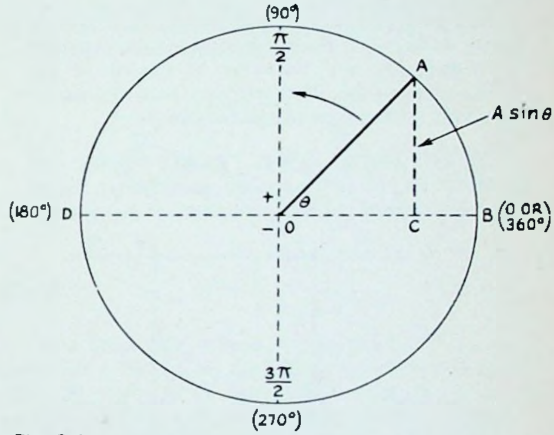


Fig. 9-1

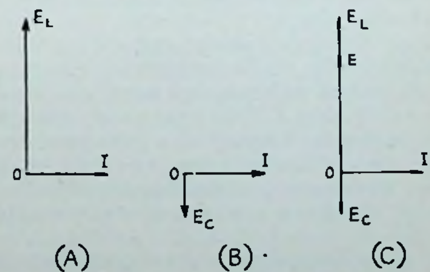


Fig. 9-2

unlike reactances. Fig. 9-2 is the phasor representation corresponding to Fig. 7-4(A), Chapter 7. Any arbitrary scale can be used; the arrowhead (not always used) marks the length of the phasor by identifying the end. The current is placed in the reference position since it is the same in all circuit elements in series. The voltage E_L across the inductive reactance is drawn 90 degrees ahead of I in the counterclockwise direction, Fig. 9-2(A). The voltage E_C across the capacitive reactance is 90 degrees *behind* the current, Fig. 9-2(B). Finally, the two voltages and current are combined in (C). The voltage phasors are seen to point in opposite directions, indicating that their polarities are opposite. Their difference, the *total* voltage, is the length E.

The parallel case corresponding to Fig. 7-4(B) is shown in Fig. 9-3. Here the voltage is common to both elements and therefore is drawn in the reference position. The two currents are drawn separately with the proper phase angles.

REACTANCE AND RESISTANCE IN SERIES: Phasor diagrams are particularly useful when both reactance and resistance are present in a

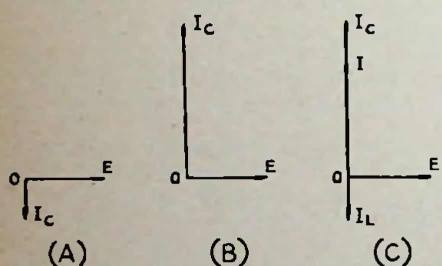


Fig. 9-3

circuit. Reactance and resistance in series are shown in Fig. 9-4. Since the current is the same throughout such a circuit it is placed in the reference position. In the resistance the voltage is in phase with the current, so the direction in which E_R is drawn coincides with that of the current, I. In (A) E_L is 90 deg ahead of the current and so is drawn at right angles upward. The sum of the two voltages is found by vector addition, which takes into account both the *direction* and *magnitude* of the phasors. The actual process involves drawing a parallelogram of which E_R and E_L form adjacent sides. The diagonal of the parallelogram is the sum, E, the total voltage.

A short-cut method of vector addition is shown at the right. If E_L is drawn from the tip of E_R rather than from the origin *O*, a line joining *O* and the tip of E_L gives E. (The hypotenuse of a right triangle is also the diagonal of a rectangle having adjacent sides of the same length as those of the triangle.) The choice between these two methods of vector addition is a matter of which is more convenient.

The angle θ between E and I is the phase angle of the circuit.

With capacitive reactance, (B), the voltage lags the current, as shown. In other respects the explanation is similar.

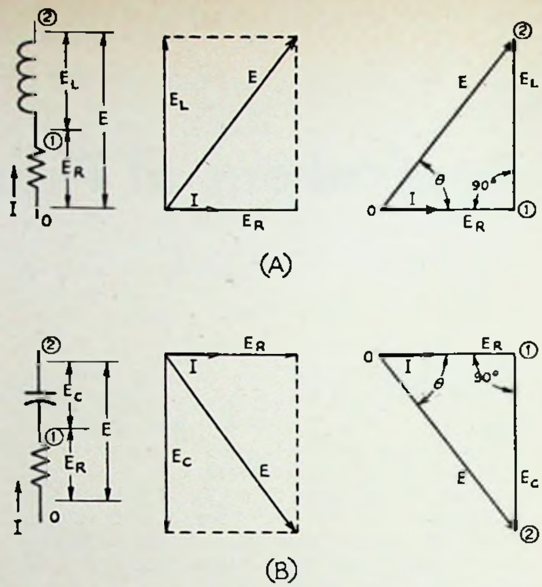


Fig. 9-4

REACTANCE AND RESISTANCE IN PARALLEL: The principles of constructing phasor diagrams are the same when reactance and resistance are in parallel. Fig. 9-5 is an example. Since the voltage is the common feature of parallel circuits, it is used as the reference.

Just as in Fig. 9-4, the directions of the current and voltage phasors for the resistances coincide. In (A), the current through the inductive reactance lags 90 degrees behind the voltage, so the sum, I, of the currents likewise is behind the voltage by some angle θ. In (B), the capacitive current leads the voltage and the total current also leads.

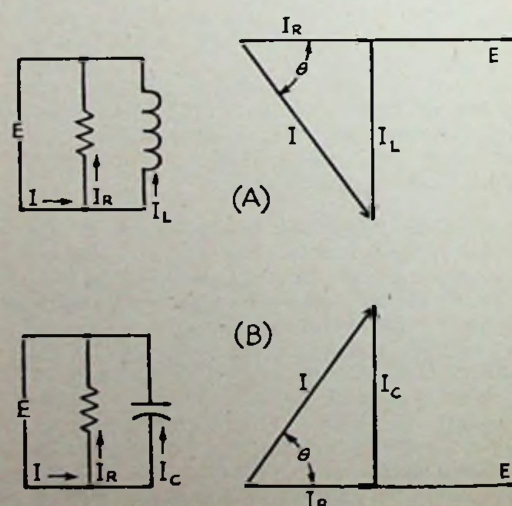


Fig. 9-5

Note that the presence of resistance causes a shift in the phase angle between current and voltage to some value intermediate between 0 deg (for pure resistance) and 90 deg (for pure reactance). This is true of either the series or parallel connection. Also, the actual phase angle depends on the *relative* values of resistive and reactive voltages (series) or resistive and reactive currents (parallel).

IMPEDANCE AND ADMITTANCE DIAGRAMS: The resemblance between phasor diagrams and the triangular relationship between resistance and reactance discussed in Chapter 8 (Fig. 8-1) is obvious. The triangle is in fact based on the phasor principle. By the same algebraic methods the phasor diagrams can be converted into impedance and admittance diagrams.

COMPLEX CIRCUITS: It must be understood that in such diagrams the resistance and reactance must be *pure*, and so must conductance and susceptance, if these quantities are to be drawn at right angles to each other. Fig. 9-6(A) is a circuit which does not meet this condition. Although this is basically a parallel circuit, X_L and X_C cannot be converted into susceptance by the formula $B = 1/X$ because the formula assumes that the reactance X is entirely in parallel with the rest of the circuit, while here there is resistance and reactance in series in both parallel arms of the circuit. Similarly, the conductances of R_1 and R_2 cannot be found from $G = 1/R$. Also, it is not sufficient to calculate the impedances of X_LR_1 and X_CR_2 and assume that the usual parallel formulas will give a correct result. The phase angles of the two branches may be quite different and cannot be ignored.

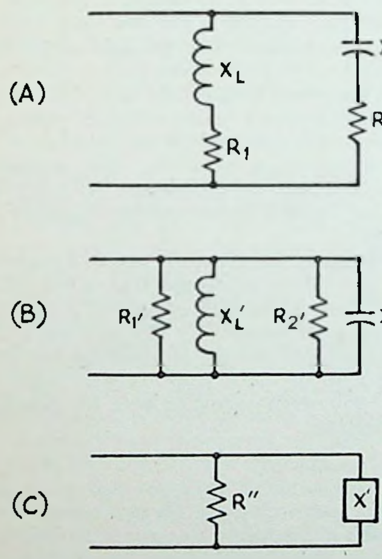


Fig. 9-6

Instead, a parallel circuit equivalent to X_LR_1 must be found; and a similar parallel circuit equivalent to X_CR_2 must likewise be determined. On substituting these equivalents, the circuit will have the form shown in Fig. 9-6(B). The separate circuit elements are now all in parallel, and may be combined as described in Chapter 8 to arrive at the overall circuit of Fig. 9-6(C). The circuit is now composed only of "pure" elements and may be solved as described in earlier chapters.

In Fig. 9-6(C) X′ is shown in block form because it could be either inductive or capacitive, depending on the relative values of X'_L and X'_C found in Fig. 9-6(B).

PARALLEL AND SERIES EQUIVALENTS: Every parallel circuit has a series-connected equivalent circuit and vice versa. The individual element values differ in the two equivalent circuits, but the overall impedance is the same.

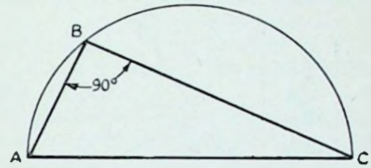


Fig. 9-7

As a simple example, a resistance of 100 ohms could be made up of, say, either 40 ohms and 60 ohms in series or of two 200-ohm resistances in parallel. So far as applied voltage and current go, the two combinations behave exactly alike. Similarly, if the impedance and phase angle of a resistance and reactance in series are duplicated by a resistance and reactance in parallel, or vice versa, the two combinations are equivalent.

Because the phase angle as well as the absolute impedance must be the same in both circuits, a unique set of values is required in forming the equivalent of an existing combination. The problem of finding exact values is simplified by making use of a geometrical principle shown in Fig. 9-7. If a semicircle is constructed on line *AC*, any pair of straight lines *AB* and *CB* drawn from the ends of AC and intersecting at the semicircle form a right-angled triangle *ABC*.

Fig. 9-8 shows how this is applied in converting graphically from a series circuit to its equivalent parallel circuit. As in the earlier example, the current I is the reference in the series circuit (A) and the resistive and reactive voltages are E_R and E_L, respectively. The phase angle is θ. The applied voltage, E, is found by combining E_R and E_L at right angles.

Insofar as the power source is concerned, this circuit is simply one which takes a current I at a particular phase angle θ when the applied voltage is E. From this viewpoint the circuit seen by the source is represented by the diagram at (B).

In (C) the applied voltage is used as the reference, here drawn in the same position as before in order to emphasize the equivalence. The current I is drawn at the angle θ and to the same scale as before. A semicircle drawn on I intersects

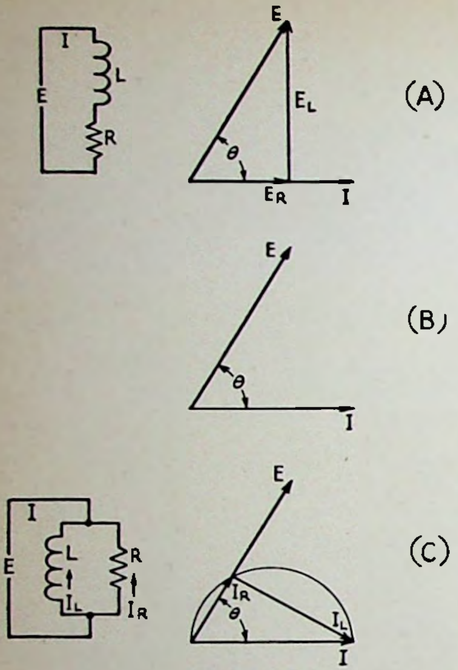


Fig. 9-8

E as shown. This intersection gives I_R, the equivalent parallel-resistance component of I, since I_R must be in phase with E (see Fig. 9-5B). Then a line from this point of intersection to the tip of I gives the reactive current, I_L, required to be combined with I_R to make up the total current, I.

In constructing an equivalent circuit the impedance and phase angle must be known. If not given, they can be derived from the known values of resistance and reactance, or conductance and susceptance, in the circuit by using the methods already detailed. Then, knowing the impedance and phase angle, an arbitrary voltage, E, can be assumed (1 volt is often a convenient value) and the current I found from E/Z. The diagram is next constructed to scale with the proper phase angle between voltage and current, and the procedure described above followed to get the equivalents for (in the series to parallel conversion) I_R and I_X. With the resistive and reactive currents known, they may be divided into E to find the resistance and reactance of the equivalent circuit.

For converting from parallel to series the same method is used, but with the current as the reference. The voltage is then broken down into the resistive and reactive components required for finding the resistance and reactance of the equivalent series circuit.

PRACTICAL CIRCUIT COMPONENTS: To emphasize again what has been said in earlier chapters, there is always some power dissipation in a practical reactive component. The effect of dissipation is equivalent to adding a resistance in series with the reactance. The resistance may be actual – an example is the resistance of the wire used in making an inductance coil – or it may represent the resistive equivalent of a power loss in materials that respond to the magnetic or electric field, such as the iron core in an inductor or the solid dielectric in a capacitor. The equivalent resistance of a reactive component is usually not constant but depends on the applied frequency and other factors.

In addition to the dielectric loss in a capacitor there is also often a leakage loss – an actual leakage of current through or around a capacitor because of imperfections in the insulation. The effect is equivalent to that of a leakage resistance in parallel with the capacitance. Leakage resistance is usually quite high compared with the reactance of the capacitor.

Questions and Problems

9-1) What is meant by a phasor?

9-2) When is it necessary to resort to equivalent series or parallel circuits in solving circuit problems?

9-3) How does vector addition differ from ordinary algebraic addition?

9-4) Can phasors of different frequencies be used in the same diagram?

9-5) Construct an impedance triangle for a capacitance of 5 μF, an inductance of 50 mH, and a resistance of 250 ohms, all in series, for a frequency of 800 Hz. What is the impedance of the circuit at this frequency? What is its admittance? What is the phase angle? (The angle can be measured by a protractor.)

9-6) Find the equivalent parallel circuit for the series CLR circuit of Q. 9-5.

9-7) Find the equivalent series circuit for a parallel circuit consisting of an inductance of 1.3 microhenrys, a capacitance of 500 picofarads, and a resistance of 600 ohms, operating at a frequency of 7000 kHz.

9-8) How can the leakage of a capacitor be represented in a circuit?

Chapter 10

Networks; source impedance; load impedance; impedance matching; matching by connecting network design; effect of impedance mismatch; the decibel; voltage and current ratios in dB; efficiency; impedance matching and efficiency; source dissipation; alternate definition of impedance matching.

Impedance Matching

A SYSTEM FOR generating power and making use of it has three essential parts – the source, the load (device that consumes the power), and the connecting network between them. This is shown in Fig. 10-1.

Fig. 10-1

NETWORKS: The connecting network may be very simple. In a flashlight, for example, it is just a short metallic path from the battery to the lamp and back, with a switch for turning the power on and off. Most networks will be much more complex, and individual parts of complex systems will be networks in themselves.

The ratio of voltage to current – impedance – becomes important in the transfer of power from a source to a load. At power-line frequencies, where voltages in a given circuit are more or less constant, it is customary (and convenient) to think in terms of voltage and current, the impedances being implied rather than specified. In communication circuits, however, the signal voltages may vary over a wide range in the course of normal operation, and it is more fruitful to think directly in terms of impedances.

SOURCE IMPEDANCE: Every power source has an internal impedance which must be considered in the transfer of power. Depending on the type of source, the internal impedance may be a simple resistance (a battery is an example of this type) or may consist of reactance along with resistance, as is typical of many ac power sources.

Current flowing through the resistance causes power dissipation inside the source. This is frequently an undesirable condition because power so dissipated not only is not available for the load but also causes heating in the power source. Such heating and the consequent rise in temperature ultimately limits the amount of power that most sources can deliver safely.

Although current flowing through a reactance does not directly cause a power loss, since reactance does not consume power, there can be other undesirable effects. One of these is a drop in output voltage because of the voltage drop in the internal reactance.

Very generally, a power source may be represented as shown inside the dashed box in Fig. 10-2. It is equivalent to a constant voltage E, an internal resistance R_1, and an internal reactance X_1, all in series. Only the terminals are accessible.

LOAD IMPEDANCE: The impedance representation of a load, also shown in Fig. 10-2, is quite similar, but since the load does not supply power there is no voltage source. X_2 is the internal reactance and R_2 is the load's resistance. In some cases R_2 may be an actual resistance (at power frequencies resistances are used for heating electric irons, toasters, and so on) but it is more likely to be the resistance *equivalent* ($R = P/I^2$ or $R = E^2/P$) of some other form of power consumption such as production of sound by a loudspeaker or radiation by an antenna.

IMPEDANCE MATCHING: The connecting network, also indicated in block form in Fig. 10-2, usually will have two functions to perform: carrying the power from the source to the load, and reconciling the differences between source and load impedances so that the desired amount of power can be transferred.

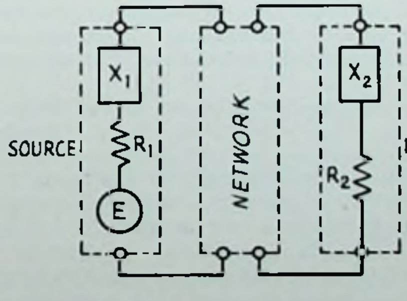


Fig. 10-2

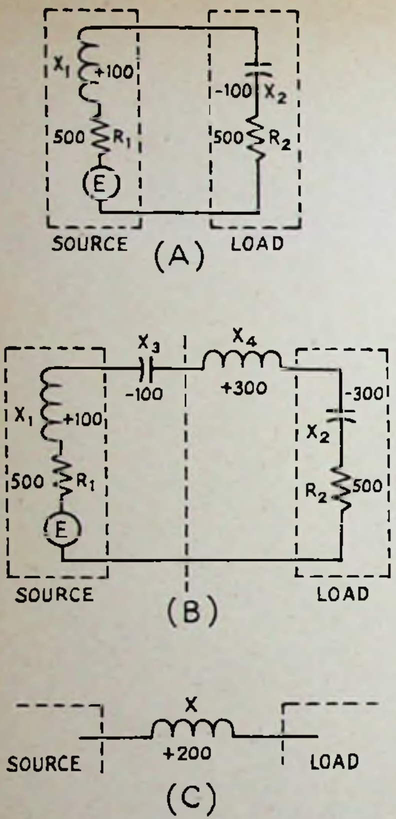


Fig. 10-3

It can be shown mathematically that the maximum power that it is possible to draw from a source (known as available power) will be obtained when the load impedance matches the source impedance. If the load impedance and the internal impedance of the source are both purely resistive, maximum power is secured when these two resistances are equal. If the source impedance is complex, (i.e., contains reactance along with resistance) the load impedance must be the conjugate of the source impedance. That is, the resistive components of both must be equal, and the reactances also must be equal but of *opposite* kinds. Since opposite reactances have opposite signs, this cancels the internal reactances of both source and load, leaving only the resistive components operative.

Conjugate matching is shown in Fig. 10-3(A), where $R_1 = R_2 = 500$ ohms, X_1 is a positive (inductive) reactance of 100 ohms, and X_2 is a negative (capacitive) reactance of 100 ohms. The reactances cancel, leaving just the resistances effective, and the impedances are matched.

MATCHING BY CONNECTING-NETWORK DESIGN: Cases of the type in Fig. 10-3(A) are rare, and it is usually necessary to make the connecting network reconcile at least the reactances. Fig. 10-3(B) illustrates. The load impedance does not match the source impedance directly. However, adding X_3 in the network will cancel the source reactance, and adding X_4 will cancel the load reactance, again resulting in a match. It is not necessary that separate compensating reactances for the source and load be used in the network; a single one of the proper overall value can be substituted as in (C).

Greater complexities result when the *resistances* do not inherently match, as would be the case if R_2 were 50 ohms instead of 500, for example. Simple cancellation of reactances does not suffice for a match in such cases, and measures have to be taken to transform the load resistance to the proper value. This can readily be done, but at the expense of using more complicated matching circuits. Networks suitable for this will be discussed in later chapters.

EFFECT OF IMPEDANCE MISMATCH: When there is only resistance in the load and source (including cases where reactances have been cancelled out, leaving only resistance) the ratio of power delivered to power available is as given in Fig. 10-4, as a function of the mismatch. The mismatch ratio is the ratio either of load resistance to source resistance or source resistance to load resistance, whichever results in a number larger than 1.

If the mismatch is not very great, most of the available power will be delivered to the load. About half the available power is realized when the mismatch is as great as 6 to 1.

THE DECIBEL: In communications work, increases or decreases in power are usually computed by multiplication or division rather than by addition or subtraction. The overall power ratios often become very large, and their direct use tends to be awkward. It is much more convenient to use a logarithmic scale, where equal steps correspond to equal ratios.

Such a logarithmic unit is the bel. One bel is an integral step in the exponents of 10, so that 10^1 (a

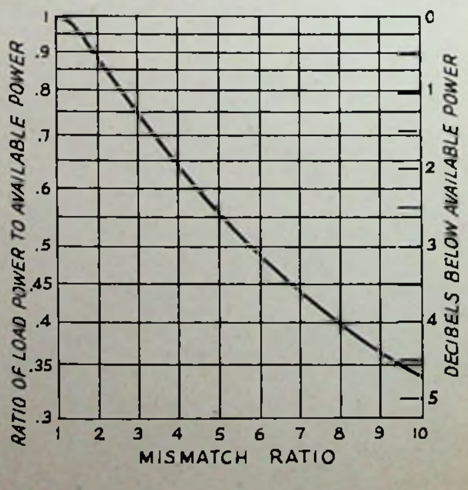


Fig. 10-4

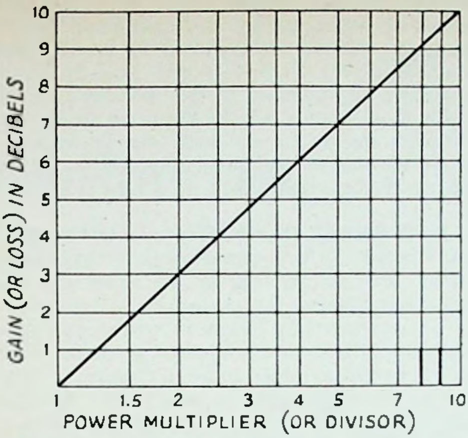


Fig. 10-5

ratio of 10) is 1 bel, 10^2 (a ratio of 100) is 2 bels, 10^0 (a 1 to 1 ratio) is 0 bels, and so on.

The bel is a rather large unit, and it is customary to divide it into ten parts, each of which is called a decibel. To convert a power ratio to decibels,

$$dB = 10 \log P_1/P_2$$

where P_1/P_2 is the power ratio. This relationship is shown in Fig. 10-5, for power ratios from 1 to 10, either multiplying or dividing. Each time the power is multiplied by 10 the gain increases 10 dB; each time it is divided by 10 the number of decibels decreases by 10 dB. For example, if the power is multiplied by 350 (3.5 × 100) the gain is +5.4+10+10 = +25.4 dB. If the power is divided by 26 (2.6 × 10) the loss is −4.2−10 = −14.2 dB.

An advantage of the decibel as a power-ratio unit is that it corresponds roughly to a just-detectable change in the loudness of a sound. (In a general way, the ear responds logarithmically to changes in sound power.) Thus the number of decibels indicates approximately the number of successive multiplications of the power by a constant ratio. For one decibel the power is multiplied by 1.26 for an increase or gain in power, or divided by 1.26 (multiplied by 0.795) for decrease or loss in power.

The right-hand scale in Fig. 10-4 shows the decibels loss caused by a mismatch between a power source and its load.

VOLTAGE AND CURRENT RATIOS IN dB: As stated above, the decibel scale represents a scale of power ratios. By Ohm's Law, $P = E^2/R$, so in terms of voltage ratio the ratio of two powers is

$$\frac{P_1}{P_2} = \frac{\dfrac{E_1^2}{R_1}}{\dfrac{E_2^2}{R_2}}$$

If $R_1 = R_2$, therefore, the power ratio varies directly as the square of the voltage ratio. By similar reasoning it can be shown that the power ratio varies as the square of the current ratio when $R_1 = R_2$.

Under these conditions the number of decibels gain or loss, when voltages or currents are specified instead or power, is found simply by multiplying the power decibel scale by 2. For example, if one voltage is 15 times another, the ratio expressed in decibels is 2 × 11.8 = 23.6 dB.

However, this is true *only* when the resistances in both cases are the same. If R_1 differs from R_2, as it often does in amplifiers at points where voltages are measured, the voltage ratio cannot be expressed in decibels. In such cases the actual powers at the two points have to be calculated so that the true power ratio can be found before converting to decibels.

EFFICIENCY: The total power generated in a source is

$$P = P_O + P_S$$

where P_O is the power delivered to the network (and, in most cases, in large part to the load) and P_S is the power dissipated in the source. The efficiency, or ratio of power output to total power generated, is

$$\text{Eff.} = \frac{P_O}{P_O + P_S}$$

Efficiency is usually expressed as a percentage. Efficiency cannot be as high as 100% because that would imply a source having zero resistance, an impossible situation.

IMPEDANCE MATCHING AND EFFICIENCY: Since equal resistances carrying the same current consume the same power, it follows from the definition of matched impedances that when the match is perfect, half the generated power is used up in the source itself and half in the load. That is, the *available* power is one-half the *generated* power. Thus the efficiency is only 50 percent with perfect matching.

The efficiency depends on the ratio of the load resistance to the source resistance, the reactances being assumed to have been cancelled. As the ratio of load resistance to source resistance is increased the efficiency also increases. The effect of raising the load resistance above the source resistance is shown in Fig. 10-6. When the ratio is *less* than 1 the efficiency decreases below 50 percent, reaching zero when the load resistance is zero (short circuit on the power source). If the source has purely resistive internal impedance and constant internally generated voltage, E, the short-circuit power is four times the available power. It is all lost in the source itself.

SOURCE DISSIPATION: The amount of power dissipated in the source in the course of delivering power to the load is an important consideration. As mentioned earlier, the effect of power dissipation in the source is to cause a temperature rise which, if great enough, can do severe damage. Internal power dissipation therefore must be kept to a safe level in relation to the source's ability to radiate heat.

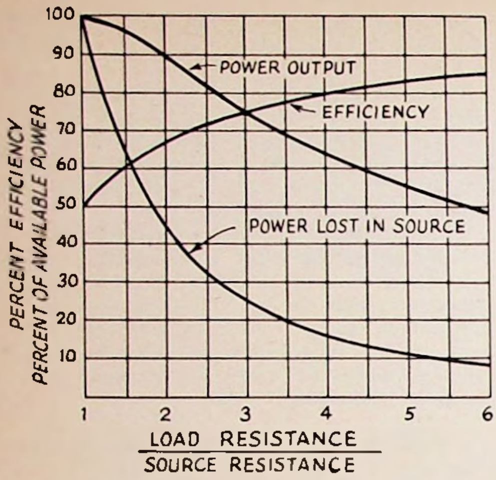


Fig. 10-6

In communications circuits handling *small* signals – microvolts or perhaps millivolts, and power levels from micromicrowatts to milliwatts – efficiency is not usually a prime requirement, and in such cases circuits are usually designed to develop all the available power. On the other hand, when *large* signals – measured in volts and watts – must be handled it is usually necessary to work at or near the highest practicable efficiency and with the lowest possible source dissipation; i.e., to work in the right-hand region of Fig. 10-6. Since it is not possible, because of limitations on source dissipation, to realize all the available power in such cases, the circuits are designed for much higher available power than ever will actually be demanded – of the order of two or more times the power that actually will be delivered to the load. Although the source and load impedances are of course mismatched in these cases, this is the normal mode of operation.

ALTERNATE DEFINITION OF IMPEDANCE MATCHING: In large-signal operation impedances usually are mismatched in the sense discussed above. Nevertheless, it is customary to use the term "impedance matching" to mean matching an actual load resistance to some value of resistance which has been predetermined to be an "optimum" value of load for a given power source. This optimum load resistance is usually a compromise between permissible power loss in the source, the amount of output power needed, and other preconditions such as overall economy of power, the permissible amount of distortion in an amplifier, and so on. The operating conditions under which such a value of load is optimum may vary from application to application, or the optimum load itself may be different in different applications.

In impedance matching of this type it is assumed that the reactances in both source and load either are negligible or will be cancelled by the matching network. The load on the power source is therefore a pure resistance, or very close to it. The design of matching networks follows the same rules for either type of impedance matching; only the impedance to be matched is different.

Questions and Problems

10-1) What is meant by conjugate matching?

10-2) When is it possible to realize all the power available from a source?

10-3) If a power source having an internal resistance of 500 ohms is delivering power to a 1000-ohm resistive load, what is the efficiency? What percentage of the available power is dissipated in the load?

10-4) If the power available from the source in Q. 10-3 is 10 watts, what is the actual power delivered to the 1000-ohm load? How much power is dissipated in the source?

10-5) What function does the connecting network between a source of power and its load perform?

10-6) The resistance looking into the input terminals of an amplifier is 50,000 ohms, and the load on its output terminals is 10 ohms. If the voltage across the load is 2.5 volts when the voltage applied to the input terminals is 0.01 volt, what is the gain of the amplifier in decibels?

10-7) What is the voltage gain of the amplifier of Q. 10-6? How does it compare with the power gain?

10-8) A power source having an internal resistance of 400 ohms is delivering power to a 1200-ohm load. What is the efficiency? If the power in the load is 100 watts, how much power is dissipated in the source?

10-9) If a source of power has an internal resistance of 800 ohms and is delivering power to a resistance of 1600 ohms, what is the efficiency? If the source is capable of dissipating 60 watts safely, how much power can be put into the load?

10-10) If the load resistance in Q. 10-9 is changed to 800 ohms, how much power can be put into it without exceeding the safe dissipation in the source; What if the load is changed to 500 ohms? To 3000 ohms?

10-11) Under what conditions is it undesirable to attempt to obtain the maximum possible power from a source?

10-12) Under what conditions is it advantageous to design for maximum available output?

Chapter 11

The transformer; primary and secondary currents; primary/secondary phase relationships; impedance ratio; transformer circuit network; voltage regulation; stray capacitances; core loss; core construction; the autotransformer; frequency considerations; shielding.

Transformers

PROBABLY MOST circuits used in communication employ magnetic induction in one way or another. The purpose may simply be to separate direct from alternating currents, or it may be to provide a means for modifying impedances for optimum transfer of power from a source to a load. The form the network takes depends on the application for which it is intended. The network discussed in this chapter is the transformer.

THE TRANSFORMER: One of the most useful networks for matching unlike values of resistance is the transformer. The operation of this device is based on the mutual inductance between two coupled inductances. The basic transformer circuit is shown in Fig. 11-1. This representation neglects a number of inevitable stray effects that become important at frequencies above (and sometimes in) the audio range; these effects are considered later.

It will be recalled from Chapter 4 that if a varying current – specifically, an alternating current – is flowing in L_1 a voltage will be induced in L_2, and vice versa, as a result of the mutual inductance. If the coefficient of coupling is 1, it is possible to ignore the actual inductances and calculate the *relative* values of the induced voltages entirely on the basis of the numbers of turns in each coil. This is because when k = 1 all the turns of both coils link with all the magnetic flux lines, so that the voltage induced *per turn* is the same in both coils.

This means that if L_1 and L_2 have N_1 and N_2 turns, respectively, the ratio of the voltages E_1 and E_2 induced in the two coils (the voltage ratio) will be N_1/N_2. That is,

$$\frac{E_1}{E_2} = \frac{N_1}{N_2},$$

or

$$E_2 = \frac{N_2}{N_1} E_1.$$

PRIMARY AND SECONDARY CURRENTS: If the secondary of the transformer is

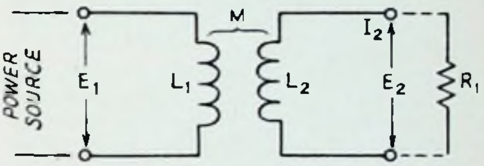


Fig. 11-1

"open" – no current flowing in the secondary circuit – the primary current is simply the current that would flow in the primary inductance, L_1, at a specified applied voltage. This current, called the magnetizing current, is quite small in a well-designed iron-core transformer because the primary inductance will have been made large enough for this to be the case. In an "ideal" transformer – one having no internal losses – this current lags the applied voltage by 90 deg and causes no power dissipation. The open-circuit secondary voltage is then established by the turns ratio as explained above.

When a load is connected to the secondary terminals and power is being delivered, the current flowing in the secondary winding sets up a varying magnetic field. By Lenz's Law there must be a corresponding primary current that will set up a field that exactly opposes the field set up by the secondary current, thus balancing out the increase in field and leaving only the field set up by the magnetizing current.

In an ideal transformer the power delivered to the load by the secondary must exactly equal the power entering the primary. That is, the product of current and voltage (excluding magnetizing current) must be the same:

$$E_1 I_1 = E_2 I_2$$

or

$$\frac{E_1}{E_2} = \frac{I_2}{I_1}.$$

Thus the ratio of secondary load current to primary load current is the same as the ratio of primary voltage to secondary voltage.

PRIMARY/SECONDARY PHASE RELATIONSHIPS: Since it is necessary that the fields set up by the secondary current and by the primary current that flows to offset it be opposite in direction, it follows that if the turns of the two coils are wound in the same way – i.e., both clockwise or both counterclockwise – the primary and secondary currents must flow in opposite directions. That is, the primary and secondary load currents are 180 deg out of phase. (If the direction of winding of one coil is reversed the fields must still cancel, so the current in that coil also must reverse direction. The currents are therefore still out of phase so far as the magnetic fields are concerned.)

Also, this 180-deg phase relationship must hold regardless of the phase between the secondary voltage and secondary current. In other words, whatever phase exists between current and voltage in the secondary circuit is transferred to the primary current and voltage.

The phasor diagram of Fig. 11-2, which assumes an ideal transformer, may make these relationships more clear. The magnetizing current, I_M, is used as the reference. I_M lags the applied voltage, E_1, by 90 deg, as explained in Chapter 7. The induced secondary voltage, E_2 lags I_M by 90 deg; its phase is the same as that of the *induced* primary voltage (see discussion on inductive reactance, Chapter 7) because both coils are in the same varying magnetic field. The secondary current, I_2, is balanced by a primary current, I_1, exactly 180 degrees out of phase with I_2, as explained above.

E_1 and E_2 are drawn to equal length in Fig. 11-2, indicating a voltage ratio of 1 to 1. I_1 and I_2, also are equal in length, since the power is the same in the secondary as in the primary. The 1-to-1 voltage ratio has the advantage that it shows more clearly the equality of the primary and secondary load currents in their effect on the magnetic field. Actually, it is the number of ampere-turns in the primary and secondary that must be equal, since the ampere-turn product represents the magnetizing force of both windings. If the phasors I_1 and I_2 are taken to represent ampere turns instead of just current, they will always be of equal length in such diagrams. In such case the voltage phasors are also equal in length, and represent *volts per turn* rather than the actual voltages.

The *total* primary current is the vector sum of I_M and I_1. I_M remains unchanged regardless of the load current, since the effect of I_2 is always balanced out by the corresponding value of I_1.

IMPEDANCE RATIO: In this simplified picture it is assumed that the transformer is perfect; therefore if a load is connected to L_2 and a source of power to L_1, whatever power appears in the load will be taken from the source without any power loss in the transformer. If the load is resistive, the power in it is E_2^2/R_L, R_L being the load resistance. Since the same amount of power is entering the primary, L_1, the primary power must equal E_1^2/R_p, where R_P is an *apparent* resistance which the primary simulates when the load is connected to the secondary. That is,

$$\frac{E_1^2}{R_P} = \frac{E_2^2}{R_L},$$

and

$$R_P = \frac{E_1^2}{E_2^2} R_L.$$

Further, since $E_1/E_2 = N_1/N_2$,

$$R_P = \left(\frac{N1}{N2}\right)^2 R_L.$$

In other words, a known actual load resistance, R_L, will be transformed into a new value of resistance, R_P, which the source of power will "see" looking into the primary terminals of the transformer. R_P is not "real," in the same sense that a resistor is real, but it will take power in exactly the same way that a resistor of the same value would. This power of course is passed on to the load via the magnetic field.

R_P/R_L, called the impedance ratio of the transformer, is equal to the square of the turns ratio. Although the impedance *ratio* is a function of the turns ratio, the *value* of R_P depends on the value of R_L. By proper choice of turns ratio any value of load impedance can be transformed into a desired load resistance for the power source.

Complex loads, containing both resistance and reactance, are similarly transformed because, as demonstrated in Fig. 11-2, the phase angle of the primary voltage and current is the same as the phase angle of the secondary voltage and current. A given value of reactance in the load is transformed to a new value in the primary by the same ratio as in resistance transformation. The transformed reactance is of the same type as the reactance in the load.

In many cases where transformers are suitable devices for impedance matching, as at audio frequencies, the reactance in the load is small compared with the resistance. For ordinary calculations in such cases it is possible to ignore load reactance.

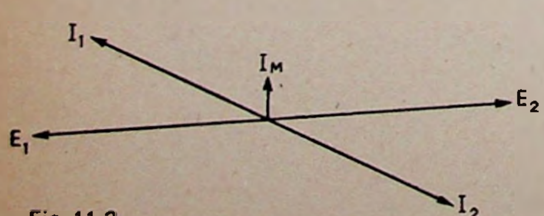


Fig. 11-2

TRANSFORMER CIRCUIT NETWORK: In practice, none of the simple formulas given above will hold exactly, although in many cases they may give satisfactory approximations. A more complete equivalent circuit of the transformer is shown in Fig. 11-3. In this diagram R_1 and R_2 represent the effective resistances of the two windings. The resistance of the wire used in the windings increases when the frequency is increased because of the "skin effect" described in a later chapter.

In actual transformer construction it is not possible for the coefficient of coupling to be

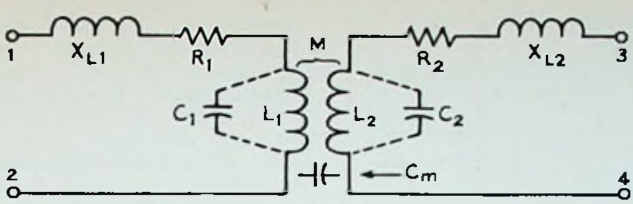


Fig. 11-3

exactly 1: there is always some magnetic flux set up by L_1 that escapes being linked with the turns of L_2. This is called leakage flux, and represents a leakage inductance associated with L_1. Similarly, there is a leakage inductance associated with L_2. These inductances result in leakage reactances X_{L1} and X_{L2} associated with the respective coils.

When an ac voltage is applied to terminals 1-2, the current flowing through X_{L1} and R_1 causes voltage drops which reduce the voltage that actually is effective across L_1. This voltage, not the supply voltage, must be used in the equation for E_1. E_2 then appears across L_2, but when a load is connected to terminals 3-4 the flow of current through R_2 and X_{L2} causes additional voltage drops between L_2 and terminals 3-4. Further, the flow of current in the secondary circuit is accompanied by increased current flow in the primary circuit, and when more power is delivered to a load, the increased primary current causes proportional increases in the voltage drops in X_{L1} and R_1. Thus the larger the load current the smaller the voltage effective across L_1.

Since only the terminals 1-2 and 3-4 are actaully accessible for making measurements, the overall effect – i.e., the actual ratio of applied primary voltage to secondary terminal voltage under various secondary loads – is the only thing that can be measured. Such a measurement can be compared with the theoretical ratio based on the turns ratio alone, but this in itself does not isolate the various reactances and resistances. There are test methods for doing so, but they are outside the scope of this qualitative treatment.

VOLTAGE REGULATION: Because there are voltage drops inside the transformer, the voltage actually delivered to a load is not the same as the voltage induced in the secondary, even if the primary applied voltage is held constant with different loads. That is, the voltage regulation (see Chapter 3) of the transformer is not perfect.

The percent voltage regulation depends on the transformer design and the applied frequency. At power and low audio frequencies it can be of the order of 5 to 10 percent, with transformers designed for the larger power levels having the better regulation figures, as a general rule. At the higher audio frequencies the regulation tends to become poorer because the voltage drop in the leakage reactances is directly proportional to frequency for a given current.

STRAY CAPACITANCES: As explained in Chapter 1, an electric field exists between any two points having a difference of potential. Because each turn in a coil has a different potential than its neighbors when current is flowing, such fields are set up between turns. There is therefore a small capacitance between turns, called distributed capacitance.

For the primary coil, L_1, in Fig. 11-3, this capacitance is indicated by C_1. It is not actually a single capacitance in parallel with the coil, as drawn, but is the overall capacitive effect of numerous tiny capacitances between turns. A similar capacitance, C_2, is associated with L_2.

Further, there is an additional capacitance, C_m, between the two windings, for a similar reason. Not shown in Fig. 11-3, but nevertheless often important, is still another capacitance that exists between the windings and nearby metal objects (such as a chassis or shield) and between the windings and the core in transformers using magnetic cores.

These capacitances are of little practical interest at power-line and low audio frequencies, but as the frequency applied to the transformer is increased the stray capacitive currents between turns and windings begin to become appreciable. Such currents cause extra heat to be developed in the resistance of the windings and thus increase the transformer losses.

If the frequency is made high enough, the capacitance of a loaded transformer will self-resonate with the leakage inductance. In the frequency region around resonance the transformer no longer follows the principles outlined above, particularly in regard to impedances. At still higher frequencies (above resonance) the transformer acts like a complex capacitive circuit and loses its normal properties entirely.

In addition to resonating with the leakage inductances, the stray capacitances also resonate (at a much lower frequency) with the primary and secondary inductances themselves. This type of resonance assumes importance only when there is no load or a very light load on the secondary. It seldom needs consideration in present-day low-frequency circuits because transformers for use in such circuits are practically always operated with a secondary load. The load resistance minimizes the resonance effect, for reasons discussed in Chapter 12.

Resonances of both types will occur in all transformers, including those designed for low frequencies as well as high. If the transformer is intended for impedance matching over a range or band of frequencies, resonances should be avoided, by proper design, if they will affect the impedance ratio at frequencies within or close to the desired range. At radio frequencies, transformers are designed so that the response to various frequencies is "shaped" by proper utiliza-

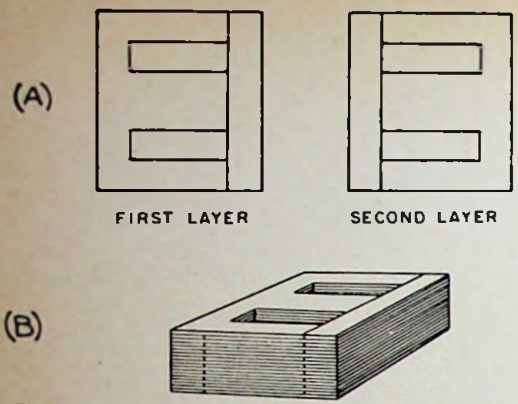


Fig. 11-4

tion of resonance effects, as described later in the chapter on rf circuits.

CORE LOSS: In addition to the power loss in the resistance of the windings, there is also a loss in the core of a magnetic-core transformer. One cause of this is the fact that energy must be expended in overcoming retentivity in the magnetic material (Chapter 4). This is the hysteresis loss; it depends on the inherent magnetic properties of the core material. A second cause is that the varying magnetic field induces voltages in the core itself. These cause circulating currents (eddy currents) in the core, and power is lost when these currents flow through the core resistance. Eddy-current loss can be reduced by making the resistance of the current paths as high as possible; the method depends on the type of core.

The core loss is caused by the field set up by the magnetizing current alone, since only one value of rate of change of magnetic flux is needed for maintaining the induced voltage in the primary equal to the applied voltage. The power needed for supplying the loss is accounted for by an increase in the magnetizing current over the value that would be required if there were no core losses. Although not shown in Fig. 11-3, the loss can be represented by an equivalent resistance (relatively high in a well-designed transformer) in parallel with L_1.

CORE CONSTRUCTION: For audio-frequency and power-frequency (60 Hz) operation the common core material is one of a number of varieties of silicon steel. Although the magnetic properties vary with the different types, the material typically has a permeability of 5000 or more and becomes saturated at a flux density of the order of 100,000 lines per square inch of cross section. The core is always built up in layers or laminations, which are thin stampings having the E and I shapes shown in Fig. 11-4(A). The purpose of laminating the core is to break up the paths through which eddy currents flow and thus reduce the power loss. The laminations often have a very thin coating of insulating varnish to assist in this.

The core is assembled by stacking the laminations as shown in Fig. 11-4(B), with alternating E and I pairs. Alternating the pairs improves the magnetic properties of the core in that each small gap in the magnetic path where an E butts up against an I is bypassed by the continuous magnetic path of the laminations above and below. This increases the flux density per unit of applied magnetic force and reduces flux leakage from the core.

The complete transformer is shown in Fig. 11-5. The coils are on the center leg of the core, and as a result are fairly well surrounded by it. This helps reduce leakage inductance by providing a relatively large ratio of iron to air for the magnetic flux.

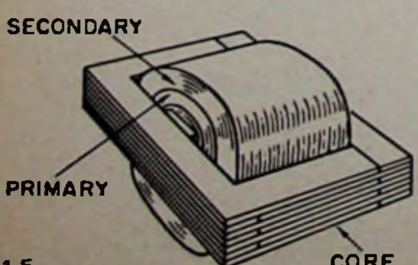


Fig. 11-5

THE AUTOTRANSFORMER: A variant of the conventional two-winding transformer is the

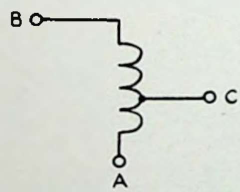


Fig. 11-6

autotransformer, which has only one. A tap on the single coil serves the same purpose as the second winding of an ordinary transformer.

Such a tapped transformer is shown in Fig. 11-6. One terminal is common to the primary and secondary circuits, making a direct connection between the two. (In the two-winding transformer the two circuits can be completely separate.) Any pair of terminals may be used as the primary, with proper voltage applied, and one of these, plus the third terminal, gives the secondary output. For example, if *A* is common, *A* and *B* may be used as the primary, and *A* and *C* as the secondary.

The voltage and impedance ratios are determined in the same way as for a transformer having separate windings, using the number of turns between taps in place of the number in a separate coil.

The section of the autotransformer winding that is common to both the primary and secondary circuits carries the *difference* between the primary and secondary currents, since these two currents are out of phase. This, plus the smaller number of windings, can result in some economy in transformer design. The disadvantage is the necessity for accepting the common connection between the two circuits. In some applications this is not feasible.

FREQUENCY CONSIDERATIONS: The general principles outlined in this chapter apply to transformers of all types. However, at different frequencies different operating features will assume major importance, and one or another characteristic will be given special emphasis.

As already mentioned, at power and audio frequencies in the communications range – roughly 50 to 3000 Hz – resonance effects are either small enough to be neglected or can be handled quite readily in circuit applications. Also, reactance in the load or in the circuit as a whole is generally negligible, or else not a serious factor. As a result, the simple relationships given early in this chapter often will be sufficient for practical applications. It must always be remembered, however, that *exact* results cannot be obtained with them.

At radio frequencies, reactances both in the transformer and in the circuit greatly modify the picture, and the simple primary/secondary voltage, current, and impedance relationships described earlier may be inapplicable. Radio-frequency considerations are covered in subsequent chapters.

SHIELDING: Because magnetic lines of force are continuous and closed upon themselves, as explained in Chapter 4, it is not possible to confine a magnetic field to a given space by the shielding method described in Chapter 1, Fig. 1-4, for electric fields. There are no "magnetic charges" to be drained off to ground.

At low frequencies (at least through the upper audio range) the solution is to offer an attractive path for the leakage flux so the flux will stay in it rather than spread out around the coil or transformer. Construction of the type shown in Fig. 11-5 helps greatly, because by far the greater portion of the total flux will stay in the iron core. Nevertheless, some flux does exist in the air surrounding the assembly, and it may be enough to cause difficulty in sensitive circuits nearby.

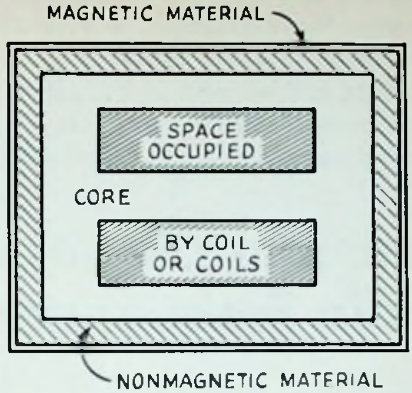


Fig. 11-7

If the transformer or inductor is enclosed in a case made of good magnetic material most of this outside flux can be "captured" and made to flow in the container itself, leaving the space outside the container substantially field-free. Such an arrangement is shown in cross section in Fig. 11-7. The effectiveness is increased if the case is separated from the core by a nonmagnetic space, because such a space has high reluctance and the flux is attenuated before reaching the case. Also, the case should surround the coil/core assembly as completely as possible.

A different method of shielding is used at radio frequencies, as described in the next chapter.

Questions and Problems

11-1) If the impedance ratio of a transformer is 100 to 1, what is the corresponding turns ratio?

11-2) Why must a different type of shield be used for low-frequency magnetic fields than for electric fields of the same frequency?

11-3) Given the turns ratio of a transformer, can you deduce the actual numbers of turns in its windings?

11-4) If a transformer is to deliver 50 volts at 1 ampere to the load on the secondary and the voltage applied to the primary is 200 volts, the frequency being 500 Hz, what would you consider to be a suitable value for the primary inductance?

11-5) Would you consider the transformer of Q. 11-4 to be suitable for a frequency of 50 Hz, under the same operating conditions?

11-6) What is leakage reactance, and what causes it?

11-7) Can dc measurements (by an instrument such as an ohmmeter) give you a fairly accurate idea of the effective resistance of a transformer winding at 1000 Hz? At 100 Hz?

11-8) Why is there a magnetizing current in a transformer? Is there a power loss associated with it?

11-9) What is the relative phase of the primary and secondary load currents?

11-10) What is the overall effect of transformer resistance and leakage reactance?

11-11) In a transformer designed for working over the audio-frequency range, would you expect the voltage regulation to be the same at, say, 3000 Hz as it is at 150 Hz?

11-12) What is the ampere-turn primary-to-secondary ratio of an ideal transformer?

11-13) Why does the primary current increase when a load is connected to the secondary of a transformer?

11-14) What effects do stray capacitances have in a transformer?

11-15) What are eddy currents? How may they be reduced?

11-16) Are hysteresis and eddy-current losses related?

Stray reactances; resonance; series resonance, Q; bandwidth; parallel resonance; definitions of parallel resonance; impedance curve of parallel circuit; resonant voltage in series circuits; circulating current; skin effect; coil resistance at rf; magnetic materials at rf; distributed capacitance of rf coils; radio-frequency chokes; inductance in capacitors; shielding rf coils.

Resonant Circuits at Radio Frequencies

A PRIMARY DIFFERENCE between an rf circuit and one for frequencies below approximately 20 kHz (the upper audio-frequency limit) is that stray reactances are important in the one case and relatively unimportant in the other. At radio frequencies it is necessary to *introduce* reactances to balance out or overcome the effects of reactances that are unavoidably present in conventional components and wiring. Failure to do this may result in a completely inoperative circuit.

STRAY REACTANCES: The reason is that stray reactances, both capacitive and inductive, vary with frequency in such a way that their values become comparable with those intended to be used in rf circuits. A shunt capacitance of 20 pF, for example (not unusual for a stray) represents almost 10 megohms at 1000 Hz but is less than 1000 ohms at 10 MHz. Similarly, an inductance of 5 μH in a wiring connection would not affect the operation at audio frequencies, but at 10 MHz the same lead inductance would represent a reactance of 300 ohms – a fatal stray in most rf circuits. Further, transformer impedance matching as described in Chapter 11 is impractical because self-resonances (as well as high losses) in a transformer of conventional design would destroy the transformer action.

Neutralization of reactance in a power source and its load already have been discussed briefly (Chapter 10). In essence, it amounts to introducing a reactance of the same value but of opposite kind to form a resonant circuit.

In addition to the "correction" of stray reactances, there are other considerations that operate at radio frequencies and not usually at low frequencies. One of these is the need for the ability to separate a desired signal from the thousands that are constantly present in the radio medium. Fortunately, the requirement for selectivity and the need for overcoming stray reactances are not incompatible; in fact, both are achieved by the same method – the resonant circuit, or a combination of resonant circuits.

RESONANCE: The compensating reactance may be placed either in series or parallel with the reactance to be suppressed. That is, either a series- or parallel-resonant circuit may be formed. While the choice often can be a matter of convenience, there are cases where it will be dictated by the circuit itself or by practical considerations which will be touched on in later chapters.

Since complete cancellation of reactance can occur at only one frequency, the external circuit will be sufficiently compensated only at frequencies close to the resonant frequency. Thus the useful range of frequencies – the bandwidth – becomes an important consideration in circuit operation.

SERIES RESONANCE, Q: The current in a series-resonant circuit such as Fig. 12-1 can be shown graphically by the resonance curve of the circuit. This is obtained by applying a voltage, E, of fixed amplitude and variable frequency and observing the value of current at various frequencies. The current as plotted on the graph is usually relative; that is, the values are shown as a decimal fraction of the maximum current. The maximum occurs at the frequency for which the inductive and capacitive reactances are equal and opposite. The *actual* current at resonance depends on the amount of resistance, R, in the circuit at this frequency, and is equal to E/R. At frequencies off resonance the current is determined by the impedance of the circuit as discussed in Chapter 8, and the larger the value of $X_L - X_C$ at a given frequency the smaller the current. If $X_L - X_C$ is

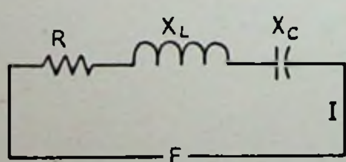


Fig. 12-1

large compared with R, the resistance has very little effect on the current amplitude.

The relationship between these three quantities, in their effect on the shape of the resonance curve, can be expressed in terms of X_L/R, where X_L is the inductive reactance *at the resonant frequency* and R is the effective resistance at the same frequency. This ratio is called the Q or quality factor of the circuit.

The way in which X_L and X_C vary with frequency is known (Chapter 7). In general, R too will vary with frequency since it represents, among other things, losses which vary with frequency. The variation in R, although not predictable with high accuracy, has an effect on the resonance curve. Fig. 12-2 shows two resonance curves, one based on the assumption that R is constant with frequency, the other assuming that Q is constant. For constant Q, R decreases with frequency in the same ratio that the inductive reactance does. The shape of the resonance curve is not greatly affected by this except around the peak, which tends to be flattened. However, the whole curve is shifted toward a somewhat lower frequency.

BANDWIDTH: The useful band of frequencies lies between two points of equal current amplitude which represent the lowest values useful for the particular application. For comparing circuits, the bandwidth is frequently defined as the band lying between the two frequencies at which the current amplitude decreases to 0.707 (or $1/\sqrt{2}$) times the maximum value. Since the power consumed in R is proportional to the square of the current, the power at these points is 1/2 the maximum power (assuming constant R). This is called the 3-dB bandwidth, and the limiting frequencies are called the half-power points. These points are indicated by the arrows in Fig. 12-3.

The difference between the two cases shown in Fig. 12-2 tends to disappear when Q is made larger. Fig. 12-3 shows the relative current vs. frequency for two larger values of Q, 10 and 100. These curves are based on constant Q. There is a slightly noticeable departure of symmetry at the peak of the curve for Q = 10, but the Q = 100 curve is quite symmetrical. Since in many applications Qs of 10 or more are common, it is usually possible to assume that resonance curves will be symmetrical.

On this assumption, the 3-dB bandwidth can be calculated from

$$\text{Bandwidth} = \frac{f}{Q}.$$

As an example, if the resonant frequency is 4 MHz and the Q is 10, the bandwidth is 4 MHz/10 or 0.4 MHz (400 kHz). If the Q is increased to 100, the bandwidth decreases by a factor of 100/10 = 10, becoming 40 kHz.

PARALLEL RESONANCE: The operation of a circuit such as is shown in Fig. 12-4(A) has been described in earlier chapters (7 and 8). The resistance R here represents the effective resistance in series with L. No resistance is shown in series with C, because the types of capacitors used at radio frequencies almost invariably have such small power losses that, with little error, their effective

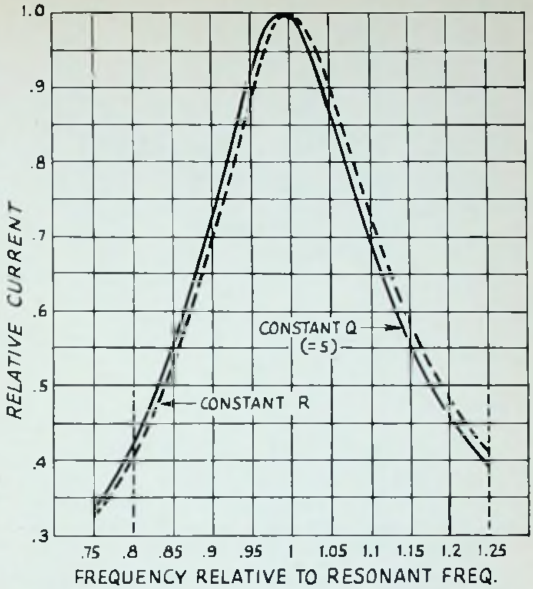


Fig. 12-2

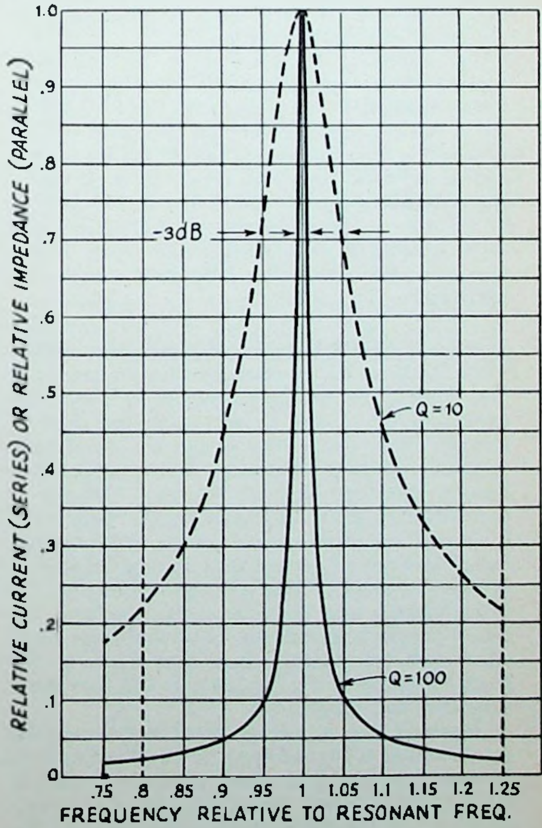


Fig. 12-3

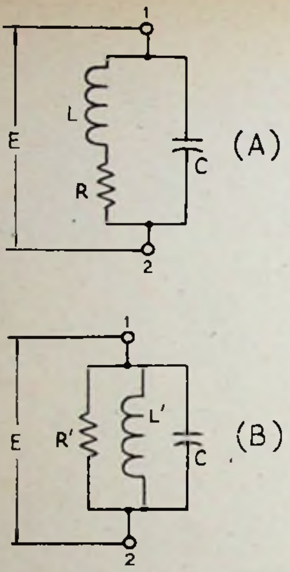


Fig. 12-4

resistance can be assumed to be zero. The capacitance is therefore a "pure" reactance. The inductive branch of the circuit is an impedance, consisting of L and R in series.

As explained in Chapter 9, an equivalent parallel circuit can be formed that will have the same impedance and phase angle as L and R in series. Such a circuit, combined in parallel with C, is shown in Fig. 12-4(B). L' and R' are the parallel equivalent of L and R in series. R' is of course larger than R, and L' is smaller than L, as explained in Chapter 9. (The identity of the parallel and series circuits is true, incidentally, only at the frequency for which the equivalent circuit was obtained.)

DEFINITIONS OF PARALLEL RESONANCE: If we define resonance for the circuit of Fig. 12-4(B) to be the frequency at which the reactances of L' and C are equal, the currents through these two elements also are equal (and opposite) at that frequency. The total current flowing into and out of the two branches is therefore zero, since the reactances are "pure." But R' is still across the circuit, and a current determined by Ohm's Law will flow through R' when a voltage E is applied. This is the line current for the entire circuit, and since it cannot be zero the circuit has finite impedance. The line current I is in phase with the applied voltage E as shown by the phasor diagram of Fig. 12-5(A), so the impedance is resistive and equal to E/I.

However, this is an *equivalent* circuit, and the *actual* circuit is Fig. 12-4(A). If Fig. 12-4(A) is to produce a current I in phase with E, the current through the inductive branch must be as represented by I_L in the phasor diagram of Fig. 12-5(B). The amplitude of I_L is greater than the amplitude of I_C, which means that the reactance of L has to be lower than the reactance of C. In other words, if the line current is to be in phase with the applied voltage so that the circuit "looks" like a pure resistance, the two reactances in Fig. 12-4(A) cannot actually be equal. When the circuit constants are adjusted so that the applied voltage and line current are in phase, the circuit is said to be in unity power factor resonance.

If the reactances of C and L are actually equal, the impedance of the inductive branch will exceed the reactance of the capacitive branch because of the resistance in series with L. I_L therefore will be smaller than I_C, leading to the phasor diagram of Fig. 12-6. (This is constructed by using the semicircle technique described in Chapter 9 to obtain the resistive and reactive voltage drops in the proper amplitude and phase relationships, and then drawing I_L in phase with the resistance voltage drop. The length of the I_L phasor is found by calculating the impedance from the triangle so found, dividing the voltage by the impedance to find the current, and then plotting the current to the same scale as that used for I_C.) The line current I is found by combining I_C and I_L by the usual parallelogram method. I leads the applied voltage in phase, and is somewhat smaller than it was in the unity power factor case, indicating that the overall circuit impedance is higher.

Further investigation will show that, depending on the values of L and R (C being fixed), still a third value of L will result in a *smallest* value of line current. How it comes about is shown in Fig.

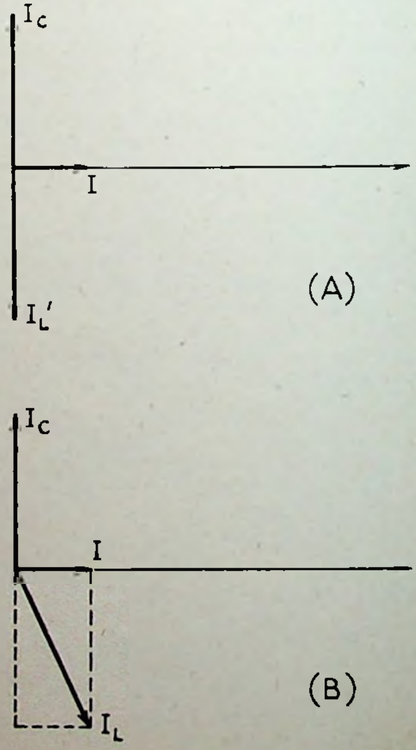


Fig. 12-5

12-7. For a fixed ratio (Q) of X_L to R the phase angle between E and I_L will be constant with different values of X_L. As X_L is varied (and with it R) I_L will take on successive values such as *A, B, D, E*, which combined with I_C result in corresponding line currents I_A, I_B, I_D, I_E. The line current is smallest (maximum-impedance resonance) when its phasor is perpendicular to the I_L phasor, as at I_D. If Q is changed a different set of values will result, but the principle is the same.

The three ways in which parallel resonance can be defined lead to three different resonant frequencies for a given set of constants (L, C, and R). However, the three frequencies will differ noticeably only when the circuit Q is appreciably below 10. The difference increases as the Q is made smaller. But at Qs of 10 or more the three resonant frequencies coincide, for practical purposes, and it is sufficient to call the "resonant frequency" the one for which $X_L = X_C$.

IMPEDANCE CURVE OF PARALLEL CIRCUIT: An important characteristic of the parallel-resonant circuit in equipment design is its impedance curve. The impedance curve shows the numerical value of the impedance over a frequency range centered on the resonant frequency. The phase angle of the impedance may also be plotted, if the information is required; however, in voice and code communication it is rarely needed. It is generally sufficient to know that the current lags behind the applied voltage at frequencies below resonance because at such frequencies the inductive reactance is lower than the capacitive reactance, and that above the resonant frequency the current *leads* the voltage because the reactance of the capacitive branch is lower than the inductive reactance.

If the circuit Q is 10 or more the curve showing relative parallel impedance as a function of frequency is practically identical with the curve for relative current in a series circuit having the same Q. Thus Fig. 12-3 may be used either for the relative current of a series circuit or the relative parallel impedance of a parallel circuit. The maximum impedance occurs at resonance and is equal to E/I, where I is the line current at the resonant frequency. On the assumption that Q is 10 or more, the parallel impedance at resonance is also

$$Z = \frac{X_L^2}{R}$$

where X_L is the reactance of the inductance at resonance and R is the effective resistance in series with it. Since X_L/R is the Q of the circuit, the equation can also be written

$$Z = QX_L.$$

The reactance X_C can be substituted for X_L in the equation since at resonance $X_C = X_L$. Signs are neglected.

As the Q is made higher, the curves become sharper; that is, the impedance of the circuit at frequencies off resonance drops off at a more rapid rate as the frequency is moved away from resonance.

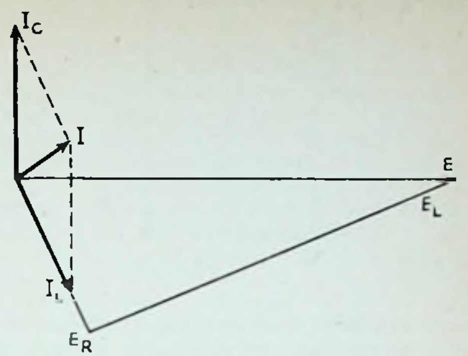


Fig. 12-6

In both the series and parallel circuits the ability to suppress frequencies on either side of resonance is known as the selectivity of the circuit. Selectivity consistent with the required bandwidth is an important feature in most rf circuit design.

RESONANT VOLTAGE IN SERIES CIRCUITS: It was mentioned briefly in Chapter 7 that the voltage developed across X_L and X_C in series circuits can become quite large compared with the applied voltage. The reason is that R (which is generally small compared with the reactance in rf circuits) is the only factor that limits current flow at resonance, and is also the principal limiting factor near resonance where the inductive and capacitive reactances nearly, but not quite, cancel out in the circuit as a whole.

In a circuit such as Fig. 12-1 the voltage developed across X_L by I is IX_L, and the voltage across R is IR. Thus the ratio of reactive voltage to resistive voltage is

$$\frac{IX_L}{IR} = \frac{X_L}{R} = Q.$$

At resonance $X_C = X_L$, so the voltage across either the inductance or capacitance at this frequency is equal to QE. This is indicated perhaps a little more

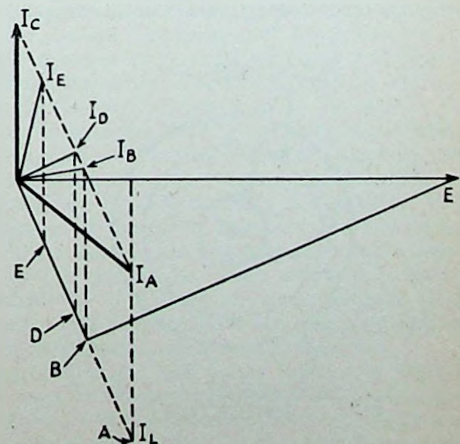


Fig. 12-7

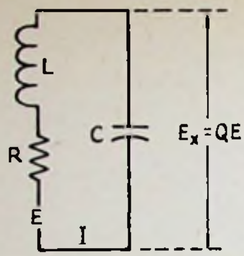


Fig. 12-8

clearly in Fig. 12-8, which is Fig. 12-1 rearranged slightly to show how the voltage (here labeled E_X) looks from "outside" – as though the circuit were parallel-resonant instead of series-resonant. (Actually, the relationship E_X = QE is not quite true. For the inductive branch as shown in Fig. 12-8, E_X is the sum of the voltage drops across L and R. But since these voltages combine in quadrature (Chapter 8) the error is only about 1 percent if Q is 10, and considerably less as Q is increased.)

Because E_X is directly dependent on the current flowing in the circuit, resonance curves such as those in Fig. 12-3 also show the relative value of E_X, because these curves are identical with the curves for relative current in series circuits of the same Q. At any frequency off resonance E_X will be equal to the voltage at resonance (that is, QE) multiplied by the factor shown by the vertical scale for that frequency.

CIRCULATING CURRENT: Since E_X in Fig. 12-8 is dependent on the current I, it follows that if a voltage E_X is applied to the parallel circuit by an outside source of power, the current I inside the circuit is related to E_X in the same way. In the parallel case I is often called the circulating current.

For Qs greater than 10, the circulating current in a parallel circuit at resonance is

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

where X is the reactance, at resonance, of either the inductance or capacitance. For any frequency off resonance the ratio of the actual current to the current at resonance can be found from curves of the type given in Fig. 12-3.

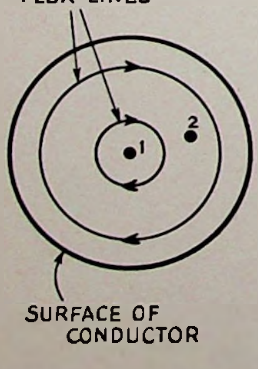


Fig. 12-9

SKIN EFFECT: As compared with the resistance for direct currents, the resistance of a conductor carrying an alternating current becomes larger as the frequency is increased. The reason for this is that the varying magnetic field set up by the ac does not induce the same voltage throughout the conductor.

Fig. 12-9 represents the cross section of a conductor, in a plane containing the magnetic field lines. It can be seen that the element or "filament" of current indicated by the dot numbered 1 is linked by the inner flux line. This same line does not link with a current element (such as No. 2) between it and the conductor's surface. However, both No. 1 and No. 2 are linked with a line between No. 2 and the surface. That is, a current element at the center is linked by a larger number of flux lines than one near the surface. The increased linkages result in increased reactance as the center is approached, hence the greater portion of the current is forced to flow near the surface through an area smaller than the actual cross section. This is equivalent to a reduction in the conductor cross section, and therefore to an increase in resistance.

This skin effect depends on several factors, the more important of which are the conductor size, shape, and material. Skin effect increases approximately with the square root of the frequency. The depth to which the current penetrates in round conductors, for a given frequency and material, is the same for all diameters, and amounts to a few thousandths of an inch at frequencies of the order of 1 MHz. Because most of the current flows so near the surface at radio frequencies a thin tube and a solid conductor of the same diameter and material have essentially equal resistances at rf.

COIL RESISTANCE AT RF: The resistance that determines the Q of a coil (and thus the effective Q of the circuit when no other resistance is deliberately introduced) is the resistance of the conductor of which the coil is made plus the resistive equivalent of any power loss in material that may be in the field of the coil. The "stray" losses, such as those in dielectric and magnetic material, if used, can be held to a minimum by choice of suitable materials and by keeping them, insofar as possible, out of the most intense part of the magnetic field.

The resistance of the conductor itself depends partly on the skin effect discussed above, but when the conductor is wound into a coil there is a further increase in resistance. This is of the same nature as the skin effect, but is caused by the close proximity of the turns (proximity effect). The current tends to concentrate in the parts of the conductors that are closest together, causing a further reduction in the effective cross section of the conductor.

Altogether, the resistance of a given coil tends to increase directly with the frequency, so that its Q remains almost constant over the frequency range for which the inductance is usable. The Qs of practical coils can range from values as low as 25 to several hundred. The lowest values are usually associated with small-diameter coils of relatively low inductance, wound with small wire and having iron cores. High values of Q can be obtained if the coil is wound with large wire or tubing, has a large diameter, and has its turns spaced to reduce proximity effect. It also helps to use a minimum of dielectric, as well as low-loss material, in the supporting form.

MAGNETIC CORE MATERIALS AT RF: Hysteresis and eddy-current losses (Chapter 11) in the iron cores used for low-frequency work increase tremendously at radio frequencies, and cores of that type become unusable above the af range. Special measures have to be taken to reduce losses in magnetic cores for radio frequencies.

One such core is the powdered-iron type. It uses a special grade of iron, having low retentivity, ground into a powder. The powder is mixed with a binder that insulates the iron particles from each other so that eddy currents are minimized. The core is usually cylindrical and has the coil wound around it on an insulating form. With this construction a large part of the path for the magnetic flux is in air, so the permeability of the iron/air core is increased only by a factor of 2 or 3 over the permeability of air alone. Nevertheless, with suitable core material the overall resistance loss can be reduced and the Q thereby increased, for the same inductance. This is because the effective resistance of the inductor is lowered more by the reduction in the number of turns (and hence wire resistance) than it is increased by the core losses. The relative positions of the coil and core are frequently made adjustable so that the inductance can be varied conveniently.

Another useful core material is ferrite, a ceramic-like material having magnetic properties along with the very high resistivity of an insulator. It is made in several varieties for various frequency ranges. Eddy-current losses are negligible in ferrites. The core is usually made in the form of a ring or toroid, thus providing a continuous magnetic path for the flux set up by a coil wound on it. The high permeability of the closed core greatly reduces the number of turns required for a given inductance, and relatively high Q values result. The closed core also reduces leadage inductance to a very low figure.

DISTRIBUTED CAPACITANCE IN RF COILS: It was mentioned in Chapter 11 that capacitance exists between turns of a coil, but that at low frequencies the effects of this capacitance do not become appreciable until the frequency is well up in the audio range.

Distributed capacitance also exists in rf coils, and because a very small capacitance has a relatively low reactance at radio frequencies it is seldom possible to ignore the capacitance entirely. As shown in Fig. 12-10, it exists not only between adjacent turns but between *any* two turns in the coil. Actual values of capacitance vary with the separation between turns. Stray capacitance also exists between individual turns of the coil and any conductors that may be in the vicinity, so the overall capacitive effect is quite complex.

Distributed capacitance causes internal resonance effects that are substantially independent of whether or not the coil is connected into a circuit. If these resonant frequencies are measured by an instrument such as a dip meter it is found that they group themselves into two types. If the coil is measured with its ends unconnected, there is a group of resonant frequencies beginning with a lowest one and followed by successively higher ones, not harmonically related to the lowest. If the coil is measured with its ends short-circuited, another such group is found. The former are usually called "parallel" resonances and the latter "series" resonances. At the parallel-resonant frequencies the coil has high impedance, mostly resistive, similar to that of a parallel-resonant circuit of the type discussed in this chapter. The impedances at the "series" frequencies are low, like the impedance of the series-resonant circuits considered earlier.

Distributed capacitance can be reduced, in coils having single-layer windings, by using a small turn diameter and by spacing the turns. The form on which the coil is wound, if one is used, should have a low dielectric constant. To reduce stray capacitance the coil should be separated from other conductors as much as circumstances permit.

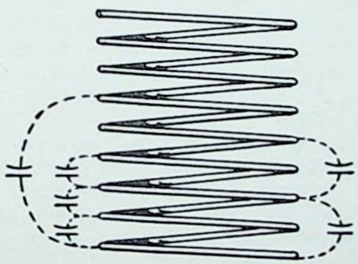


Fig. 12-10

RADIO-FREQUENCY CHOKES: Coils having inductances much larger than those used in the tuned circuits that function as described earlier in this chapter are frequently used for isolating a dc source of power for an amplifying device, as well as in other cases where one circuit must be isolated from another for radio frequencies but not for dc or low frequencies. Such a coil is called a radio-frequency choke. Ignoring other circuit details, the coil is usually in parallel with a tuned circuit, as shown by the simplified circuit of Fig. 12-11. The impedance of the power supply is assumed to be very low at the rf operating frequency; this usually will be the case, if the power supply is properly bypassed by a low-reactance capacitance. The resonance effects discussed in the preceding section become very important in such applications.

At frequencies well below the lowest parallel resonance the reactance is inductive – that is, the coil behaves like a normal inductance. Since the coil is in parallel with L in Fig. 12-11, the total circuit inductance is reduced to the value repre-

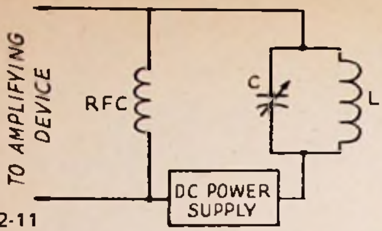


Fig. 12-11

sented by the inductances of RFC and L in parallel. This reduction is generally small, since RFC is ordinarily many times larger than L.

Around its parallel-resonant frequencies the choke simulates a very high resistance. If this resistance is large compared with the load resistance required by the rf power source, very little power is lost in the choke, and RFC serves its intended purpose. When the operating frequency is somewhat *above* the lowest parallel resonance the choke shows a high capacitive reactance (the effective capacitance is 1 or 2 pF, usually) in parallel with a high resistance. This is, in fact, generally the case at any frequency above the lowest parallel resonance *except* around the higher resonances. Around any *series*-resonant frequency the choke exhibits low resistance in series with low values of reactance, and becomes ineffective because it absorbs a relatively large amount of rf power. Also, its effect on the normal tuning behavior of the LC circuit is very pronounced, while at the frequencies at which the choke is operating properly the principal such effect is the addition of the 1 or 2 pF mentioned above to the tuning capacitance.

The series-resonant points are known as "holes" or "dead spots" in the choke. Their position in the frequency spectrum depends on the construction of the choke, and cannot be predicted. Experimental methods have to be used to design chokes that will work well over a wide range of frequencies.

INDUCTANCE IN CAPACITORS: Practical capacitors have inductance inherent in the plates and in the leads used to make connection to them. The internal inductance cannot be reduced, in a given component, and at some frequency the capacitance and inductance will become series resonant. At this frequency the capacitor acts like a short-circuit. Above the series-resonant frequency the capacitor becomes increasingly inductive.

Internal inductance is most serious in capacitors having foil/paper construction and in the larger values of capacitance. Where such capacitors are used as rf bypasses, the bypassing is effective at frequencies up to and somewhat beyond the series-resonant frequency, but eventaully, as the frequency continues to be raised, the inductance destroys the bypassing action. With "paper" capacitors this can occur at various frequencies in the hf range, depending on the capacitance and the internal construction. Capacitors with ceramic dielectric, which have a high dielectric constant, start becoming resonant at about the low-frequency end of the vhf range, in the larger capacitance values available. The length of the external leads has a considerable bearing on the resonant frequency.

Because the effective impedance is very low at and near resonance, the bypassing action is best at the resonant frequency, regardless of the capacitance. As the operating frequency is raised, the optimum bypass capacitance becomes smaller, because the internal inductance does not differ greatly in the various sizes of capacitance.

SHIELDING OF RF COILS: The magnetic shield described in Chapter 11 for low-frequency coils and transformers is not desirable at radio frequencies, because the losses in ordinary iron or steel become excessive at high frequencies. A different principle is used at radio frequencies. A high-conductivity nonmagnetic container (usually copper or aluminum) is used. The magnetic field of the coil induces a voltage in the shield, causing an eddy current to flow. This current, by Lenz's law, generates a magnetic field which opposes the field producing it, so that substantially no field is detectable outside the shield if the container is completely closed. If the shield is grounded, there is also no external *electric* field, as discussed in Chapter 1.

Good conductivity is essential because the resistance to the flow of eddy currents must be as low as possible if the field cancellation is to be complete. Also, the amount of energy needed to supply the losses in the shield will be small if the resistance is low: this energy is equivalent to an increase in the resistance of the coil itself and reduces its Q. The inductance is also reduced by the field set up by the shield. For these reasons the shield should not be too close to the coil, especially in those regions where the magnetic field is strongest. If the shield diameter is at least twice the coil diameter and the shield is at least a coil diameter from each end of the coil, these effects amount to only a few percent.

Questions and Problems

12-1) Why is a nonmagnetic shield preferred to a magnetic one at radio frequencies? What properties should the shield have?

12-2) Define the three conditions that can be called parallel resonance.

12-3) When is it necessary to differentiate between the three conditions mentioned in Q. 12-2?

12-4) What is the effect of circuit Q on the parallel-resonance impedance curve?

12-5) The rf voltage applied across a parallel LC circuit is 350 volts. If the circuit Q is 35, what is the current inside the circuit at resonance?

12-6) What is the cause of skin effect?

12-7) A circuit has inductive and capacitive reactances of 250 ohms each and a series resistance of 10 ohms. What is its parallel impedance at resonance? If the frequency is 2 MHz, what is the 3-dB bandwidth?

12-8) If a voltage of 650 at the resonant frequency is applied to the parallel circuit of Q. 12-7, what is the circulating current?

12-9) A parallel circuit consists of a 100-pF capacitor and a 2-μH inductor. If the circuit Q is 12, what is the parallel impedance at resonance, and what is the 3-dB bandwidth?

12-10) If a low-Q parallel circuit is to be tuned to take the smallest possible line current when a fixed rf voltage is applied, what type of parallel resonance is required?

12-11) What effect does distributed capacitance have in a radio-frequency choke coil?

12-12) What determines the voltage across the terminals of the inductor in a series-resonant circuit? Across the capacitor terminals? What happens when the applied frequency departs from the resonant frequency?

12-13) Is there a limit to the frequency at which an inductor can be used? If so, what determines it?

12-14) What is the means used for combating stray inductance and capacitance at radio frequencies?

12-15) What is meant by the half-power point?

12-16) A 20-μH inductor has a resistance of 5 ohms at 2 MHz. What is its Q? If the resistance increases to 8 ohms at 3 MHz what is its Q at that frequency?

12-17) If the -3 dB frequencies of a parallel circuit are to be 3.5 and 4 MHz and the maximum impedance within the band is to be 1200 ohms, what values of L, C, and R are required? (Hint: assume that the impedance curve will be symmetrical about the midfrequency of the band.)

12-18) Why does the effective resistance of a conductor increase when it is wound into a coil?

Direct coupling; Q of parallel-loaded circuit; tapped circuits; transformer coupling; transformer coupling, one tuned circuit; transformer coupling, two tuned circuits; effect of Q on coupling.

Radio-Frequency Coupling

THIS CHAPTER deals with the transfer of power from one point to another in a radio-frequency circuit, particularly in those cases where impedance matching in either of the two senses discussed in Chapter 10 is required.

DIRECT COUPLING: When a load has approximately the right value of impedance to be connected directly to the source, no other impedance matching is needed; the only requirement at radio frequencies is that the effects of the inherent reactances of the source and load be minimized.

Coupling ofthis type occurs in certain cases where power is to be transferred from the plate of one vacuum-tube amplifier to the grid of another. The circuit is shown in simplified form in Fig. 13-1. The power source is equivalent to an rf generator having an open-circuit voltage E and an internal resistance R_S, with its output side shunted by internal capacitance C_S. The load, R_L, has an internal capacitance C_L. C_L and C_S are in parallel and therefore add together. Their total reactance, X_C, must be balanced by an equal inductive reactance, X_L, which is supplied by inductance L_1. Capacitance C_1 can be omitted if L_1 is adjustable to the proper value to resonate with $C_L + C_S$ and thus cancel their reactance at the operating frequency. L_1 often is used alone in vhf circuits.

However, by using C_1 in conjunction with L_1 it becomes possible to raise the Q of the circuit and thereby increase its selectivity. C_1 adds to the stray capacitances already present, and thus the value of L_1 required for resonance is reduced. C_1 also offers a convenient means for adjusting the circuit to resonance; i.e., for exact cancellation of reactance.

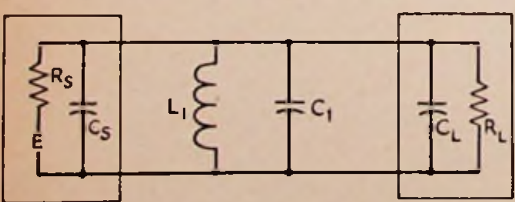


Fig. 13-1

Q OF PARALLEL-LOADED CIRCUIT: Viewed from circuit L_1C_1, R_S and R_L are in parallel and across the tuned circuit consisting of L_1, C_1, C_S, and C_L in parallel. The equivalent resistance of R_S and R_L constitutes a load *for the parallel-tuned circuit*, although the load for the *source* is simply R_L, when L_1C_1 is resonant. (This assumes that the inherent Q of the tuned circuit is high enough so that its parallel impedance is very large compared with R_L.)

The operating Q of the circuit, under these conditions, is determined by R_S and R_L in parallel. The operating Q is quite distinct from the *internal* Q of the circuit, which is determined solely by the Q of the inductance. If the total resistance of R_S and R_L in parallel is R, then on the assumption that the resonant impedance of the circuit itself is very high compared with R.

$$\text{Operating } Q = \frac{R}{X_L},$$

and the selectivity of the circuit is determined by this value of Q.

Since this relationship applies only at or close to resonance, where X_C can be considered to be equal to X_L, either X_C or X_L can be used in the equation if stray as well as intended capacitances are included in X_C. Note that under the assumed conditions (internal Q very high) the operating Q *increases* if the parallel-circuit reactance is made *smaller.*

TAPPED CIRCUITS: When the match between R_S and R_L is not good enough for adequate power transfer, a tapped circuit can be used to change R_L to a suitable value to be seen by the source. Either the inductive or capacitive branch of the circuit may be tapped. Figs. 13-2(A) and (B) show a tapped coil, while (C) and (D) are alternative circuits with divided capacitance. In (A) and (C) the load impedance, R_L, is stepped up to a higher value to be seen by the source; in (B) and (D) it is stepped down.

At and close to resonance, which is the usual operating condition, the transformed load impedance, R, in the stepped-up case is

$$R = R_L \frac{X_1{}^2}{X_{Y-Z}}$$

and for the stepped-down case

$$R = R_L \frac{X_{Y-Z}{}^2}{X_1}$$

X_1 is the reactance of L_1 in (A) and (B) and is the sum of the reactances of C_1 and C_2 in (C) and (D).

The formulas are approximate, and assume resonance along with a Q of 10 or more. The impedance curve looking into *Y-Z* has the same shape as the curve for the whole circuit, but the magnitudes at all points are reduced, as compared with the values across the whole tuned circuit.

TRANSFORMER COUPLING: The basic transformer operation at radio frequencies can be described by using Fig. 13-3. The leakage reactances are quite high in most types of transformers suitable for rf work, and the coefficient of coupling between the primary and secondary windings, L_1 and L_2, respectively, is relatively small. As a result, the turns ratio does not have the significance it has with low-frequency transformers, and the voltage induced in the secondary depends on the mutual inductance, M.

In the simplified transformer circuit at (A), voltage from the source is applied to L_1 and a voltage is induced in L_2 by means of M. A current then flows in the secondary circuit, which consists of the reactance of L_2 and the load resistance R_L. If X_{L2} is the reactance of L_2 without L_1 present, the impedance of the secondary circuit is

$$Z_2 = \sqrt{R_L{}^2 \; X_{L2}{}^2}$$

(see Chapter 8). The effect of this impedance on the primary is equivalent to a coupled impedance in series with L_1. This coupled impedance, shown in the dashed enclosure in Fig. 13-3(B), has the same numerical value of phase angle as the secondary-circuit impedance, but the sign of the reactance is reversed. The impedance relationship is

$$Z_1 = \frac{(2\pi f M)^2}{Z_2},$$

where Z_1 is the impedance introduced into the primary circuit. The quantity $2\pi fM$ is the "mutual reactance" between the reactances of the primary and secondary coils, and can be designated X_M.

The equation for transformed impedance involves both reactance and resistance, and cannot be solved by simple algebra. However, if the circuits are made resonant the reactance is eliminated, and in such cases relatively simple methods suffice for obtaining solutions that are of adequate accuracy for practical work. In the cases considered below it will be assumed that where a circuit is tuned, its *operating* Q will be 10 or more, and that the Qs of individual components will be much higher – enough so that their resistance is a negligible factor in the determination of the impedance of externally loaded circuits.

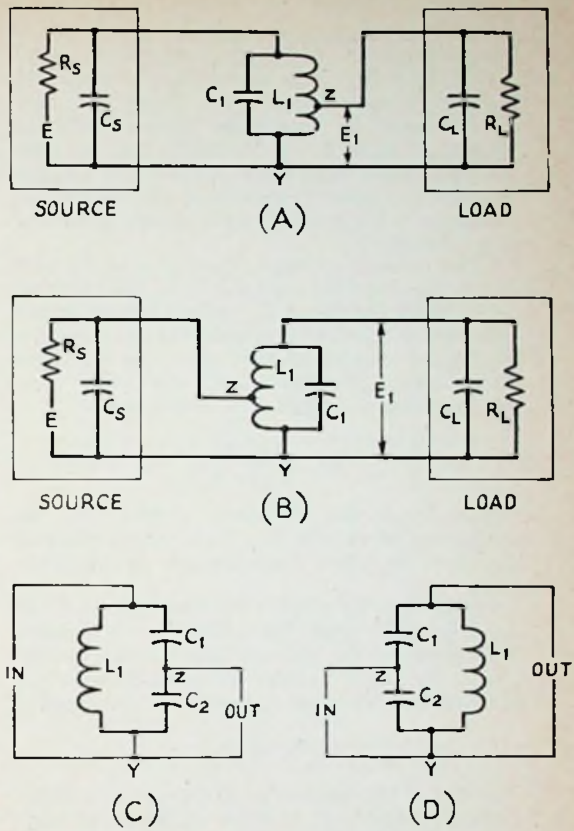


Fig. 13-2

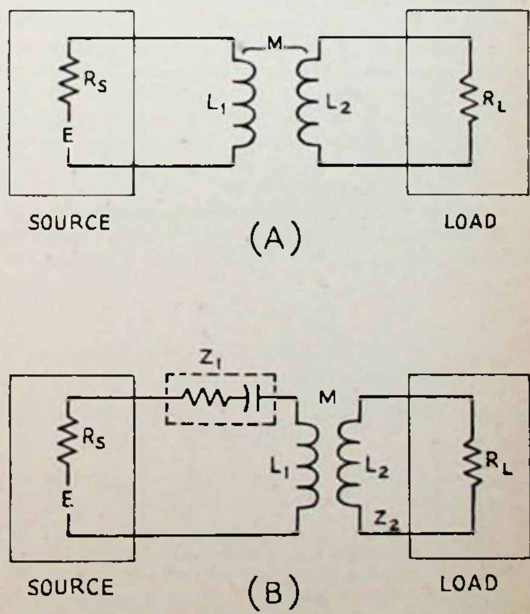


Fig. 13-3

TRANSFORMER COUPLING, ONE TUNED CIRCUIT: Two cases of single-tuned transformer coupling that arise frequently in practice are shown in Fig. 13-4. In (A) the reactance coupled into the primary (in series with L_1) from the secondary can be eliminated by tuning the primary circuit to resonance *while R_L is connected to L_2*. (The effective tuning capacitance in this circuit includes C_S and C_1 in parallel.) This leaves only the coupled resistance in the primary tuned circuit, considered as a series circuit.

The presence of stray capacitance in the load can be ignored if the reactance of C_L is 10 or more times the resistance of R_L. This is commonly the case with a low-resistance load, the usual application of this type of circuit. Under these conditions the following simplified formula will give the value of the resistance coupled into the primary:

$$R_1 = \frac{X_M{}^2 R_L}{X_2{}^2 + R_L{}^2}$$

where R_1 is the resistance coupled into the primary in series with L_1, X_2 is the reactance of L_2, and the other quantities are as previously given.

Note that R_1 is the *series* resistance, but the load for the source of power is the *parallel impedance* of the primary tuned circuit. If the primary circuit is resonant, this parallel impedance is a pure resistance (see Chapter 12) and is equal to

$$R = \frac{X^2}{R_1},$$

where R is the resistive parallel impedance of the primary circuit and X is the reactance, at resonance, of either L_1 or the total parallel capacitance. R is the load resistance offered to the source of power.

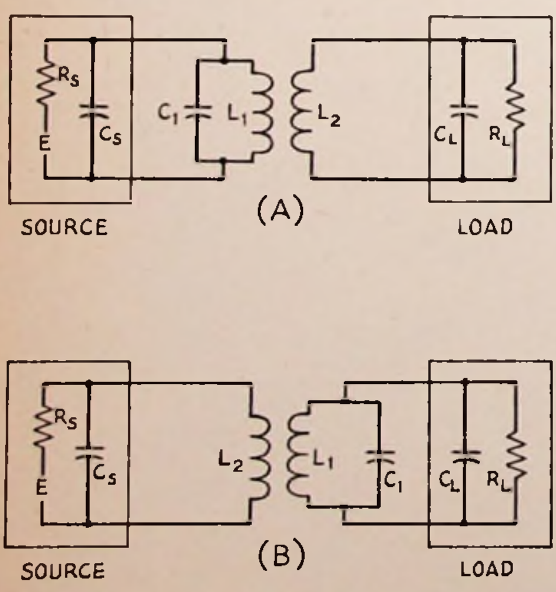


Fig. 13-4

Fig. 13-4(B) operates on the same general principles, but the tuned and untuned circuits are interchanged. Again assuming that the operating Q of the secondary circuit is to be at least 10, the reactances of L_1 and C_1 must not exceed 1/10 the value of the parallel load resistance, R_L. This limits the application of this circuit to cases where R_L is high enough so that the required values of L_1 and C_1 can be realized with practicable components.

To find the effect of the secondary-circuit load on the primary, R_L must be converted to its equivalent resistance R'_L in *series* with L_1 and C_1. The equivalent series resistance is

$$R'_L = \frac{R_L}{X^2},$$

where X is again the reactance of either L_1 or C_1 at resonance. The total series reactance of the secondary circuit is zero at resonance, which simplifies the equation given earlier to

$$R_1 = \frac{X^2_M}{R'_L},$$

where R_1 is the resistance coupled into the primary circuit. The *impedance* of the primary circuit, which is also the load on the source, consists of R_1 in series with the reactance of L_2 in this case, and is not a pure resistance as in Fig. 13-4(A). (It can be made resistive, or nearly so, by detuning L_1C_1 so that capacitive reactance is coupled into the primary circuit in series with L_1 to neutralize L_1's inductive reactance. Although the formula just given for coupled resistance is no longer quite accurate when this is done, the error is not ordinarily serious in practical cases.)

TRANSFORMER COUPLING, TWO TUNED CIRCUITS: The double-tuned transformer is shown in Fig. 13-5(A) in the version suitable for coupling between relatively high values of source and load resistance. For coupling between a low-impedance source and a low-impedance load the series version in Fig. 13-5(B) can be used. Also, one circuit may be series tuned while the other is parallel tuned, if the associated source and load impedances require it. As before, the operating Qs of the circuits are assumed to be at least 10.

The method of calculating the performance of these circuits closely parallels the methods just discussed. For example, in (A) the parallel load resistance, R_L, should first be converted to the equivalent resistance R'_L in series with L_2C_2, and then used in the formula for coupled *series* resistance given in connection with Fig. 13-4(B):

$$R_1 = \frac{X^2_M}{R'_L}.$$

R_1 is then used for finding the resonant *parallel* impedance, R, of L_1C_1; this is the load placed on the source by R_L through the double-tuned circuit. In (B) it is unnecessary to make the parallel-to-series conversion in the primary circuit, because the coupled resistance, R_1, is the load for the source of power. Reactances disappear since the circuits are series-tuned to resonance.

Fig. 13-5(A), where the impedance levels and practical components permit its use, has the advantage that the stray capacitances C_S and C_L

can be "absorbed" in the tuning capacitances, C_1 and C_2, by adjusting the latter so that the total capacitances in primary and secondary, respectively, tune the circuits to resonance.

EFFECT OF Q ON COUPLING: The resistance coupled into the primary circuit represents the power-absorbing effect of the coupled secondary circuit. The power taken from the source and transferred by the primary to the secondary is

$$P = I_1{}^2R_1 \,,$$

I_1 being the primary current and R_1 the resistance coupled into the primary. The product $I_1{}^2R_1$ must be constant for constant power transfer. Also, for constant power the source must see a constant value of load resistance.

In Fig. 13-4(A) the primary is a parallel-resonant circuit, L_1C_1, in which (assuming negligible loss in the circuit itself) the following relationship holds:

$$R_1 = \frac{X^2}{R},$$

where R is the resistive parallel impedance of L_1C_1 at resonance and X is the reactance of either L_1 or C_1. If X is changed by some factor n while R is to be held constant, R_1 must change in proportion to n^2; e.g., if n is 1/2, R_1 will have to be reduced to 1/4 its original value. Since the power input is constant with constant R, the current I_1 will automatically change (by a factor of 2 in the example) to maintain the $I_1{}^2R_1$ product constant. As the original value of R_1 was produced by some value of mutual inductance, M, M must be changed by the factor n (from the relationships previously given) to produce the required new value of resistance, n^2R_1.

Changing X by the factor n, without changing anything else, will reduce the mutual inductance by $\sqrt{n}$, but this is not enough of a change. There has to be a further change by $\sqrt{n}$ again, in order to couple the required new value of resistance into the primary. This can be brought about by changing either the coefficient of coupling or the secondary inductance, or a combination of both. In other words, the coupling between the primary and secondary has to be changed in order to maintain the same load resistance on the source, when the operating Q of the primary is changed.

If the secondary is parallel-tuned, as in Fig. 13-4(B), a change in the operating Q of the secondary circuit changes its equivalent series resistance, R_L, in the same way as in the case of the primary circuit just discussed. By the relationship given earlier ($R_1 = X_M{}^2/R'_L$) the resistance R_1 coupled into the primary will also change under these conditions, and the circuit constants must be modified so that R_1 will remain unchanged and thus place the same load on the source of power. By reasoning similar to that in the tuned-primary case, the mutual inductance must be changed by the same factor, n, by which the inductance of the secondary coil is changed.

When the primary and secondary circuits both are tuned, Fig. 13-5, changes in both circuits must be taken into account. If the primary inductance is changed by a factor n_1 and the secondary inductance by n_2, the overall change in mutual inductance required is n_1n_2. The approximation is good if the individual operating Qs are 10 or more and the circuits are tuned to resonance.

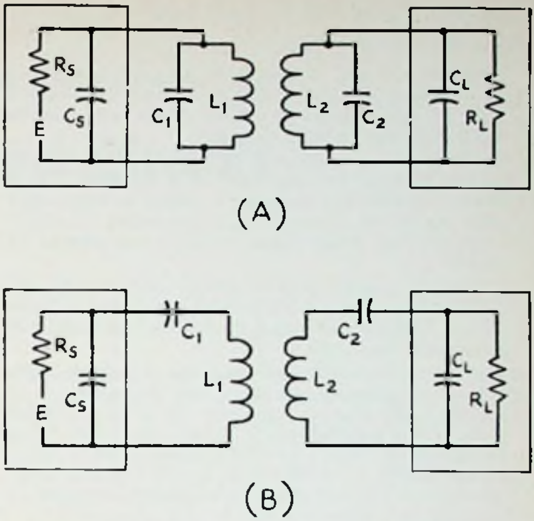


Fig. 13-5

Since the operating Q is proportional to 1/n, it becomes apparent that the higher the operating Qs of the tuned circuits, the smaller the mutual inductance required. That is, the coupling will be "looser" for the same power transfer. Also, two tuned circuits can be much more loosely coupled than a single one coupled to an untuned circuit.

In the series-tuned case, Fig. 13-5(B), the value of R_1 required for placing a given load on the source remains fixed regardless of the primary Q. Likewise, R_L is constant no matter what the value of the secondary Q. In this case the mutual inductance must remain constant for the same power transfer, and the coefficient of coupling between the two coils has to be changed so that this condition is met when the values of L_1 or L_2, or both, are changed, to make a change in the Q. If the inductance of only one circuit is changed, by a factor n, the new coefficient of coupling required is

$$k' = k\sqrt{n}$$

k being the original coefficient. When the inductances in *both* circuits are changed, by factors n_1 and n_2 respectively, the required coefficient is

$$k' = k\sqrt{n_1n_2}.$$

The overall result is the same with either series or parallel tuning – an increase in Q must be accompanied by looser coupling if the same load resistance for the source is to be maintained. Physically, it is perhaps helpful to think in terms of the fact that there is a resonant rise in current or voltage in resonant circuits, and that the rise becomes greater when the Q is increased.

Questions and Problems

13-1) If a tuned circuit is to be used in parallel with a resistance of 10,000 ohms and the required Q is 15, what values of reactance are called for in the inductance and capacitance?

13-2) If a load resistance of 50 ohms is to be fed power through the medium of a coupled tuned circuit, should the load be in series or in parallel with the tuned circuit, in your judgment? If a Q of 10 is needed, what values of reactance should be used?

13-3) Can practicable values of reactance be obtained for a Q of 10 if the load resistance is 2000 ohms? Should the tuned circuit be connected in series with the load?

13-4) A source of power requires a load of 1500 ohms for optimum operation, and the resistance of the load which is to consume the power is 4000 ohms. Two parallel-tuned circuits are to be used for coupling. What values of reactance are required in each, if both are to have Qs of 12, and what should the mutual inductance and coefficient of coupling be?

13-5) If a 25-ohm load is to be fed power through a series-tuned circuit and the Q is to be 20, what values of inductance and capacitance are needed if the frequency is 7.2 MHz?

13-6) A load of 8000 ohms is to be fed power through a tapped tuned circuit from a source which takes 5000 ohms as an optimum load resistance. Which of the circuits in Fig. 13-2 should be used? What ratio should the reactance of the tapped portion have to the total reactance?

13-7) If it is found that a source of power is overloaded — i.e., the load resistance presented to it is too low — what should be done to correct the condition if the coupling circuit is

a) Fig. 13-4(A),
b) Fig. 13-2(D),
c) Fig. 13-5(B),
d) Fig. 13-1?

13-8) What effect does circuit Q have, in inductively coupled circuits, on the mutual inductance required?

13-9) If the Q of the tuned circuit in Fig. 13-1 is doubled, what is the effect on the load resistance offered to the source of power?

13-10) Suppose a source of power requires an optimum load resistance of 6000 ohms and the actual load resistance is 75 ohms. Which circuits, or combinations of circuits, would be practicable for transferring the power?

Chapter 14

Effect of coupling on circuit Q; selectivity of tuned circuits; tuned primary, untuned secondary; untuned primary, tuned secondary; double-tuned circuits; alternative methods of coupling.

Selectivity of Coupled Circuits

BESIDES TRANSFERRING power, coupled circuits provide much-needed selectivity in radio communication. This selectivity is dependent on the Qs of the tuned circuits; the untuned ones, or untuned sections of complete coupling circuits, contribute only more-or-less incidental effects. (For example, the reactance of C_L in Fig. 13-4(A) will decrease with frequency and thus cause some reduction in the current in R_L as the applied frequency is increased, but this effect is small compared with the selectivity of L_1C_1.)

EFFECT OF COUPLING ON CIRCUIT Q: Chapter 13 was concerned primarily with the way in which a load can be placed on a source of power. As shown there, it is the loading on the tuned circuit that determines its Q, assuming that internal resistance can be neglected. This loaded Q, in turn, determines the selectivity of the circuit.

It was explained in connection with the transformer-coupled circuit of Fig. 13-4(A) that the coupled resistance, R_1, is converted by the tuned circuit into an equivalent parallel resistance, R, which becomes the load for the power source. An alternative way of looking at it is that the tuned circuit converts R_S, which is in parallel with L_1C_1, Fig. 14-1(A), into an equivalent resistance in series with the tuned circuit. Fig. 14-1(A) is a simplified diagram in which stray capacitances, as well as the source of voltage, E, are not considered. The equivalent circuit, Fig. 14-1(B), contains R_1, the resistance coupled into L_1C_1 from the secondary, and R'_S, the series resistance equivalent to R_S in parallel with the circuit. From the relationships discussed previously, R'_S is

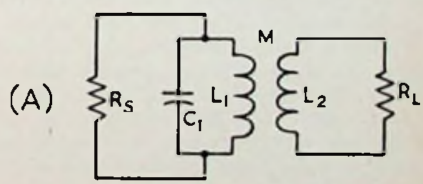
$$R'_S = \frac{R_S}{X^2},$$

where X is the reactance of L_1 (or C_1) at resonance.

Since the total series resistance is $R_1 + R'_S$ (assuming, as always, that the resistance of L_1 itself is negligible in comparison) the operating Q of the circuit is determined by their sum. If there is no coupling between L_1 and L_2, R_1 disappears and the Q is determined by R'_S alone. If the coupling between L_1 and L_2 is such as to give maximum power transfer, $R_1 + R'_S = 2R_S$ and the operating Q is just one-half its value with no coupling. If R_1 is smaller than R'_S, the Q is higher than it is when $R_1 = R'_S$, but the power transferred is less than the maximum. If R_1 is larger than R'_S (overcoupling) the power is again reduced, and because the total series resistance is larger the Q is poorer than at maximum power transfer.

The equivalent circuit is simply a tuned circuit having specified values of series resistance, and it makes no difference how power is introduced into it. Hence R_S and R_L can be interchanged in Fig. 14-1(A) – that is, the source of power and R_S could occupy the position now occupied by R_L, and the latter could be put in the place now occupied by R_S – and if the circuit constants are adjusted so that the same values of series resistance and X are maintained, the selectivity will be the same either way.

SELECTIVITY OF TUNED CIRCUITS: The inherent Qs of coils used in tuned circuits may run as high as 100 or more, although some coils are considerably poorer. However, if the equivalent series resistance *introduced* into the tuned circuit is at least 10 times as large as the resistance of the coil itself, the relationships discussed above will be sufficiently accurate for practical work.



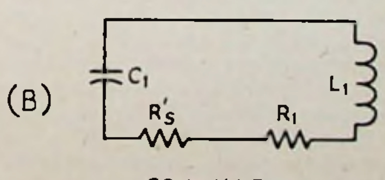


Fig. 14-1

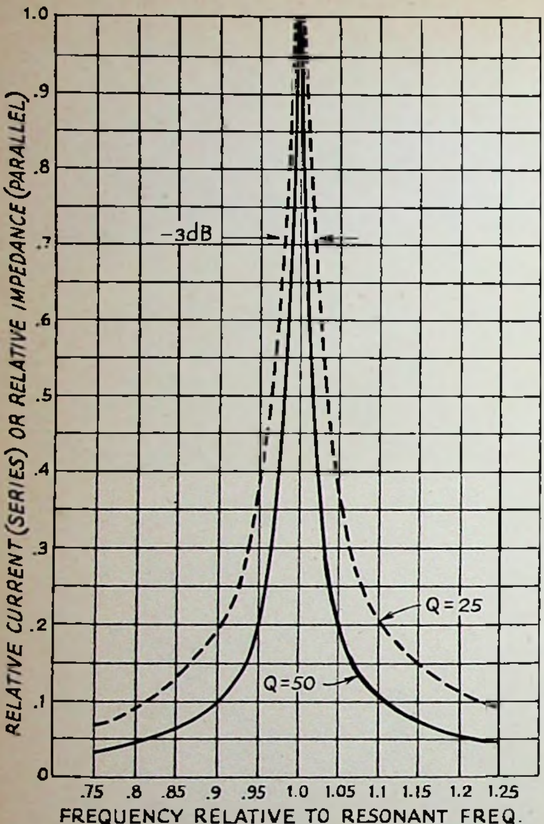


Fig. 14-2

As examples of the degree of selectivity afforded by various values of Q, Fig. 14-2 shows selectivity curves for Q = 50 and Q = 25, supplementing those shown in Fig. 12-3. However, since the audible effect of a change in signal level is approximately logarithmic (Chapter 10), curves in which the response is plotted to a logarithmic scale, which is linear in decibels, make it easier to estimate the aural effect of varying degrees of selectivity. This type of scale is used in Fig. 14-3. The four values of Q shown cover most of the practical range of operating Qs, and the response at intermediate values can be interpolated with fair accuracy.

Several practical cases are discussed below.

TUNED PRIMARY, UNTUNED SECONDARY: This circuit, a simplified version of which is shown in Fig. 14-1(A), is used principally for coupling a source having an internal resistance of a few thousand ohms to a load having a resistance of 50 to a few hundred ohms. An example is coupling between bipolar transistors. The coupling usually is designed for maximum power transfer, so that R'_S and the resistance coupled from the secondary into the primary are equal. The operating Q is therefore one-half the Q with L_2 removed.

The L/C ratios customarily used for receiving circuits will result in operating Qs that are necessarily low because R_S is not a high resistance. Consequently the selectivity of such circuits is not high. Some improvement in this respect can be made by decreasing the L/C ratio of the tuned circuit, but usually this cannot be carried vary far if the circuit must be tuned over a considerable frequency range.

An alternative is to tap the source of power down on the primary coil as shown in Fig. 14-4, in which case the primary tuned circuit is not loaded so heavily by the source and therefore has a higher operating Q. The equivalent circuit for Fig. 14-4 is the same as Fig. 14-1(B) with the understanding that the value of R'_S differs because the source of power is tapped down on the coil. The new value of R'_S is

$$R'_S = \frac{X^2_{AB}}{R_S}.$$

The operating Q of the circuit is determined by the resonant reactance, X, and the sum of R'_S and the coupled load resistance, R_1. As the tap is moved down from the top of L_1 the Q of the circuit increases in the ratio

$$\frac{X^2}{X^2_{AB}},$$

provided the mutual inductance between L_1 and L_2 is changed in order to maintain the same power

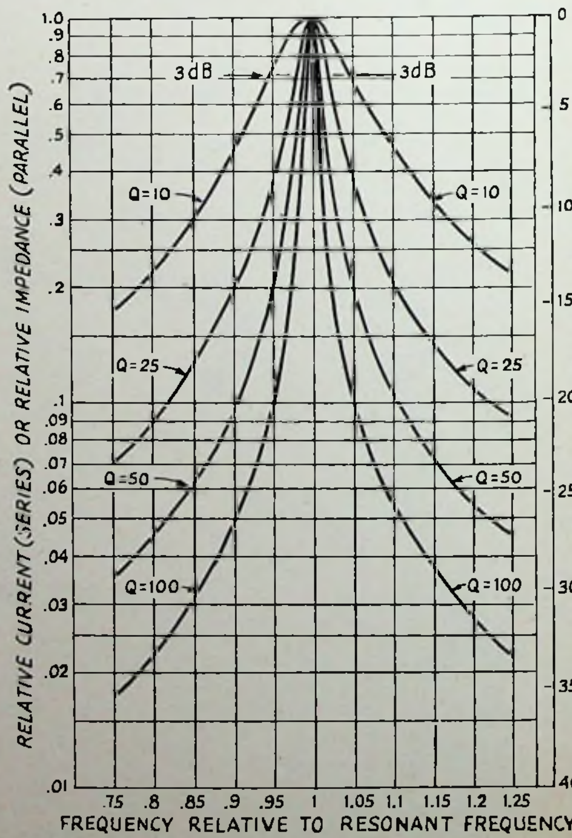


Fig. 14-3

transfer. A smaller value of X_M is required because R'_S becomes smaller as the tap is moved down to increase the Q.

UNTUNED PRIMARY, TUNED SECONDARY: This circuit, Fig. 14-5(A), is one commonly used in receiving-type tuned-rf amplifiers, C_2 being varied to resonate the secondary circuit to signals of various frequencies. As explained earlier, it does not matter which way the power moves through the equivalent circuit, so Fig. 14-5(C) is basically the same as Fig. 14-1(B). The remarks concerning Fig. 14-1(B) also apply.

The source of signal may be an antenna, transmission line, or the output circuit of a vacuum-tube or transistor amplifier. The circuit often will be working into a load which has extremely high resistance, such as the grid-cathode circuit of a tube amplifier or the gate-source circuit of a field-effect transistor. In such case R'_L is the resistance of L_2 itself; that is, the circuit is its own load. (A high resistance, 1 or 2 megohms, may be in parallel with L_2C_2 – usually through a coupling

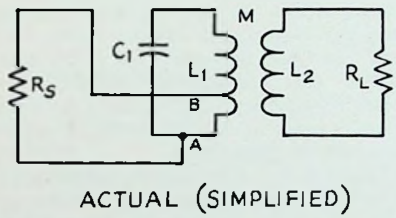


Fig. 14-4

capacitor which does not affect the rf operation but simply serves to isolate dc voltages – but the series equivalent of this resistance usually is small enough, in comparison with the coil resistance, to be neglected.)

If the source of power (in receiver circuits) is the plate circuit of a vacuum-tube pentode, R_S is very high, since the plate resistance of a tube of this type is 1/4 megohm or more. It is impossible to design circuits (with practicable components) that will match such high values of R_S, because an inductance having the required value for X_1 would have so much distributed capacitance that it would not operate as a pure inductance. In this case it is customary to use as much inductance in L_1 as the circumstances permit. Secondary output has to be sacrificed to some extent, but in compensation the operating Q is higher because the coupling cannot be tight enough to be optimum.

If the load resistance, R_L, is too low to permit desired values of operating Q, it can be tapped down on L_2 as shown in Fig. 14-5(B). The operating Q with such a circuit is increased in the same way as described for Fig. 14-4.

DOUBLE-TUNED CIRCUITS: The selectivity of two tuned circuits, independently adjusted to be resonant at the same frequency and then coupled together without changing the tuning, differs considerably from the selectivity of single-tuned circuits such as have been considered so far. The simplified circuit is shown in Fig. 14-6(A). The

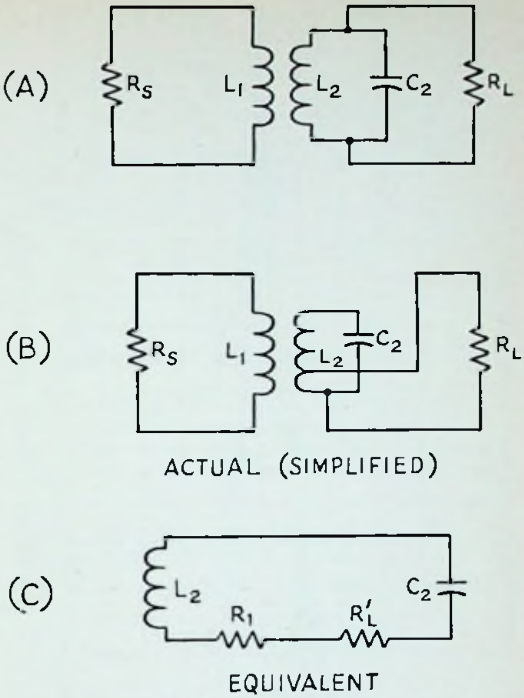


Fig. 14-5

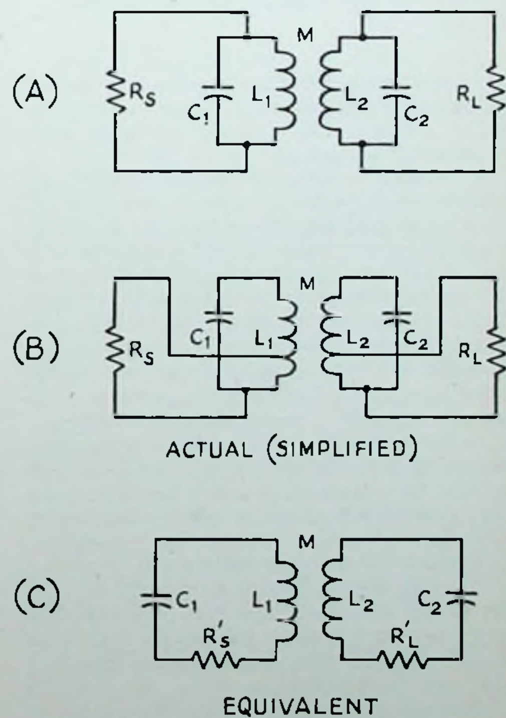


Fig. 14-6

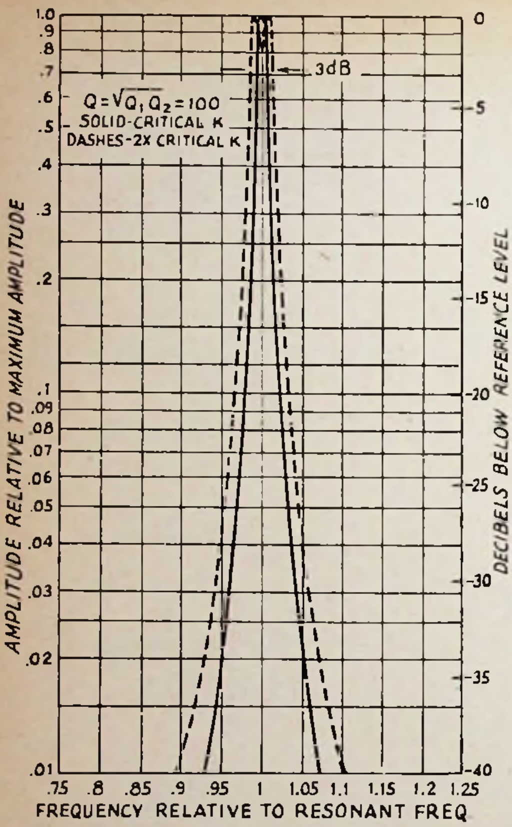


Fig. 14-7

equivalent circuit at (C) does not attempt to show the effect of coupling, but simply converts the parallel resistances R_S and R_L into their series equivalents, R'_S and R'_L.

When the two circuits are coupled together the overall frequency response depends both on the Qs of the individual circuits and on the degree of coupling between them. If the coupling is very loose, the overall response is approximately that of one circuit multiplied by the response of the other, but the output from the secondary is small. As the coupling is made tighter a higher value of resistance is coupled into the primary, and when the coupled resistance equals R'_S the power transfer is maximum. At this point – known as critical coupling – the overall selectivity curve becomes slightly flattened at the peak and the curve, as compared with the curve for a single-tuned circuit having the same Q, will show an increase in relative response at frequencies somewhat removed from resonance. This is shown by the solid curve in Fig. 14-7. This curve may be directly compared with the Q = 100 curve in Fig. 14-3, and inspection will show that near the peak it is broader, but that the "skirts" of the curve drop off more steeply as the frequency is moved away from resonance.

At critical coupling,

$$X_M = \sqrt{R'_L R'_S}.$$

If the coefficient of coupling, k, is increased beyond the critical value (overcoupling) the response at the resonant frequency will *decrease*, and the overall selectivity curve will show two "humps" or peaks, about equally spaced on either side of resonance. This is shown by the dashed curve of Fig. 14-7. The spacing between these peaks increases with increasing k, and the peaks themselves become more prominent.

The humps appear because at frequencies off resonance reactance is coupled into the primary from the secondary, as mentioned in Chapter 13. Since the coupled reactance is opposite in type to the reactance of the primary itself at the same frequency, a degree of reactance cancellation takes place. This leads to minimum values of reactance (and consequent peaks in primary current) at two frequencies determined by M and the circuit Qs. The larger primary current in turn induces a larger voltage in the secondary. At first, as M is increased beyond the critical value, the peaks are quite small, but when k is about 1.5 times critical they become distinct. In the region from critical coupling to about twice critical the overall selectivity curve will be essentially flat-topped (within a few decibels, although there will be a small dip at the resonant frequency). The substantially uniform response over a small frequency range is useful for transmitting signals, such as voice-modulated ones, that occupy a *band* of frequencies.

The form of the selectivity curves in the region of higher response is shown more clearly in Fig. 14-8, in which the frequency scale has been expanded by a factor of 5. The flattening at critical coupling and the double-humped response at twice critical coupling are quite evident.

If the circuit Qs are reduced the peaks, for a given ratio of actual k to critical k, occur at frequencies farther from the resonant frequency, and become less "sharp." Thus the bandwidth of the circuit is controlled by the Qs of the individual circuits; if a wider band is wanted, the operating Qs must be reduced, or if a narrower band is desired they must be increased. Since these effects are proportional to the value of the resonant frequen-

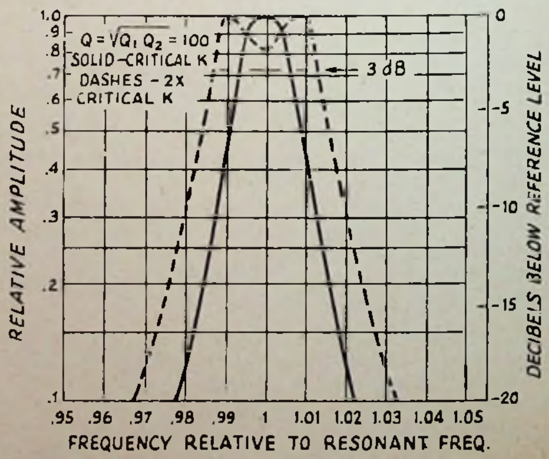
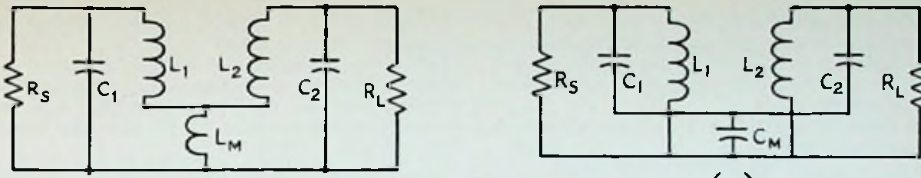
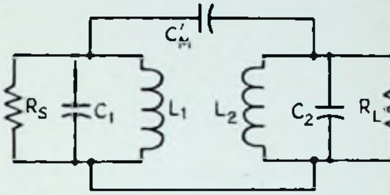


Fig. 14-8



(A) (B)

Fig. 14-9



(C)

cy, narrow bandwidths are usually associated with lower frequencies – for example, the narrow bandwidths at intermediate frequencies such as 455 kHz. Wide bandwidths can be obtained by increasing the resonant frequency as well as by control of the circuit Qs.

The curves shown are based on a number of simplifying assumptions, but are representative of practical results so long as the Qs of the primary and secondary are about equal to each other. The approximate figures in the following table are useful as a guide to the order of bandwidth that may be expected:

	$Q=25$	$Q=50$	$Q=100$
3-dB bandwidth, critical coupling	8	4	2
30-dB bandwidth, critical coupling	32	16	8
3-dB bandwidth, 2X critical coupling	10.8	5.4	2.7
30-dB bandwidth, 2X critical coupling	51.2	25.6	12.8
Peak separation, 2X critical coupling	6.8	3.4	1.7

The figures give the bandwidth in percentage of the resonant frequency.

The tapped circuits of Fig. 14-6(B) behave the same way. The difference is that the circuit Qs are established as described in connection with Figs. 14-4 and 14-5(B).

ALTERNATIVE METHODS OF COUPLING: The coupling between two tuned circuits does not have to be through mutual inductance, as in the preceding discussion, but can be by other means without changing the general principles involved. Three fairly common methods of coupling without mutual inductance are shown in Fig. 14-9. The approximate coefficient of coupling, k in each case is given by

(A) $$k = \frac{L_M}{\sqrt{L_1 L_2}}$$

(B) $$k = \frac{\sqrt{C_1 C_2}}{C_M}$$

(C) $$k = \frac{C'_M}{\sqrt{C_1 C_2}}.$$

Critical coupling occurs, as before, when the mutual reactance, X_M, is equal to $\sqrt{R'_S\ R_L}$, where the latter are the equivalent series resistances of the tuned circuits.

Figs. 14-9(B) and (C) are equivalent, but for the same degree of coupling C_M will be large compared with C_1 and C_2 while C'_M will be small. Note that in (B) C_M is in series with both C_1 and C_2, so that the capacitances effective in tuning L_1 and L_2 are smaller than C_1 and C_2, respectively. A comparable situation occurs in (A), where L_1 and L_2 are in series with L_M. Here the *sum* of L_M and L_1, or L_M and L_2, must be used in establishing the value of tuning capacitance needed for resonance.

Questions and Problems

14-1) A source of rf power can develop 20 volts across a load resistance of 25,000 ohms. Its internal resistance is 5000 ohms. Another power source having an internal resistance of 15,000 ohms can develop exactly the same voltage across the same load resistance. Which source has the greater effect on the Q of a parallel-tuned circuit which represents a parallel impedance of 25,000 ohms, or is there no difference?

14-2) What is meant by critical coupling?

14-3) If the critical coupling between two tuned circuits is exceeded, what happens to the selectivity? What happens if the coupling is *below* critical?

14-4) A single circuit tuned to 3.5 MHz is required to suppress a signal at 4 MHz by 25 dB. Approximately what value of operating Q is required, from Fig. 14-3?

14-5) Assuming that internal circuit resistance can be neglected, what value of resistance can be coupled into a tuned circuit if its operating Q is to be 25 and it consists of 75 pF in parallel with 5 μH at the resonant frequency?

14-6) Why is there a "double-humped" resonance curve when two tuned circuits are coupled so that k exceeds the critical value? If double-humped response is found to exist in a coupled circuit and only a single response peak is desired, what can be done to obtain it?

14-7) In terms of impedance matching, what does critical coupling mean?

14-8) Two 100-kHz tuned circuits are to be coupled so that the approximate 3-dB bandwidth will be 3 kHz. If twice critical coupling is to be used and the Q of each circuit is 100, what center frequency must be used?

14-9) What advantage, if any, in terms of bandwidth is there in using two coupled tuned circuits in preference to one, assuming that one tuned circuit has high-enough Q to give the required bandwidth?

14-10) When would it be necessary, or desirable, to use a tapped circuit if two tuned circuits are inductively coupled?

Chapter 15

Principles; the L network; L network design; L networks in cascade; the π network; π network design; the T network; load impedance variations.

Impedance-Matching Networks

THE MATCHING circuits described in Chapter 13 represent a group easily recognized as simple tuned circuits. Those to be considered in this chapter also utilize resonance to eliminate unwanted reactances, but the resonant circuits are not always immediately obvious.

PRINCIPLES: The basis on which such networks operate is the series/parallel equivalence described in Chapter 9, Fig. 9-7. If a reactance X_S and resistance R_S are in series, Fig. 15-1(A), there is an equivalent circuit of parallel reactance X_P and resistance R_P which will have the same impedance. Also, the phase angle will be the same in both circuits. The parallel circuit is shown at (B) in Fig. 15-1. As pointed out in Chapter 9, X_S and X_P do not have the same values, although both are alike – i.e., both are inductive or both are capacitive. Also, R_S and R_P have different values. There is therefore the possibility of transforming one resistance value into the other by making use of the equivalence.

If a load resistance is to be transformed into a resistance that will provide the desired load for a source of power, the relationship between the two resistances is very simple:

$$R_P = R_S\,(Q^2 + 1),$$

or

$$\frac{R_P}{R_S} = Q^2 + 1.$$

where Q is defined, as usual, by

$$Q = \frac{X_S}{R_S}.$$

Thus

$$X_S = QR_S.$$

The expression above for equivalent parallel resistance involves no approximations. Since the phase angle is the same in both circuits the Qs are likewise the same, so the parallel reactance is

$$X_P = \frac{R_P}{Q}.$$

Because R_P is invariably larger than R_S, R_P/R_S is always a step-up ratio. Therefore R_P is always the larger of the two resistances to be matched, whether it happens to be the actual load or the resistance that provides the optimum load for the power source.

The reactance may be either inductive or capacitive.

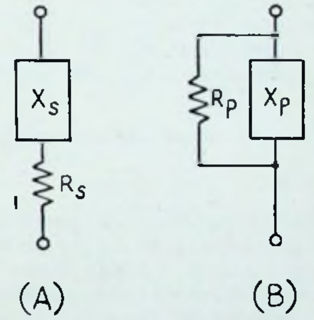


Fig. 15-1

THE L NETWORK: The series/parallel equivalent circuits are always reactive, and it is necessary to eliminate the effects of reactance in the impedance-matching circuit. This is done by connecting a reactance of equal value but of opposite type in parallel with X_P, as shown in Fig. 15-2.

This circuit is known as an L network. The L is the simplest possible impedance-matching network. It is also the most efficient one, the reason being that the operating Q has the lowest value that can possibly be used. That is, the reactance required

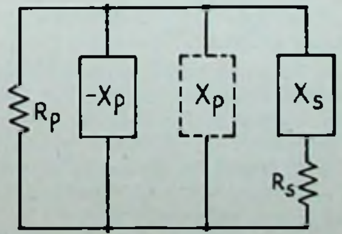


Fig. 15-2

for effecting a match has its smallest possible value, and therefore the losses in the reactances (in practical components) are minimum.

In Fig. 15-2 the dashed box which symbolizes X_P indicates that X_P is a transformed reactance, not a real component. The actual component is $-X_P$. The negative sign simply indicates that $-X_P$ and X_P are reactances of opposite types, and has no other significance

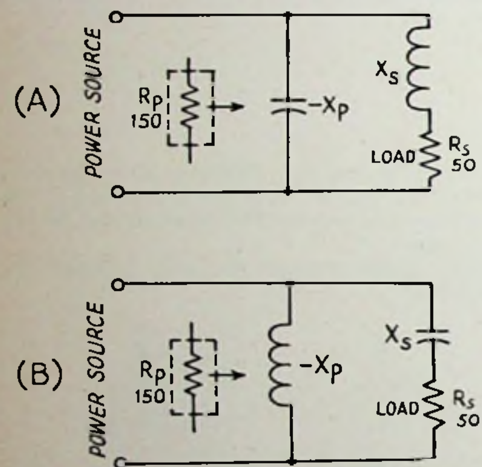


Fig. 15-3

L NETWORK DESIGN: The design of an L network to match two known resistances is simple. As an example, Fig. 15-3(A) shows a load resistance, R_S, of 50 ohms which is to be matched to a resistance R_P of 150 ohms required as a load by a power source. Then

$$\frac{R_P}{R_S} = (Q^2 + 1) = \frac{150}{50} = 3.$$

Since $Q^2 + 1 = 3$,

$$Q^2 = 3 - 1 = 2$$

and

$$Q = \sqrt{2} = 1.41.$$

Therefore

$$X_S = QR_S = 1.41 \times 50 = 70.5 \text{ ohms}$$

and

$$X_P = \frac{R_P}{Q} = \frac{150}{1.41} = 106 \text{ ohms.}$$

Since X_S is chosen to be inductive in this example, $-X_P$ will be a capacitive reactance of 106 ohms. Solving for actual circuit values requires that a definite frequency be specified, of course. For instance, if the operating frequency is to be 7250 kHz, the inductance will be 1.55 μH and the shunt capacitance, $-X_P$, will be 207 pF.

Fig. 15-3(B) matches the same two resistances, but uses the opposite types of reactance, X_S being capacitive and $-X_P$ therefore inductive. The reactance values are exactly the same, but since X_S

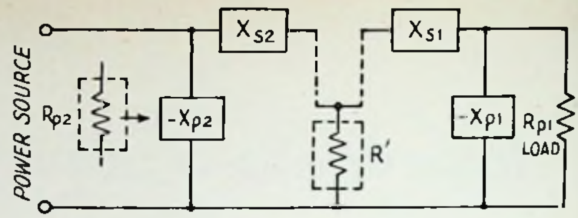


Fig. 15-4

and X_P are not equal, the values of inductance and capacitance differ from Fig. 15-3(A). For the same frequency the inductance is now 2.26 μH and the capacitance is 312 pF.

In either case R_P is a pure resistance, so the circuit is in unity-power-factor resonance (see Chapter 12). If matching requires a Q of more than 10, the three resonances described in that chapter merge to the point where, practically speaking, all three occur with equal reactance values. In such case the same inductance and capacitance can be used in either circuit.

As stated earlier, the positions of the power source and load can be interchanged if the load resistance is the larger of the two resistances.

L NETWORKS IN CASCADE; THE π NETWORK: The fact that there is no control over the operating Q of the L network may be a disadvantage when the Q is low and selectivity is a consideration. Putting two L networks in cascade allows a great deal more freedom in design because introducing a third – "virtual" – resistance permits the Q of one of the two networks to be chosen at will. A variety of combinations is possible, depending on the types of reactance chosen for the circuit elements.

A cascade using two L networks is shown in Fig. 15-4. Network 1 transforms the actual load resistance, R_{P1}, into a virtual resistance R' by means of X_{S1} and $-X_{P1}$. Network 2 then transforms R' into R_{P2} using X_{S2} and $-X_{P2}$. This effects an overall match between R_{P1} and R_{P2}. R' does not appear physically in the circuit, just as the resistance the power source sees does not appear physically in any of the circuits discussed.

R' is in the series arms of both L networks, and so must be smaller than either R_{P1} or R_{P2}. Its value is usually chosen so that some desired value of Q will be used in the network connected to the power source. X_{S1} and X_{S2} are in series and may

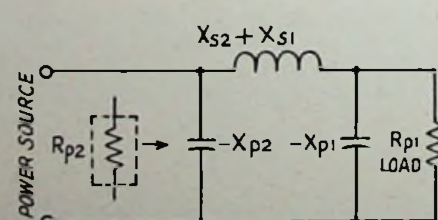


Fig. 15-5

be replaced by a single reactance equal to their sum.

If inductive reactance is chosen for both series arms, the actual circuit will be as shown in Fig. 15-5. This is one form of π network, so called because the arrangement of the reactances resembles the Greek letter π. Fig. 15-5 is commonly used for the output circuits of transmitting power amplifiers because it permits using a reasonably high value of Q in the input L network to reduce harmonic output. Harmonic reduction is further enhanced by the additional selectivity provided by the output L network.

π **NETWORK DESIGN:** The procedure for designing a π network includes 1) selecting a value of Q (Q_2) for the input L network; 2) determining the values of X_{P2} and R' from the selected Q value; 3) finding the value of series reactance, X_{S2}, required in the input L, and 4) finding the value of series and parallel reactance required in the output L for matching R_{P1}. (Note: if Steps 1 and 2 lead to a value of R' that is higher than R_{P1}, a suitable value of R' must be selected and the Q of the input L calculated accordingly so that the smaller value of R' will be matched to the required load, R_{P2}.)

As a practical example, assume that a transmitting amplifier requires a load, R_{P1}, of 1200 ohms, and that this is to be matched to an actual load R_{P2} of 50 ohms. Then if a Q of 10 is selected the required parallel reactance, $-X_{P2}$, is immediately determined:

$$-X_{P2} = \frac{R_{P2}}{Q_2} = \frac{1200}{10} = 120 \text{ ohms.}$$

By rearrangement of the equation given earlier,

$$R' = \frac{R_{P2}}{Q_2^2 + 1} = \frac{1200}{100 + 1} = 11.9 \text{ ohms}$$

and

$$X_{S2} = Q_2R' = 10R' = 10 \times 11.9 = 119 \text{ ohms.}$$

The output L follows the design procedure outlined earlier:

$$\frac{R_{P1}}{R'} = \frac{50}{11.9} = Q^2{}_1 + 1 = 4.2$$

$$Q^2{}_1 = 4.2 - 1 = 3.2$$

$$Q_1 = \sqrt{3.2} = 1.79.$$

From this,

$$X_{S1} = Q_1R' = 1.79 \times 11.9 = 21.5 \text{ ohms,}$$

and

$$X_{P1} = \frac{R_{P1}}{Q_1} = \frac{50}{1.79} = 27.9 \text{ ohms.}$$

The design is completed by adding X_{S1} and X_{S2} algebraically to find the total series reactance (here the sum is 21.5 + 119 = 140.5 ohms) and then converting the reactances to inductance and capacitance at the operating frequency. Note that because the value of Q_2 was chosen to be 10, X_{P2} and X_{S2} are very nearly equal.

If the reactances in the series arms are of opposite types, which might be the case if the circuit requirements make it desirable, the algebraic sum is their *difference*. In such case X_{P1} and X_{P2} would be of opposite types, since the series and parallel reactances in each L must differ in type.

More than two L networks can be placed in cascade. If an L follows a π, the network is a π-L; if two πs (four Ls) are used the network is a double π. The design procedure follows the same principles, taking one L network at a time, but there is increased freedom in selecting Qs and values of R' when more Ls are added. The more complicated networks are useful when the selectivity must be increased, or when practical considerations dictate selection of component values which may not be readily available in the sizes required by the simpler networks.

THE T NETWORK: In the π network, Fig. 15-4, the two series arms meet in a common virtual resistance, R'. If the two Ls are connected so that their *parallel* arms meet, the overall circuit is a T network, Fig. 15-6. In this case the actual load resistance, R_{S1}, is stepped *up* to a virtual resistance R' in Network 1. R' is then the load for Network 2 and is stepped down to a resistance R_{S2} which the

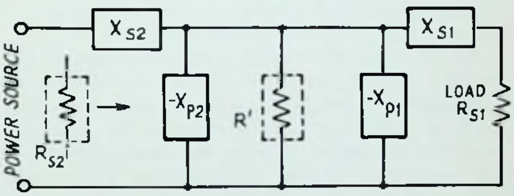


Fig. 15-6

source of power wants to see. The overall circuit therefore matches R_{S1} to R_{S2} (in contrast to the π network, which matched R_{P1} to R_{P2}).

The type of reactance to be used in each L network may be chosen at will, as before, with the usual proviso that the reactance added to compensate must be of the opposite type. In this case $-X_{P1}$ and $-X_{P2}$ are in parallel and may be combined into one. The resultant is found as described for two reactances in parallel in Chapter 7. They will not be of the same kind, if X_{S1} and X_{S2} are of opposite kinds.

Fig. 15-7 is one practical form of the circuit, in this case using inductive reactances in the series arms of the Ls. Here, the symbol $-X$ is used to indicate the resultant of the parallel compensating reactances $-X_{P1}$ and $-X_{P2}$.

Each L is designed according to the procedure given earlier. Depending on the requirements to be met in the application of the circuit, the Q of one L may be selected arbitrarily, and the necessary value of R' for that L, as well as the reactance values, may then be determined. Then, using R' as the resistance to be matched by the second network, the values for that L can be found.

Additional L networks may be added to the T, following similar principles. Since the series arms

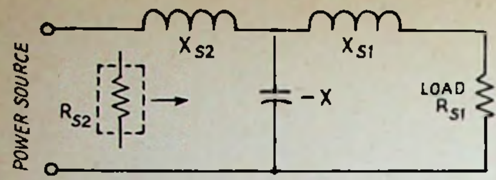


Fig. 15-7

of the T are on the outsides, the next L is likely to be connected with its series end to the end of the T. (This depends on the resistance values, but in general a series resistance will be low and a parallel resistance high in value, in relation to the reactances.) Each pair of Ls in such a case forms either a π or T, depending on the point of view.

The series reactances, X_{S1} and X_{S2}, should not be coupled except through the series connection. In particular, if they are inductances there should be no magnetic coupling between them.

BANDWIDTH: The design formulas for the L, π, and T networks considered in this chapter will apply at any frequency, since they are in terms of resistance and reactance. However, a physical circuit will match only at one frequency if the inductance and capacitance values are fixed. The reason for this, of course, is that inductive and capacitive reactances vary with frequency.

If one reactive element such as the capacitance in an L network is made variable, the change in network reactance with frequency can be cancelled (that is, resonance can be restored) by retuning. However, the match (that is, the resistance ratio of the network) cannot be restored unless the inductance also can be adjusted. This change in the matching ratio is not usually serious, in practice, if small differences in the load on the power source are tolerable.

Restoring resonance is usually indispensable for reasonably efficient operation, however, in circuits handling appreciable power. The reason is that unbalanced reactive current in the network may reduce the power factor of the load on the source to the point where source has to supply a great many volt-amperes in order to deliver a few watts. This will cause overheating in amplifying devices.

In the π network both resonance and loading can be controlled, to some extent, over a frequency range if both capacitances in Fig. 15-5 are variable and only the value of inductance is fixed. The control is possible because the circuit Qs will change along with the variation in reactance of a fixed inductance. For covering a wide band a compromise design is necessary. A suitable compromise on the inductance value may be found by first working out the network designs separately for the band-end frequencies, just as though the proper values of reactance were to be used for both frequencies. A compromise value of inductance can then be selected about midway between the two values found for the band ends. The capacitances should be variable between somewhat greater extremes than are indicated by the values at the band ends. The compromise design should be tested – and modified, if necessary – experimentally to determine whether the performance will be acceptable over the desired frequency range.

Generally speaking, there should be little difficulty in covering a bandwidth less than 5% of the mid-frequency when the capacitances are variable. Wider bands present more of a problem, especially when they approach or exceed 10%. These estimates are based on a network used in matching a 50-ohm transmission line to a required load resistance of 1000 to 2500 ohms for a transmitting power amplifier. A design value of 10 is assumed for the Q of the input L section of the π network, and a low value (of the order of 1 or 2) for the output L. The higher the Qs the narrower the bandwidth, just as in the case of the tuned circuits discussed in earlier chapters.

LOAD IMPEDANCE VARIATIONS: All of the foregoing discussion has been based on the assumption that the load will remain a pure resistance of constant value over the frequency range to be covered. In one of the applications where networks of this type are widely used, feeding power to a transmission line which is terminated in an antenna, this is seldom the case.

If the load remains largely resistive over the desired band of frequencies and does not change much in value, the variations may be small enough to be compensated for by adjustment of the variable capacitances used in the network. However, when this is not possible a match can be brought about by connecting another impedance-matching circuit between the load and the output terminals of the network.

The simplest workable circuit is generally adequate for this. For example, if there is considerable reactance but the resistive component is approximately the right value, conjugate matching (Chapter 10, Fig. 10-3) often will suffice. In other cases it becomes necessary to convert from series to parallel equivalents, or vice versa – in other words, to use yet another L network between the main network and the load. Alternatively, inductively coupled matching circuits can be used.

Questions and Problems

15-1) Given two resistances to be matched through an L network, can the Q be set at any desired value?

15-2) Describe how a T network is constructed.

15-3) An L network is to match a 15-ohm load to 2000 ohms. What values of reactance are required? If the series reactance is capacitive and the parallel reactance is inductive, what values of C and L are required for a frequency of 21 MHz?

15-4) Design a π network to match a 75-ohm load to a resistance of 2000 ohms, using a Q of 15 on the input side. Can a Q of 5 be used?

15-5) A matching network has been designed for transferring power to a 50-ohm load. How could this be used for a load consisting of a resistance of 300 ohms in series with a reactance of 175 ohms?

15-6) Design a π network to match a 50-ohm load to 1500 ohms, using an input Q of 10. What inductance and capacitance values are required for this circuit if the frequency is 4 MHz?

15-7) If the frequency in Q. 15-6 is 3.5 MHz what L and C values are required?

15-8) A T network is to be used to match 50 ohms to 300 ohms. If the Q of the L section on the 50-ohm side is to be 3, what values will the network have at 14 MHz? The series arms are to be capacitive and the shunt arm inductive.

15-9) Which is more flexible with respect to selection of Q, the π network or the T network?

15-10) Are the circuits in L, π, and T networks resonant?

Chapter 16

Types of filters; attenuation; filter sections; the half section; image impedance; attenuation in the stop band; constant-k filter design; m-derived half sections; image impedance of m-derived sections; a design example; balanced construction; general remarks; band-pass filters, form factor; mechanical resonators and filters.

Low- and High-Pass Filters

THE FUNCTION of a filter is to transmit a desired band of frequencies without attenuation and to suppress all other frequencies. The resonant circuits discussed in previous chapters all do this. However, the term "filter" is frequently reserved for those networks that transmit a desired band with little variation in output, and in which the transition from the "pass" band to the "stop" band is more rapid than is the case with simple resonant circuits.

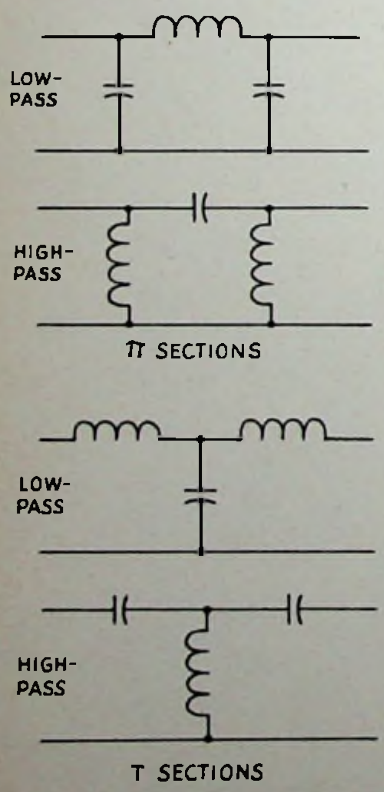


Fig. 16-1

ATTENUATION: The term attenuation refers to the ratio of the voltage applied to the filter's input terminals to the voltage that appears across its output terminals. Attenuation is usually expressed in decibels, which measure a power ratio rather than a voltage ratio; the relationship has been described in Chapter 10.

Although the use of decibels may seem to imply that more real power enters the filter than reaches the terminating resistance connected to the output terminals – in other words, that the attenuated power is dissipated in the filter – this is not the case, as explained later. The attenuation in decibels expresses the ratio of the *power that would have been dissipated in the terminating resistance had the filter not been there*, to the power that *actually* reaches the terminating resistance through the filter.

TYPES OF FILTERS: Filters are classified into three groups. A low-pass filter is one in which all frequencies below a specified frequency called the cut-off frequency are passed without attenuation. Above the cut-off frequency there is attenuation which changes with frequency in a way that is determined by the network design. A high-pass filter is just the opposite: in it there is no attenuation *above* the cut-off frequency, but attenuation does occur *below* cut-off.

The third type is the band-pass filter. In this type there are two cut-off frequencies, one at the upper-frequency edge of the band to be transmitted and one at the lower edge. Frequencies on both sides of the pass band are attenuated. An overcoupled pair of resonant circuits, both independently tuned to the same frequency (Chapter 14, Fig. 14-8), is an example of a simple band-pass circuit, although its design differs from that of a filter made according to the general principles described in this chapter.

FILTER SECTIONS: A filter section is either a π- or T-type network in which the series and shunt

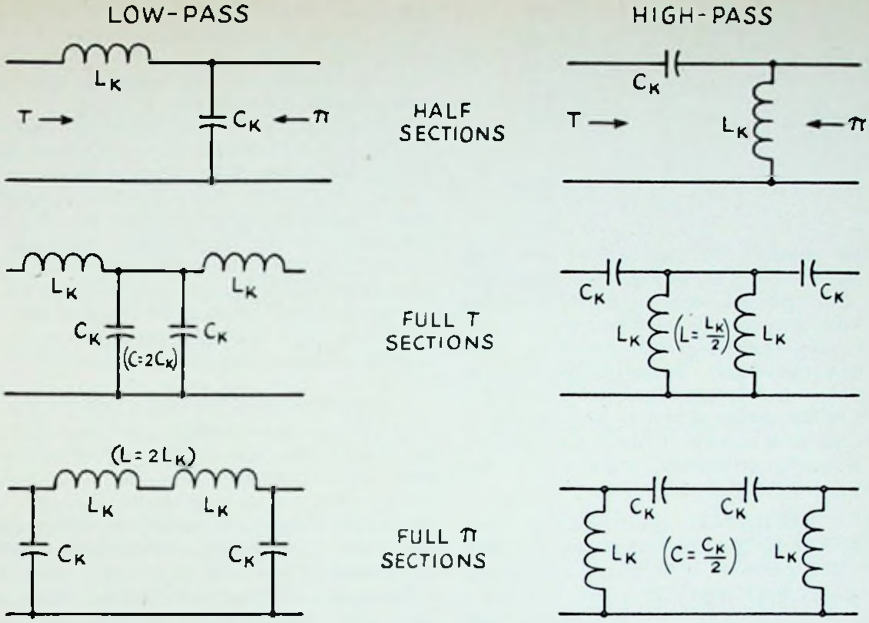


Fig. 16-2

reactances are of opposite types, as in Fig. 16-1. Unlike the π and T networks described in Chapter 15, these sections are not designed for transforming a given value of load resistance into a *different* value that will be suitable for the source of power. Instead, the design is such that the load resistance on the output side of the section will also be seen looking into the input side.

To do this the values of L and C in the section must be chosen so that the section itself will have a characteristic or image impedance (in the pass band) that equals the load resistance. When this is so, power is transferred from the source to the load without attenuation. Another consequence of the repetitive nature of the load and input impedances is that one loaded section may be used as the load for another; i.e., sections may be cascaded with no change in the input impedance shown to the source of power in the pass band.

The advantage in cascading filter sections is that in the stop band – the range of frequencies in which attenuation occurs – the attenuation increases with the number of filter sections while remaining zero in the pass band. If the attenuation of a section at a given frequency is expressed in decibels, each similar section that is added to a filter also adds the same number of decibels to the overall attenuation at the same frequency.

THE HALF SECTION: The basis on which filters are designed is the half section, shown in Fig. 16-2. The inductance and capacitance, usually designated L_k and C_k, in the half section are related to a desired value of resistance, R, which is the termination or load for the output side of the filter, by the following equations:

$$L_k = \frac{R}{2\pi f_c}$$

and

$$C_k = \frac{1}{2\pi f_c R}$$

where f_c is the frequency – the cut-off frequency – at which attenuation begins. (It is easy to remember these relationships because the reactances of L_k and C_k at the cut-off frequency both are equal to R.)

Looking into the half section from the series end shows the beginning of a T section, and looking into it from the opposite side, the shunt end, shows the beginning of a π section. The full

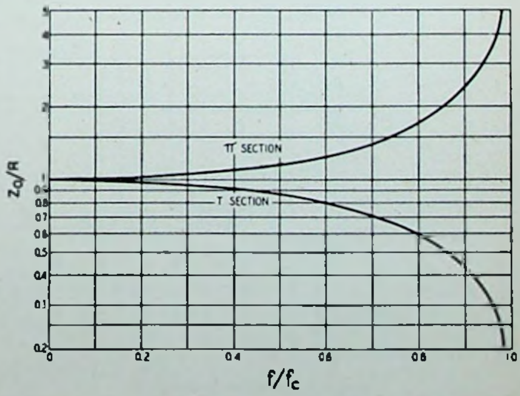


Fig. 16-3

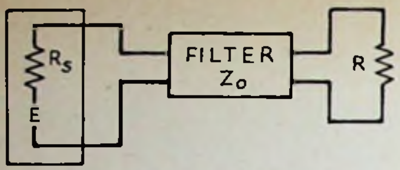


Fig. 16-4

sections are formed by connecting two half sections together as shown in Fig. 16-2. When two reactances are in parallel, as in the shunt arm of the full T sections, they may be combined into a single physical element. Similarly, where two are in series they may be combined into a single element having a value equal to the two in series, as in the full π sections. However, series elements separated by a shunt arm cannot be combined, nor is any coupling between elements permitted, other than the combinations just mentioned.

IMAGE IMPEDANCE: Unfortunately, the image impedance of a given filter section is not constant at all frequencies in the pass band, nor are the image impedances alike for π and T sections using the same L_k and C_k. The terminating impedance R is therefore a "nominal" value. The variation in image impedance in the pass band, expressed as a ratio of the image impedance, Z_o, to the nominal impedance, R, is shown in Fig. 16-3 as a function of the applied frequency, f, and the cut-off frequency, f_c.

In a practical circuit, Fig. 16-4, using a source of power having an internal resistance R_S, the terminating resistance will be chosen either to match R_S (for maximum available power) or to match some optimum value of load impedance for the source. Since the filter's image impedance, Z_o, does not match R perfectly in most of the pass band, the impedance looking into the filter from the power source will not be either R or Z_o. The input impedance that results from the mismatch cannot be determined from R and Z_o alone, but can be found for any selected frequency by analyzing the actual circuit.

In any event, the mismatch loss (see Chapter 10) is entirely between the source and the input impedance of the filter. *There is no loss of power in the filter itself.* Since the filter contains only pure reactances any power that enters it is either delivered to the terminating resistance R or returned to the source. If there is an adjustable impedance-matching network between the power source and the filter the mismatch loss can be overcome. This is usually the case when the filter follows a transmitting power amplifier.

ATTENUATION IN THE STOP BAND: In the stop band the impedances are deliberately mismatched to prevent transfer of power to the load. In this band the attenuation, with the type of filter section so far considered, increases progressively as the applied frequency is moved away from the cut-off frequency. Fig. 16-5 shows the theoretical attenuation of a single filter section, either π or T, high-pass or low-pass. The curves are based on the assumption that the load impedance equals the image impedance of the filter, a condition that cannot be fulfilled in practice because there is no way to construct a load whose impedance would vary in the ways shown in Fig. 16-3. However, the curve is reasonably accurate if the actual load is a resistance having the nominal value, R.

Note that the transition from the pass band to the stop band is rather abrupt. With an actual filter section terminated in R there will be variations in attenuation in the stop band near the cut-off frequency, and the sharp transition will be smoothed off. At frequencies somewhat removed from cut-off the real curve becomes free from

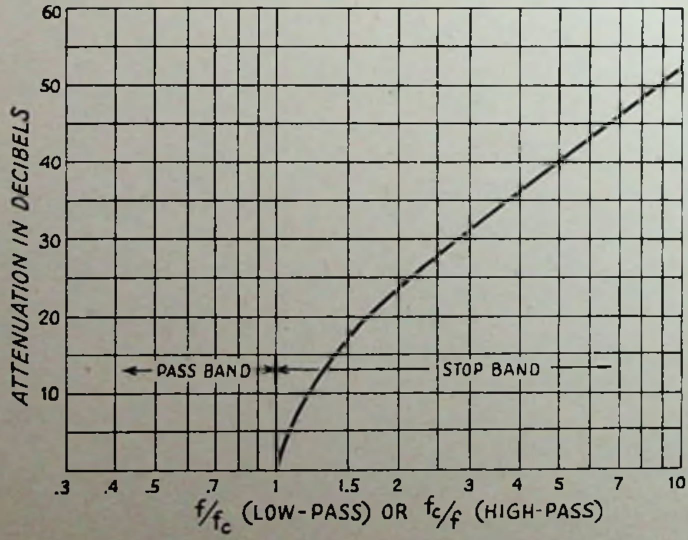


Fig. 16-5

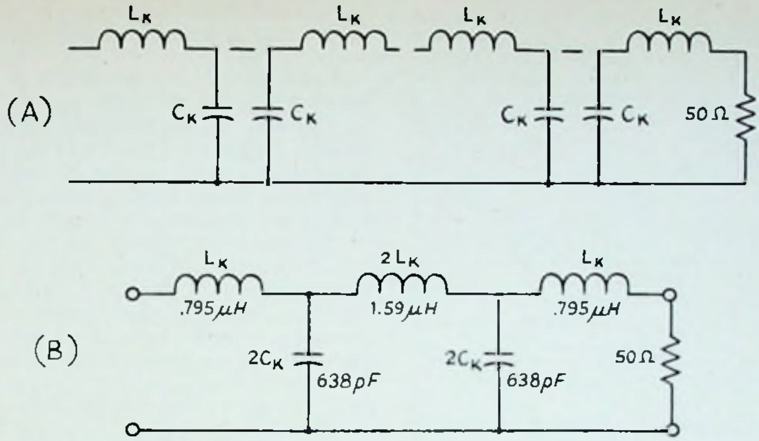


Fig. 16-6

these variations and approaches the theoretical curve. Beyond about twice the cut-off frequency in the low-pass case, and about half the cut-off frequency in the high-pass case, the attenuation increases uniformly at 12 dB per octave. (An octave is equal to a frequency ratio of 2 to 1.)

Again there is no loss of power in the filter itself. The loss is a mismatch loss between the source of power and the input impedance of the filter; i.e., the filter simply refuses to accept power from the source.

CONSTANT-k FILTER DESIGN: Filter sections of the type considered so far are known as constant-k sections. Constant-k filters are easily built up from the basic sections. As an example of design, assume that a low-pass filter using two T sections is to be built for a load impedance of 50 ohms and a cut-off frequency of 10 MHz (10×10^6 Hz, or 10^7 Hz). Then

$$L_k = \frac{50}{2\pi \times 10^7} = \frac{50}{6.28} \times 10^{-7}$$

$$= 7.95 \times 10^{-7} \text{ H or } 0.795\ \mu\text{H}.$$

and

$$C_k = \frac{1}{2\pi \times 10^7 \times 50} = \frac{1}{6.28 \times 10^7 \times 50}$$

$$= \frac{1}{314 \times 10^7} = 0.00319 \times 10^{-7} \text{ F or } 319 \text{ pF}.$$

The values also could be found, without calculation, using reactance charts, by remembering that the reactances of L_k and C_k must be 50 ohms at the cut-off frequency.

The construction of the complete filter is shown in Fig. 16-6(A) and the final component arrangement, giving the actual values of inductance

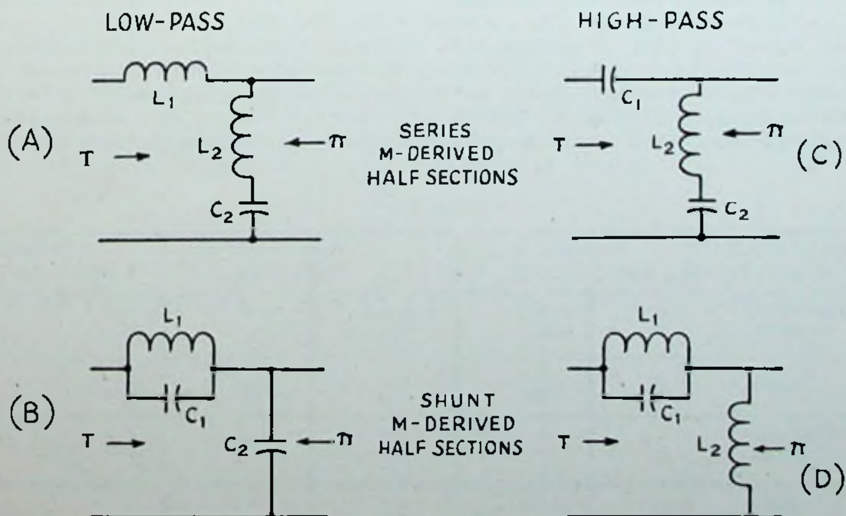


Fig. 16-7

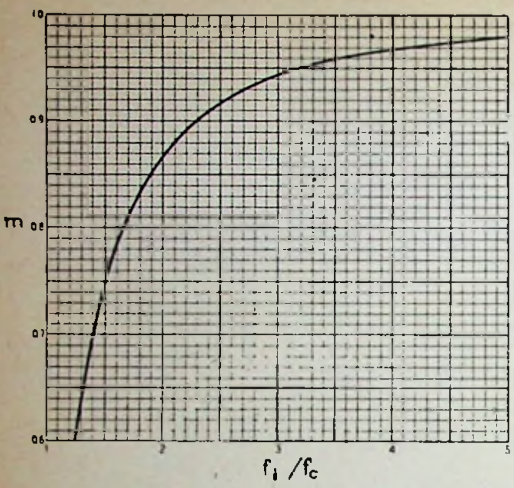


Fig. 16-8

and capacitance, is shown at B. The principles on which the filter is assembled were discussed earlier.

If the component values are within 5% of the target figures the performance would satisfy most requirements. Such a filter would be suitable, for example, for suppressing the harmonic output of a transmitter operating in the 7-MHz region, well below the cutoff frequency. The most important harmonic, the second (at 14 MHz, 1.4 × the cut-off frequency) would be attenuated about 30 dB by the two sections, as shown by Fig. 16-5.

m-DERIVED HALF SECTIONS: When it is necessary to have additional attenuation at some particular frequency in the stop band, m-derived half sections or full sections can be used instead of the constant-k type. There are two general types of m-derived sections. In the series m-derived type additional reactance is introduced in series in the shunt arm to form a circuit that is series-resonant at the frequency to be suppressed This short-circuits the output at that frequency and (theoretically) the attenuation is infinite. In the shunt m-derived type a series arm of the half section contains a circuit that is parallel-resonant at the frequency to be suppressed. As such a circuit has theoretically infinite impedance, the undesired frequency is prevented from reaching the output end of the filter.

The basic low- and high-pass m-derived sections are shown in Fig. 16-7. The π and T relationships are the same as for the constant-k half sections already discussed. Full sections, as well as multisection filters, can be formed by cascading sections as before. Also, m-derived and constant-k sections can be used in the same filter.

The factor m, which lies between 0 and 1, ensures that the image impedance of the m-derived section will match that of a constant-k section in the pass band and thus permit sections to be connected together without mismatch. To do this the values of the inductances and capacitances in the m-derived sections, Fig. 16-7, must be related to L_k and C_k, Fig. 16-2, as shown in Table 16-1.

The relationship between m and the frequency of maximum attenuation is shown by Fig. 16-8. This frequency, f_1, is given as a ratio to the cut-off frequency, f_c. For example, if the cut-off frequency of a low-pass filter is 10 MHz and it is particularly desired to suppress 14 MHz as much as possible, the frequency ratio is 14/10 = 1.4. From Fig. 16-8 the value of m for this ratio is approximately 0.7. The formula for determining m in the low-pass case is

$$m = \frac{f^2_c}{\sqrt{f^2_1}}$$

and in the high-pass case is

$$m = \frac{f^2_1}{\sqrt{f^2_c}}$$

Fig. 16-9 shows the theoretical stop-band attenuation curves for full sections with various values of m. When m = 1 the filter is a constant-k type. The pass band is not shown in Fig. 16-9 because the attenuation is zero in this band.

IMAGE IMPEDANCE OF m-DERIVED SECTIONS: The image impedance of an m-derived section is not constant over the pass band. In this it resembles the constant-k section; in fact, the variation in image impedance of a series m-derived T section is exactly the same as that of a constant-k T section for all values of m, if L_k and

Table 16-1

Type	*Fig. 16-7*	L_1	L_2	C_1	C_2
Low-pass series	A	mL_k	$\frac{1-m^2}{m}L_k$		mC_k
shunt	B	mL_k		$\frac{1-m^2}{m}C_k$	mC_k
High-pass series	C		$\frac{L_k}{m}$	$\frac{C_k}{m}$	$\frac{m}{1-m^2}C_k$
shunt	D	$\frac{m}{1-m^2}L_k$	$\frac{L_k}{m}$	$\frac{C_k}{m}$	

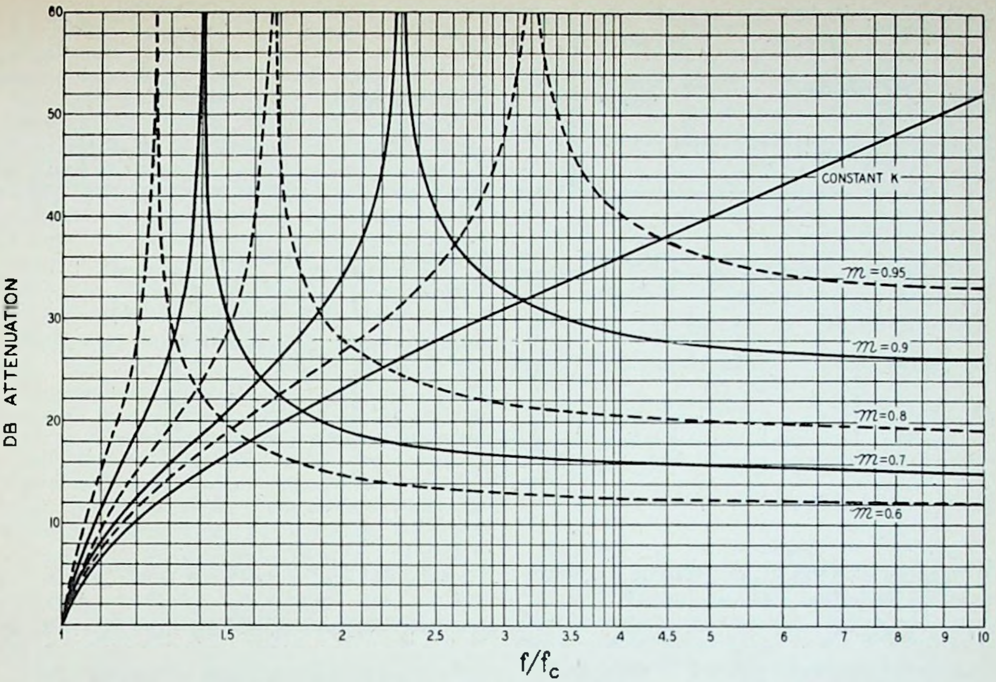


Fig. 16-9

C_k are the same in both cases. Thus series m-derived and constant-k T sections will match at all frequencies when connected together.

This identity of impedances is not true of series m-derived π sections. In these, depending on the value of m, the image impedance tends to stay closer to the nominal value, R, over much more of the pass band. In the particular case where m = 0.6, the image impedance is essentially equal to R for all pass-band frequencies except those within about 10% of the cut-off frequency.

As explained earlier, if the half section is viewed in one direction it is the beginning of a T section, and viewed from the other direction it is the beginning of a π section. The T direction can be used for maintaining a match to other full T sections that may be used in the filter, and the benefits of the smaller variation in image impedance can be obtained if the half section is viewed from the π direction from *outside* the filter. This dictates that for the best overall match to an external resistance, R, the arrangements in Fig. 16-7 as viewed from the π direction should be used as *end* half sections and that the internal sections of the filter should be T sections.

Such a filter can be constructed as shown in Fig. 16-10. A low-pass circuit is used as an example. Simply for illustration, one of the two intermediate T sections is a constant-k type and the other is m-derived. Both could be of the same type, and more sections could be used as required for a given overall attenuation. The value of m for the intermediate section or sections can be chosen at will for desired maximum-attenuation frequencies, because these sections and the adjoining Ts will always be matched.

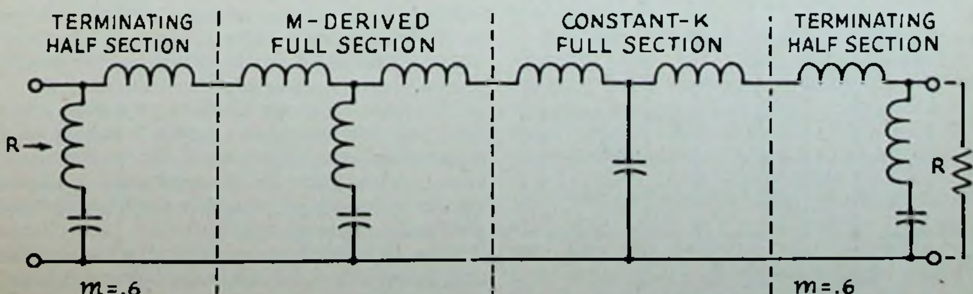


Fig. 16-10

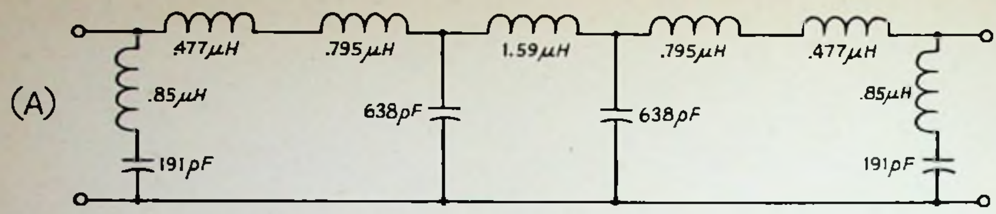


Fig. 16-11

A DESIGN EXAMPLE: To illustrate the design method for m-derived sections, suppose that the low-pass filter of Fig. 16-6 is to have m-derived half sections (with m = 0.6) added to the ends in order to improve the impedance characteristics. With m = 0.6 the ratio of the maximum-attenuation and cut-off frequencies, F_1/f_c, will be 1.25, from Fig. 16-8, and the filter also will have a sharper rise in attenuation just beyond the cut-off frequency (Fig. 16-9).

Since L_k = 0.795 μH and C_k = 319 pF, the values in the end half sections, using Table 16-1, are found to be

$$L_1 = mL_k = 0.6 \times 0.795 = 0.477\ \mu H$$

$$L_2 = \frac{1 - m^2}{m} L_k = \frac{1 - 0.36}{0.6} L_k = 1.07\ L_k = 0.85\ \mu H$$

$$C_2 = mC_k = 0.6 \times 319 = 191\ pF$$

L_2 and C_2 are series resonant at the frequency of maximum attenuation, 12.5 MHz.

The end half sections would be joined to the other two sections as shown in Fig. 16-11, following the procedure outlined earlier. The final circuit after combining series inductances is shown at (B), Fig. 16-11. Either end can be used as the input or output. In use, the filter should be terminated in a 50-ohm resistance, in which case the impedance looking into the other end would be 50 ohms over most of the pass band.

The attenuation curve can be sketched by adding the attenuations of the three full sections, using Fig. 16-9. The result is shown in Fig. 16-12. For comparison, the attenuation of the filter before adding the end half sections is shown by the dashed curve.

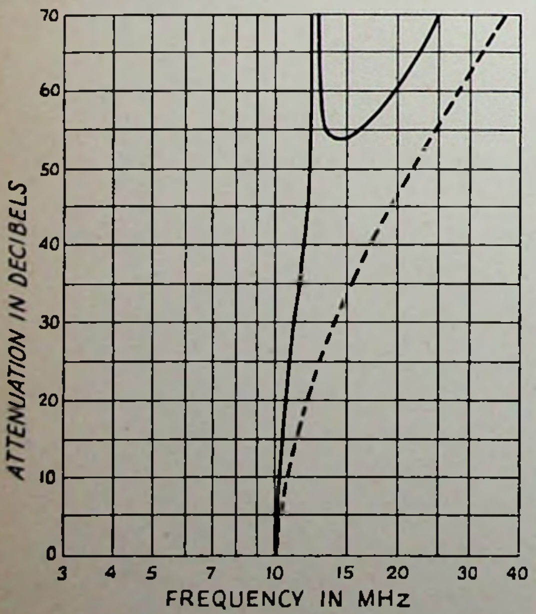


Fig. 16-12

BALANCED CONSTRUCTION: The filter circuits discussed so far in this chapter have been single-ended or unbalanced. That is, one input terminal and one output terminal have been connected directly together. This common connection usually is maintained at ground potential, so that this type of construction is suitable for working between a power source and load which themselves have one terminal grounded.

However, there are cases where a filter cannot have one side grounded, as when it is used with a transmission line that has both its conductors above ground potential. This calls for a balanced circuit arrangement, in which the filter terminals are equally above ground.

The conversion from unbalanced to balanced construction is easily made. In essence, it is done by designing an unbalanced filter by the methods previously given, but basing the design on *one-half* the required image impedance. Two such filters are then connected "back-to-back" on each side of a ground plane (which may be either an actual or virtual ground, depending on circumstances).

An example is shown in Fig. 16-13, where the unbalanced filter of Fig. 16-6(B), reproduced here in Fig. 16-13(A), is redesigned in balanced form. The filters above and below the ground connection in Fig. 16-13(B) are designed for an image impedance of 25 ohms. If an actual ground connection is not required, the shunt reactances (capacitances in this case) can be added together as in (C). The series reactances (inductances in this case) must remain at the values given in (B).

GENERAL REMARKS: The process of filter design described in this chapter is known as the image-parameter method. Its limitations with respect to constant impedance have been pointed out. Also, the performance of a practical filter will be affected to some extent by the losses in inductances and capacitances, particularly the former. However, for the applications in which the requirements with respect to such things as flatness of response in the pass band, sharpness of cutoff, and so on, are not too strict, the overall performance will be sufficiently close to the theoretical in most practical cases where low- and high-pass filters are used.

There are other methods of approach when specific requirements such as flatness of response and rapidity of attenuation beyond cutoff must be met. With the method known as modern filter design, calculations have been made with electronic computers to result in optimum selection of components. These computations have been summarized in both tabular and graphical form, but the subject of modern filter design is outside the field of the present treatment.

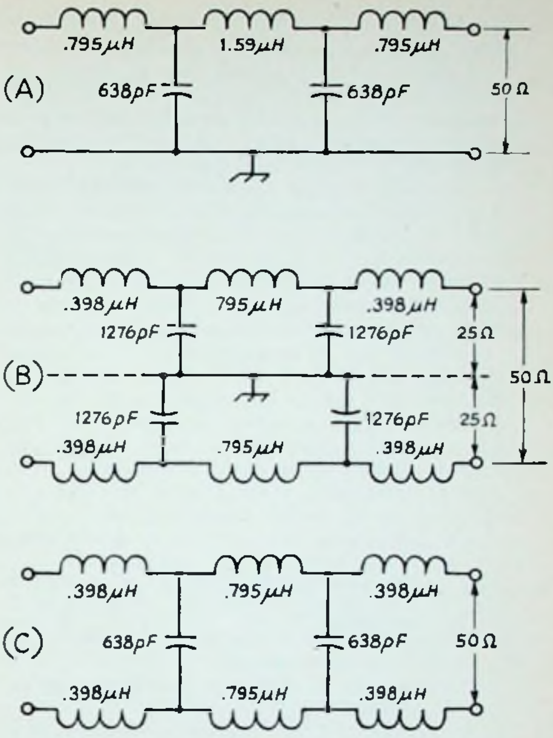


Fig. 16-13

BAND-PASS FILTERS: FORM FACTOR: Band-pass filters can be designed by the image-parameter method, but the working-out becomes considerably more complicated. Except for the simple types using coupled tuned circuits described in Chapter 14, band-pass filters consisting of inductances and capacitances are seldom used, because considerably better performance can be obtained from filters using the electromechanical devices described briefly in the next section.

A figure of merit for the selectivity of a bandpass filter is its form factor. This is the ratio of the filter's bandwidth at some specified value of *high* attenuation to its bandwidth at some *small* value of attenuation. The latter value is chosen so that a signal occupying a band of frequencies can be transmitted through the filter without undue loss at the ends of the band occupied. The ratio of the bandwidth at 60 dB attenuation to the bandwidth at 6 dB attenuation is frequently used. The smaller the form factor the greater is the ability of the filter to suppress signals that are close to, but not in, the desired pass band. Form factors of the order of 3 to 1 or less are typical of the better electromechanical filters operating in the intermediate-frequency range. LC-type filters have much higher values of form factor.

MECHANICAL RESONATORS AND FILTERS: A familiar form of mechanical tuning which corresponds in many ways to the electrical resonance in an LC circuit is found in many musical instruments. When a metal bar, for instance, is tapped or struck, the bar will vibrate at a rate or frequency which depends on its dimensions. An equivalent electrical circuit can be assumed for such a vibrator, with inductance corresponding to mechanical mass, capacitance to elasticity, and resistance to the frictional energy loss during vibration. With optimum design, including selection of material, the equivalent Q of such a mechanical resonator can be very much higher than that of an ordinary LC circuit, running 1000 or more as against a few hundred. Further, by proper dimensioning the range of vibration frequencies can be extended well up into the radio-frequency range.

To make use of the mechanical vibration it is necessary that some form of electromechanical coupling exist so that the vibrations will cause electrical voltages and currents and vice versa. The most widely used mechanical resonator is the piezo-electric crystal. A small voltage is generated at the surface of the crystal when the crystal is set into vibration, and this voltage can be taken off by means of metal electrodes on the surfaces (the electrodes frequently are plated on the surfaces). Conversely, if a voltage at the mechanical frequency of resonance is applied to the electrodes the crystal is set into vibration. The crystal vibrator thus provides its own electromechanical coupling. Both quartz and Rochelle salt crystals exhibit this piezoelectric effect in a useful degree. Quartz will vibrate at frequencies as high as the low vhf range, depending on the particular mode of vibration used (there are several possible modes). The Rochelle salt crystal is useful at audio frequencies.

A second type of mechanical vibrator is the magnetostriction resonator. This is usually a small nickel or ferrite bar or toroid which resonates in the low-frequency range (up to a few hundred kilohertz) and sets up a magnetic field varying at its vibration frequency. The coupling in this case is by means of a coil wound around the bar or toroid.

A third type in common use is a mechanical resonator consisting of metal or metallic alloy in the form of a rod, or sometimes a disk. Separate electromechanical coupling is required by this type, because it generates no electric or magnetic field. The necessary coupling is supplied by a magnetostrictive transducer coupled to the resonator. A series of rods or disks, mechanically coupled by much thinner rods or wires, forms a multiple-resonator band-pass filter. The mechanical band-pass filter is widely used at intermediate frequencies up to about 500 kHz. Band-pass filters also are constructed using magnetostrictive devices and piezoelectric crystals. With one type or another, the range from audio frequencies to well up in the hf spectrum can be covered. There are also ceramic band-pass filters, which make use of the electromechanical properties of certain ceramic materials, and are useful in the intermediate-frequency range. These do not have as good form factors as the others, however.

The design of mechanical resonators and the filters using them is beyond the scope of this book. The subject is complex and highly specialized.

Questions and Problems

16-1) A low-pass filter designed for 50 ohms has 0.02 volt across the terminating resistance at a certain frequency in the stop band. If the attenuation is 20 dB at that frequency, how much power is dissipated in the filter?

16-2) How does an m-derived filter section differ from the constant-k type?

16-3) Compute L_k and C_k for a low-pass filter having a characteristic impedance of 75 ohms and a cutoff frequency of 40 MHz.

16-4) Using the values found in Q. 16-3, determine the L and C constants for a low-pass filter consisting of a π half section, an intermediate T section, and a final π half section.

16-5) Design a balanced high-pass filter consisting of two T sections for a cutoff frequency of 45 MHz and an impedance of 300 ohms.

16-6) A 50-ohm low-pass filter having a cutoff frequency of 35 MHz is to have m-derived π half sections with $m = 0.6$ and a constant-k full T section between them. What is the circuit, and what values should be used?

16-7) What is the meaning of image impedance?

16-8) What separates the stop band of a filter from the pass band?

16-9) Why is it desirable for the cutoff frequency of a filter to differ somewhat from the nearest frequency to be transmitted in the pass band?

16-10) How does a filter differ from an ordinary tuned circuit?

Chapter 17

Frequency and wavelength; significance of wavelength in circuits; the transmission line; characteristic impedance; open- and short-circuited lines; reflection; quarter-wave lines; reactance of open and shorted lines; resonance; loading with reactance; power transfer with resonant-line circuits; Q and selectivity of resonant lines.

Transmission Lines as Circuit Elements

THE OPERATION of the circuits considered in previous chapters has been based on the assumption that energy could move from one place to another with no time delay. Actually, no form of energy can move faster than light, which travels at very nearly 300,000,000 meters per second in a vacuum.

Circuits such as have been considered earlier are called lumped-constant circuits. This chapter takes up some of the properties of circuits in which the speed at which energy moves is a definite factor.

FREQUENCY AND WAVELENGTH: In earlier chapters the electric and magnetic fields have been treated much as though they existed separately, but in actual fact they are independent only when there is no variation in either type. With alternating current, which is always changing, the two fields are coexistent, forming the electromagnetic field.

An important question is how far the electromagnetic field can travel in the time taken for one cycle of the alternating current to complete itself. If the speed of propagation is 300,000,000 meters per second, a current having a frequency of 100 Hz will have completed 100 cycles by the time the field has reached a distance of 300,000,000 meters. On the other hand, a current having a frequency of 100 MHz will have completed 100,000,000 cycles by the time its field has traveled the same distance.

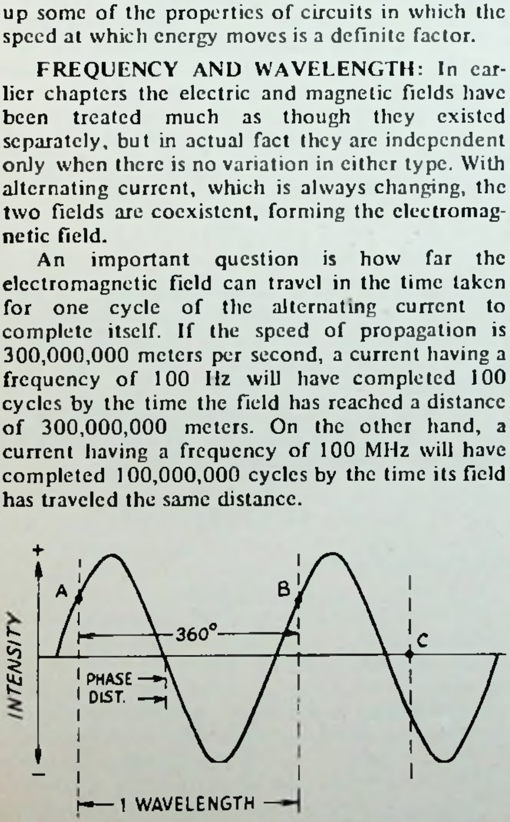


Fig. 17-1

At any given instant the field intensity will have a different value at every point on the line along which the field moves. This is shown in Fig. 17-1. If the sine curve of intensity is visualized as moving at uniform speed to the right, an observer at a fixed point such as C would see regular sinusoidal variations in intensity each cycle, indicating a wave-like motion. The distance over which the intensity variations are spread at any instant - for example, the distance from *A* to *B* in Fig. 17-1 – is called the wavelength. Wavelength and frequency are related according to the following formula:

$$\lambda = \frac{V}{f}$$

where λ is the wavelength, V is the velocity, and f is the frequency in Hz. The wavelength will be in meters, the customary measure, if V is in meters per second.

Since phase (i.e., time) and distance are directly proportional when V is constant, the phase difference between any two points along the path can be expressed as a fraction of the wavelength: 180 degrees equals 1/2 wavelength, 90 degrees equals 1/4 wavelength, and so on. The phase angle repeats every 360 degrees along the line of travel.

In the circuits considered in this chapter, V is assumed to have the same value as in space.

SIGNIFICANCE OF WAVELENGTH IN CIRCUITS: As the operating frequency is increased the dimensions of a circuit in terms of wavelength become greater. At 300 MHz, for example, a connecting lead 1/4 meter long (less than 10 inches) is 1/4 wavelength long, and there is therefore a phase difference of 90 degrees between its ends. The circuit operation described in earlier assignments has been based on the assumption that there would be no difference in phase between one end of a lead and the other, and that the field set up by a current in one circuit would have the same phase in another coupled to it, but a short distance away.

A distance of a few degrees (that is, a distance that is more than a small fraction of a wavelength) may introduce enough phase shift so that circuit

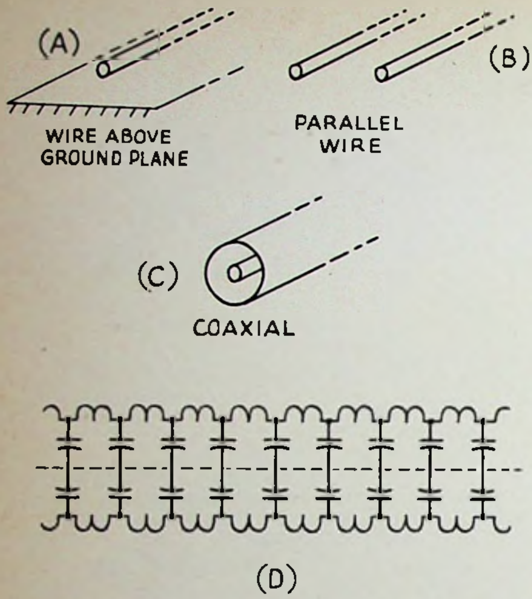


Fig. 17-2

operation differs appreciably from what would be anticipated. In some cases the phase shift can be compensated for by adjustment of the lumped constants, but if this is not feasible the lead has to be treated as a transmission line rather than as a simple connection of assumed zero resistance and zero reactance.

THE TRANSMISSION LINE: In a transmission line energy is transferred from one place to another by means of a traveling electromagnetic field. The movement of the field induces currents and voltages along the conductors, which serve to guide the energy along the desired path.

Transmission lines can take a number of forms, three of which are shown in Fig. 17-2. In the "single wire" case (a metal strip can be used instead of a round wire) the ground plane is the conducting surface of a chassis, in ordinary circuit work. There are actually two conductors, the ground plane being the second; this is necessary since the electric field lines must terminate on something that can supply an opposite charge, as discussed in Chapter 1. The parallel wire (B) and coaxial (C) lines are commonly used where power is to be transported over a considerable distance, in terms of wavelength, although they are not at all restricted to such use.

In all types of lines it is essential that there be little or no phase difference when the field set up by one conductor reaches the nearest point of the other. This means that the distance between conductors must be small in terms of wavelength. When this condition is met the line can be treated by the methods used for ordinary circuits.

Also, at any point along the line the currents in the two conductors must be equal in amplitude and flowing in opposite directions. This ensures that the fields surrounding the two conductors will be substantially equal and opposite everywhere except for a very small region near (mostly between) the conductors themselves. The opposition of the two fields eliminates electromagnetic effects in space external to the vicinity of the conductors, for all practical purposes. In the coaxial line the cancellation of fields is enhanced by the fact that the outer tube acts as a shield to prevent fields inside the line from escaping.

CHARACTERISTIC IMPEDANCE: Provided the spacing between conductors is small compared with the wavelength, the distribution of electric field lines between the conductors does not depend on frequency. This means that the capacitance per unit length of line is constant with frequency. Similarly, the shape of the magnetic fields around the conductors does not vary with frequency, so the inductance of the line per unit length is also constant with frequency. The line can therefore be visualized as being made up of a series of very tiny inductances with a tiny capacitance across the line between each pair of elemental inductances, as shown in Fig. 17-2(D). Unbalanced lines such as those at (A) and (C) would be represented by the arrangement shown on *one* side of the dashed center line in (D). A balanced line (B) would include the entire L/C circuit, with each L and C duplicated with respect to the center line. (This leads to twice as much inductance and half as much capacitance per unit length as compared with the unbalanced line at (A).)

The resemblance of such a circuit to the constant-k filter described in Chapter 16 is apparent. Like the filter, the line has a characteristic impedance. It is given by

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

where Z_o is the characteristic impedance, L is the inductance, and C the capacitance per infinitesimal unit length of line. The relative values of L and C depend on the shape and spacing of the conductors.

For lines in which the insulation or dielectric is air, the characteristic impedance of a two-wire line using round conductors is approximately

$$Z_o = 276 \log \frac{2D}{d}$$

and for coaxial line is

$$Z_o = 138 \log \frac{D}{d}$$

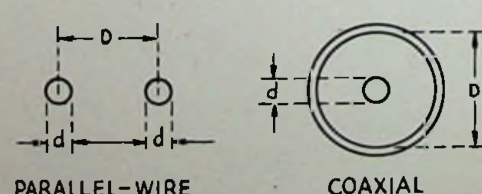


Fig. 17-3

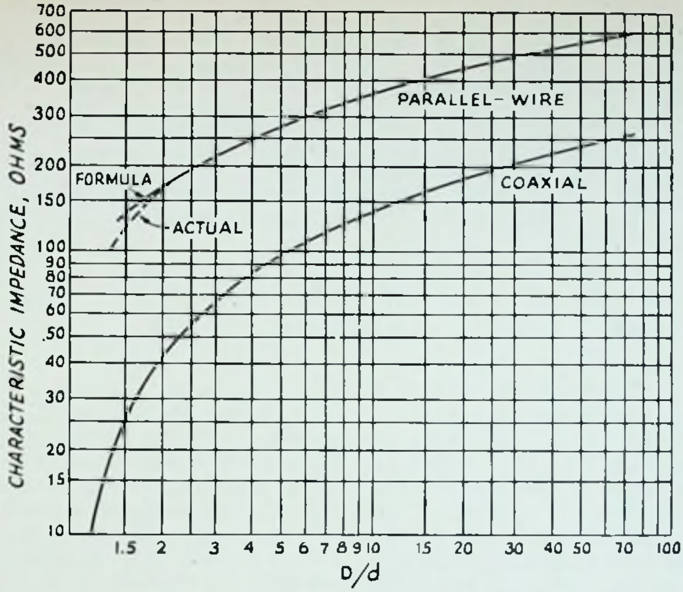


Fig. 17-4

The symbols D and d have the meanings given in Fig. 17-3. The diameters and spacings must be expressed in the same units (inches, centimeters, etc.). Note that the *inner* diameter of the outer conductor is used in the coaxial formula.

Fig. 17-4 shows the characteristic impedance graphically as a function of the D/d ratio. For the two-wire line with values of D/d of less than about 2, Z_o as given by the formula is increasingly in error, as indicated on the graph.

If the ground plane in (A) has high conductivity and is large compared with the spacing between it and the wire, it acts like an electrical mirror and forms an image of the wire just as far below the surface as the wire is above it. The impedance of such a line is equal to half the Z_o of a two-wire line having a spacing D equal to twice the actual separation between the wire and ground plane.

The characteristic impedance of a transmission line does not vary with frequency (in contrast to the filter, Chapter 16) so long as the conductor cross section and spacing are constant. If the conductor resistance can be ignored (as it can in the line circuits considered in this chapter) the line consists of pure reactances and its Z_o is a pure resistance, meaning that power can be transmitted along it without loss (again this is similar to the filter in its pass band). If the conductors and dielectric do have losses, Z_o becomes a complex impedance, and the mathematics required for expressing the various relationships is outside the sphere of simple algebra.

From a physical viewpoint, the characteristic impedance can be looked upon as a voltage/current ratio which indicates the amplitude of current that will flow when a given voltage is applied and the power is traveling in one direction along the line. For example, if the characteristic impedance is 200 ohms, the application of 100 volts to the input terminals will cause a current of 1/2 ampere. The current and voltage will travel along the line together until the far end of the line is reached. If there is a load or termination at the far end, all the power traveling outward along the line will be absorbed in the load if its resistance is equal to the line's characteristic impedance. If the termination does not match Z_o, only part of the power will be absorbed; the remainder will be reflected, as described in the following section, and the reflected part will travel back to the input end, again at the voltage/current ratio established by the characteristic impedance.

OPEN- AND SHORT-CIRCUITED LINES; REFLECTION: The limiting cases of termination or load impedance, zero and infinity, are important in line circuit operation. The first represents a short circuit at the output end of the line, while the second represents an open circuit. If the short circuit has zero resistance there can be no voltage across it no matter what value of current it may carry. On the other hand, there can be voltage across the end of an open-circuited line, but no current can flow because there is no path for it.

When an ac voltage is applied at the input end of either type of line, voltage and current are propagated along the line by the electromagnetic field at approximately the speed of light (if the dielectric is air) until the termination is reached. The power reaching the termination must *all* be reflected. In either of these special cases there is no power absorbed in the termination. It cannot leave the line, and therefore must reverse direction and travel back to the source. It goes back at the same velocity and the same voltage/current ratio because the line is the same in either direction.

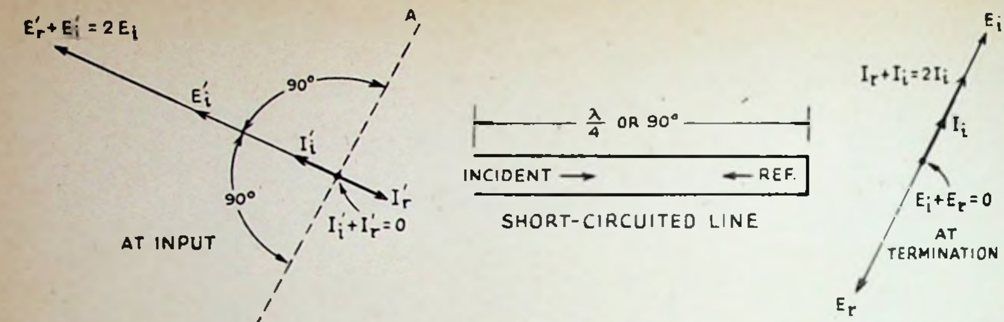


Fig. 17-5

For example, if Z_0 is 100 ohms and the applied voltage is 100, the outgoing or incident current is 1 ampere and the incident power is 100 watts. On reaching the termination, the power turns back toward the source, again at 100 volts and 1 ampere. If the termination is a short circuit the voltage across it must be zero, and can be represented by +100 volts incident and −100 volts reflected, leaving a net voltage of zero. The reflected power is therefore −100 watts, with the negative sign indicating its change in direction of flow. If the reflected power is negative and the reflected voltage is −100, the reflected current in the termination must be positive; i.e., −100 volts times +1 ampere = −100 watts. The total current in the termination is therefore 1 + 1 = 2 amperes. Similar reasoning applied to an open-circuited line shows that the reflected current must be −1 ampere (using the same figures) and the reflected voltage +100 volts, making a total current of zero and a total voltage of 200 across the termination.

These relationships are shown by phasor diagrams (at the right of the terminations) in Figs. 17-5 and 17-6. The subscripts i and r mean incident and reflected, respectively. The phasors rotate at the ac rate, going through a complete circle in one period, so the drawings represent only a single instant (here arbitrarily chosen) in the cycle.

QUARTER-WAVE LINES: If a line segment is a quarter wave long electrically, there is a time phase difference of 90° between the input and output ends, for both voltages and currents. Taking Fig. 17-5 as an illustration, if the incident-voltage phasor E_i at the termination is in the position shown at the particular instant chosen, rotating it as shown at the left will give its position E'_i at the input end. (The dashed line *AB* shows the position at the termination.) The reflected-voltage phasor E_r at the termination must also be rotated 90° to find its position E_r at the input end, but this voltage is traveling in the opposite direction so its direction of rotation is clockwise rather than counter-clockwise. The two voltages are in phase at the input end, therefore.

The incident current, I_i, is always in phase with E_i, and the reflected current, I_r, is always 180° out of phase with E_r. This being the case, the positions of the current phasors are automatically determined, and the two components of current are 180° out of phase at the input end of the shorted quarter-wave line. The input impedance is infinite because there is no current, although there is voltage.

The same considerations show that in the open-circuited case, Fig. 17-6, the voltage components at the input end are 180° out of phase and the currents are in phase. The input impedance is zero since there is current but no voltage. That is, the input impedance of a quarter-wave open-circuited line simulates a short circuit.

REACTANCE OF SHORTED AND OPEN LINES: A line segment shorter than one-quarter wavelength and having either a short circuit or open circuit as a termination has an input impedance which is a pure reactance. Fig. 17-7(A) shows a shorted line segment having a length of 60

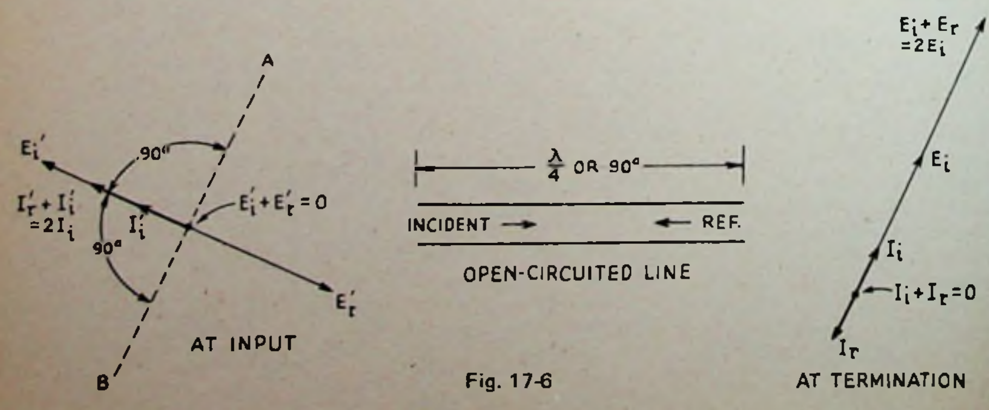


Fig. 17-6

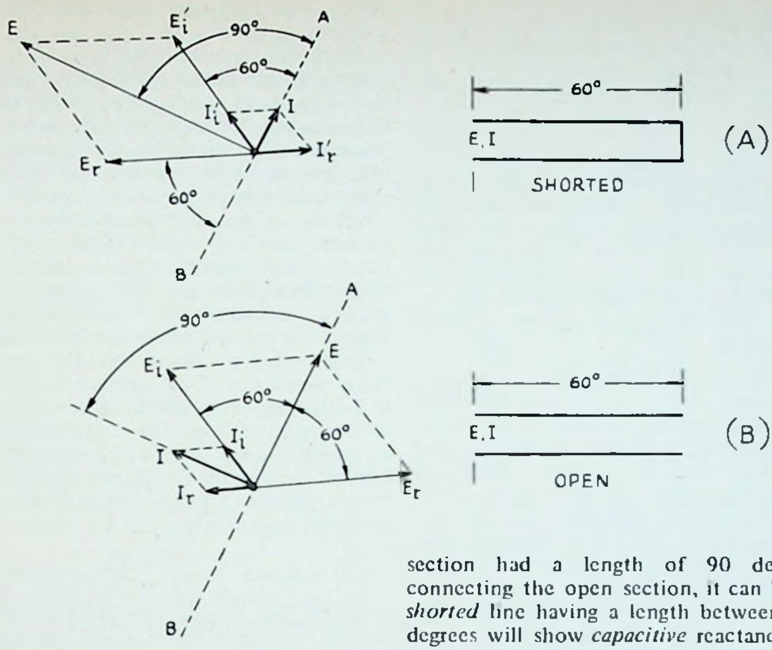


Fig. 17-7

degrees, together with the phasor diagram for the input end. (The diagram for the termination is identical with that in Fig. 17-5.) The input voltage and current components are determined in the same way as in Fig. 17-5, but using the actual line length, 60°. The two voltage components are not in phase nor are the current components 180° out of phase, so the resultant amplitudes are of intermediate value. However, the resultant current, I, is exactly 90° behind the resultant voltage, E, indicating that the line acts like a pure inductive reactance.

In the open-circuited case, (B), the resultant current I leads the voltage E by 90°, showing that this type of line, in a length shorter than a quarter wavelength, acts like a pure *capacitive* reactance.

The 90° relationship between voltage and current is true for all line lengths between 0 and 90°, as can be confirmed by constructing similar phasor diagrams for any such lengths. The value of the reactance can be calculated from

$$X_L = Z_o \tan l$$

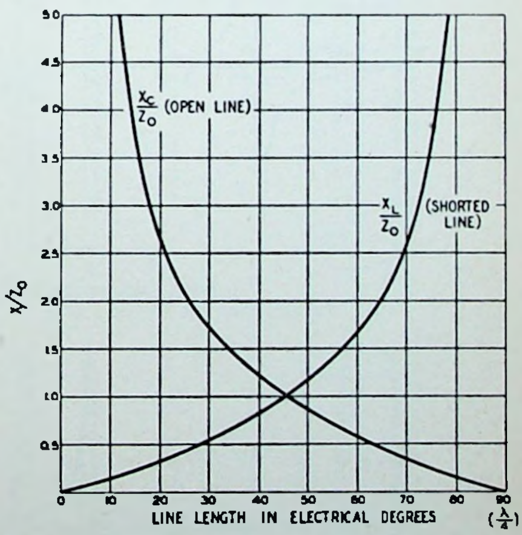
for the shorted line, and for the open line

$$X_C = Z_o \cot l,$$

where l is the line length in electrical degrees. The normalized values are shown in Fig. 17-8.

If a shorted quarter-wave section is added to the end of an open section less than 1/4 wavelength long, as in Fig. 17-9, the original voltage/current relationships in the open section will not be affected. This is because the input impedance of the shorted quarter-wave line is infinite, so connecting it across any circuit will not change the circuit conditions. Since the added section had a length of 90 degrees before connecting the open section, it can be said that a *shorted* line having a length between 90 and 180 degrees will show *capacitive* reactance at its input end.

In fact, the input reactance of any line alternates between inductive and capacitive in each quarter-wave section, measuring section lengths from the termination. If the termination is an open circuit the reactances are of the opposite type to those in the short-circuited case. Fig. 17-10 shows



Length	Short-Circuited Line	Open-Circuited Line
90 to 180 deg.	Capacitive	Inductive
180 to 270 deg.	Inductive	Capacitive
270 to 360 deg.	Capacitive	Inductive

The sign of the reactance reverses in each consecutive 90-degree (¼-wavelength) section.

Fig. 17-8

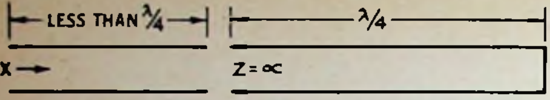


Fig. 17-9

this behavior. (The termination is at the left in these drawings.) The value of reactance in any case can be found either from the equations just given or from Fig. 17-8, simply by subtracting the appropriate multiple of 90 degrees from the actual line length, leaving a section less than 90 degrees long at the input end. The type of reactance can be determined from Fig. 17-10.

RESONANCE: There is a marked similarity between the curves of Fig. 17-10 and those for series- and parallel-resonant lumped-constant circuits, Figs. 7-5 and 7-7, Chapter 7. If a shorted line is a quarter wave long at a given frequency, lowering the applied frequency will make the same line shorter, in terms of wavelength, and the line becomes an inductive reactance. *Raising* the frequency will make the line longer, in wavelength, and the reactance is capacitive. The frequency for which the line is exactly a quarter wavelength is the resonant frequency, and since the input impedance is infinite at this frequency the shorted quarter-wave line is equivalent to a parallel-resonant LC circuit.

If the line is terminated in an open circuit, the frequency that makes it a quarter wavelength long makes it have zero input impedance. If the frequency is lowered the line shows capacitive reactance; if it is raised, the reactance becomes inductive. An open quarter-wave line is therefore equivalent to a *series*-resonant circuit.

There is an important difference between resonant lines and lumped LC circuits. In the latter there is only one resonant frequency (neglecting stray effects discussed in Chapter 12) for a given set of constants. In the line, resonance occurs for every frequency at which the length is an integral multiple of 1/4 wavelength. For any given line there is no limit to the number of such resonances if the frequency is carried high enough. There is no *lower* resonant frequency, however, than the one at which the entire line is a quarter wavelength long. If L is the physical length of the line and V is the propagation velocity (in the same length units), resonances occur when

$$f = \frac{nV}{4L},$$

n being any integer. A short-circuited line is parallel-resonant when n is odd and series-resonant when n is even. If the line is open-circuited, odd

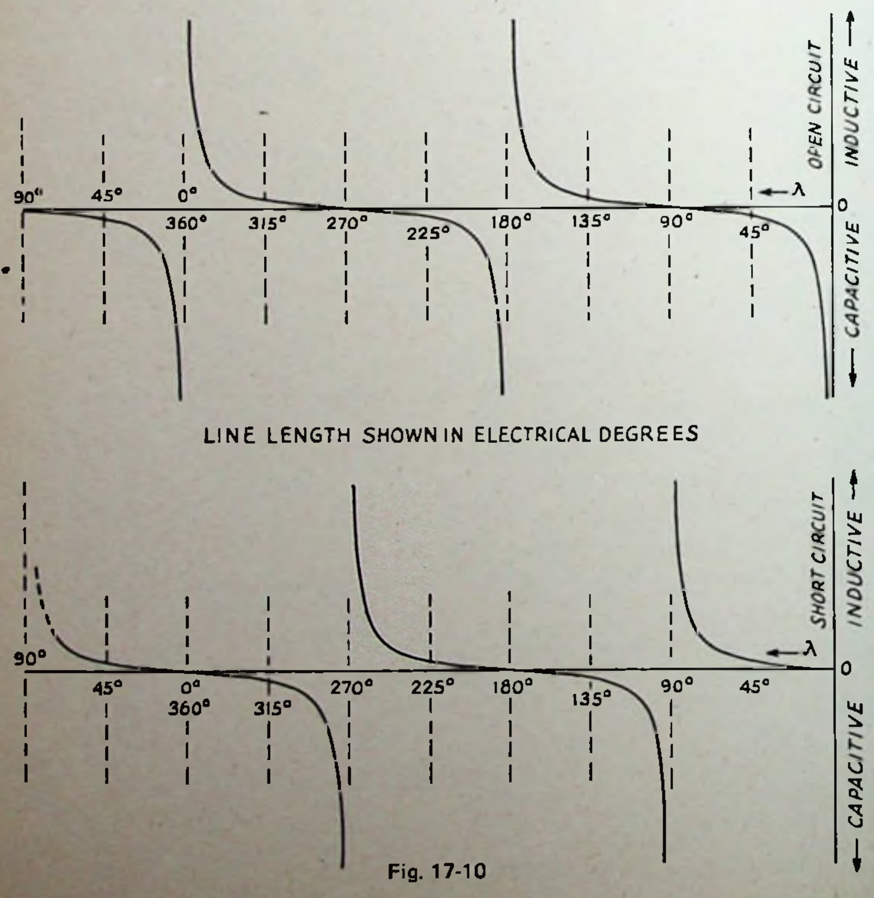


Fig. 17-10

values of n represent series resonance and even values parallel resonance.

LOADING WITH REACTANCE: Cases often arise in practice where external reactance is unavoidably across the line or is deliberately introduced. An amplifying device, for example, may have internal reactance of either type or a combination of both; or the construction of a line physically large enough to be a quarter wavelength long may be inconvenient.

Fig. 17-11 is an example of the former. The internal leads and structure give the amplifying device a shunt capacitive reactance across its output terminals. This reactance can be balanced by a line segment that shows the same numerical value but the opposite type of reactance at its input terminals, thus making the circuit as a whole resonant. Any line length, L, that will give the required reactance may be used. As explained earlier, this length, fundamentally less than 1/4 wavelength, may be made physically longer by adding sections which are integral multiples of 1/2 wavelength.

Loading with lumped constants is shown at (B) and (C). The principle is the same, that of supplying the missing reactance required to make the line segment resonant.

Loaded lines usually are adjusted to be resonant (or to a required value of reactance) by experiment. The length of the line may be varied, or, in the case of lumped-constant loading, the values of loading inductance and capacitance can be varied until the desired result is achieved.

POWER TRANSFER WITH RESONANT-LINE CIRCUITS: Quarter-wave shorted lines (including ones with loading to make them equivalent to a quarter wavelength) often are used as alternatives to coils and capacitors in vhf and uhf circuits. Since the line segments have little or no external field there is no convenient equivalent of the double-tuned inductively coupled resonant circuit, described in Chapters 13 and 14. However, the voltage along the line varies from zero at the short circuit to a maximum value at the open end, so the line can be tapped for extracting or introducing power. The principles are similar to those described for lumped LC circuits in Chapter 13, Fig. 13-2.

In circuit work the shorted end of the line is usually grounded to a chassis. This can serve as a second or "return" conductor for a single-wire connection to a source or load. Such an arrangement is shown in the coaxial case in Fig. 17-12(A). The transmission-line character of such a connection cannot be ignored, however, if the lead is more than a small fraction (perhaps 1/20) of a wavelength long.

Low-impedance inductive coupling, usually through a loop that fits inside the line as shown in Fig. 17-12(B), also can be used. The coupling loop has to be in the largely uncancelled part of the magnetic field associated with the line conductors. With the two-wire line this means that the coupling will be greatest if the loop is square, with two sides close to the opposite conductors. In the coaxial line the magnetic field fills the space between conductors and the position of the loop is not so critical.

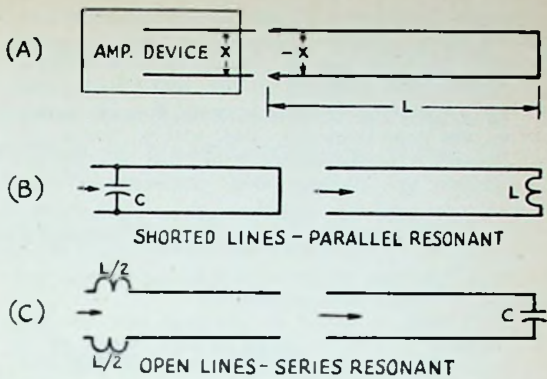


Fig. 17-11

Either type of line can have its open end connected to a high-impedance power source or load. Also, combinations of methods can be used, as required by the impedance levels to be handled.

Q AND SELECTIVITY OF RESONANT LINES: A well-constructed line, using large, high-conductivity conductors and a minimum of insulation, will have a considerably higher effective Q than an LC circuit made for the same frequency. The operating characteristics are much the same as

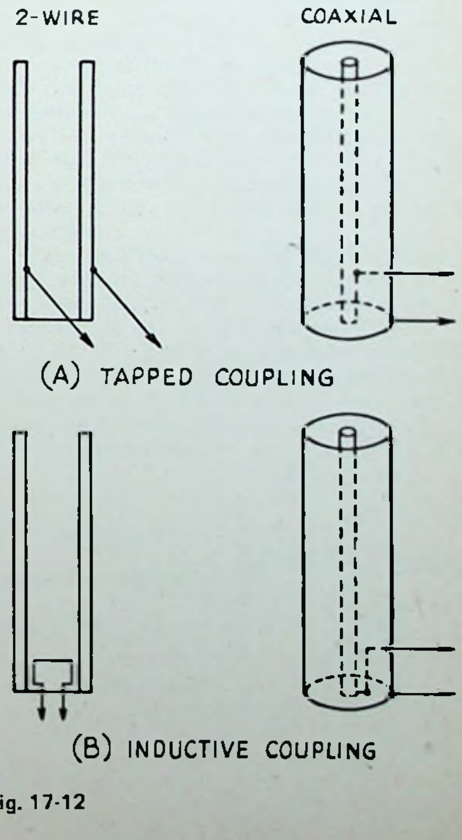


Fig. 17-12

those of lumped circuits for frequencies near resonance, although the reactances do not vary in quite the same way as they do with LC resonant circuits. The differences are more marked at frequencies considerably removed from resonance on the high-frequency side, and at twice the frequency the impedance is inverted as explained earlier. At three times the quarter-wave resonant frequency there is another resonance of the original type. In contrast, the response of the LC circuit continues to decrease along with an increase in frequency. The possibility that the multiple resonances of line circuits may lead to undesired responses must not be ignored when lines are used.

Questions and Problems

17-1) What are the input impedances of open- and short-circuited lines, respectively?

17-2) What is meant by incident and reflected voltages?

17-3) If a shorted line is 1/8 wavelength long what is its reactance, if Z_O is 80 ohms? If Z_O is 300 ohms?

17-4) With what length can an open-circuited line be made to "look like" an open circuit at its input end?

17-5) How long, in inches, is a quarter-wave line at 200 MHz?

17-6) What types of coupling to an external circuit may be used with a short-circuited quarter-wave coaxial line?

17-7) If a copper tube having an outside diameter of 3/8 inch is centered in a larger tube of the same material having 1-1/4-inch outside diameter and a wall thickness of 1/16 inch, what is the characteristic impedance of the line so formed?

17-8) An amplifier having an internal shunt capacitance of 12 pF is to be used with a shorted line section at 145 MHz. What length of 75-ohm coaxial air-insulated line is required for resonance?

17-9) How can the characteristic impedance of a two-wire line be changed, using the same conductors?

17-10) What would be the advantages and disadvantages of using a line-type resonant circuit at 29 MHz?

Chapter 18

Line terminated in a finite load; reflection coefficient; standing waves; voltage standing-wave ratio; input impedance; impedance transformation; quarter- and half-wave lines; parallel reactance, matching stubs; cancellation of reactance; frequency response; losses in practical lines; loss caused by mismatched termination; loss caused by input mismatch; operating characteristics of lossy lines; radiation from lines; wavelength in lines.

Power Transmission Over Lines

THE DISCUSSION of lines in Chapter 17 was confined to the special cases of lines that are either short-circuited or open-circuited at the output end; short segments of line are used in this way as circuit elements to replace inductors, capacitors, and resonant circuits.

The transmission of power over lines is the subject of this chapter. Lines used for this purpose may be either short or long, in terms of wavelength.

LINE TERMINATED IN A FINITE LOAD: It was mentioned in Chapter 17 that if a line is terminated in a load having an impedance equal to the line's characteristic impedance, there will be no reflection from the termination. When this is the case the input impedance of the line will be the same as its characteristic impedance. This will be true for any length of line, no matter how short or how long.

The characteristic impedance of a line that has losses of its own is complex, and the load likewise must be complex to match it. However, if line losses are small enough to be neglected, the characteristic impedance is practically a pure resistance (the line constants are practically pure reactances), and when this can be assumed to be the case the impedance relationships are much simplified. In practice, sufficiently accurate results can be obtained by assuming that the line losses are negligible, for most line lengths used at rf. Corrections can be made subsequently for power actually dissipated in the line.

Under this assumption of zero line loss, then, the load impedance required for a reflection-free termination becomes a pure resistance equal to the characteristic impedance (actually, the characteristic resistance) of the line. Such a termination will absorb all the power reaching it, and since there can be no reflected voltage or current, the rms voltage is the same everywhere along the line, and so is the current. A line operating in this way is called flat. The flat line is shown in Fig. 18-1, where R, the termination, matches Z_0. The power, P, flows in only one direction – toward the termination.

If the load (which may be either complex or purely resistive) does not match Z_0, only part of the power reaching it will be absorbed; the remainder will reverse direction and flow back to the source. The greater the mismatch between the termination and Z_0, the smaller the portion of the incident power, P_i, absorbed in the load. Thus the proportion of reflected power, P_r, increases with increasing mismatch.

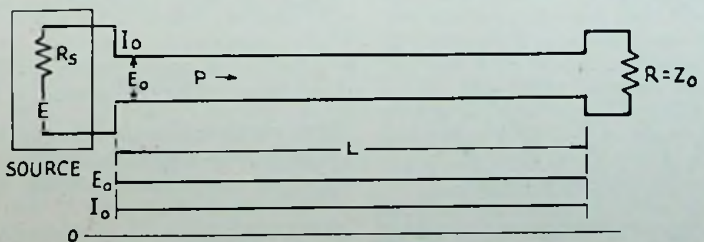


Fig. 18-1

Under steady-state conditions the *real* power entering the line and delivered to the load is equal to $P_i - P_r$ because (assuming a loss-free line – that is, P_i at the load is the same as P_i at the input end) that amount of power is all the load will accept, for a given voltage applied to the input end. *This is the total power that the source is called upon to supply.*

At the input end the reflected voltage and current combine with the incident voltage and current in a way which depends on the termination and line length. Except for the special case where $R = Z_0$, the impedance seen by the source of power is *not* equal to Z_0. Also, the input impedance of a mismatched line will, in general, contain reactance along with resistance. The reactive part simply represents the effect of the reflected voltage and current on the input impedance. The mechanism by which this comes about is similar to that operating in Figs. 17-5 and 17-6, Chapter 17, but is more complicated because of the finite resistance of the termination.

REFLECTION COEFFICIENT: When only part of the power reaching the termination is reflected, the amplitudes of the reflected voltage and current are smaller than the amplitudes of the incident voltage and current. The ratio of reflected voltage to incident voltage is called the reflection coefficient. For a loss-free line and a pure-resistance termination,

$$\frac{E_r}{E_i} = k = \frac{R - Z_0}{R + Z_0}$$

where E_r is the reflected voltage amplitude, E_i is the incident voltage, k is the reflection coefficient, R is the load resistance or termination, and Z_0 is the characteristic impedance of the line. If R is larger than Z_0, k is positive; if smaller, k is negative. The signs indicate whether the reflected voltage at the load adds to or subtracts from the incident voltage at the termination. In this respect the two cases resemble the open- and short-circuited lines, respectively, considered in Chapter 17. Open- and short-circuited lines, in fact, represent limiting cases of $R > Z_0$ and $R < Z_0$.

The reflection coefficient for current is the same as for voltage, except that the sign of the coefficient is always opposite to that of the coefficient for voltage in any given case. Ordinarily, except for the physical significance at the load, as mentioned above, the sign of the coefficient can be ignored.

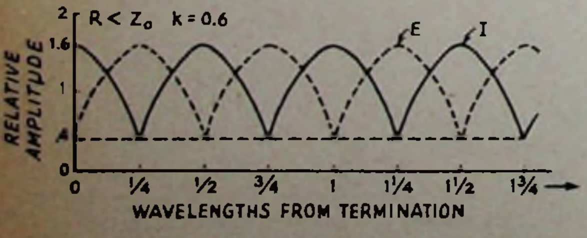


Fig. 18-2

STANDING WAVES: The incident (E_i) and reflected (E_r) components of voltage along the line combine by the same principles as in the open- and short-circuited cases discussed in Chapter 17. If R is less than Z_0 the current is larger and the voltage smaller at the load than would have been the case had R matched Z_0; in this, $R < Z_0$ resembles the short-circuited case. When R is greater than Z_0 the opposite is true, and there is a resemblance to the open-circuit case.

When $R < Z_0$ the total voltage at a distance of 1/4 wavelength from the termination is not $2E_i$ (as with a short circuit) but is something less because E_r is less than E_i ($k < 1$). Neither do the incident and reflected components of current cancel completely at this distance, because I_r is less than I_i. The case for $R < Z_0$ is shown graphically in Fig. 18-2, which plots the amplitudes of resultant current and voltage along the line for a reflection coefficient of 0.6. The voltage maxima and current minima occur simultaneously at odd multiples of 1/4 wavelength from the termination. The voltage minima and current maxima occur simultaneously at *even* multiples of 1/4 wavelength from the termination. These statements are true only when R is a pure resistance. A point of minimum current or voltage is called a node and a maximum point is called a loop.

If the voltage and current curves are interchanged – i.e., if the dashed curve represents current and the solid curve represents voltage – Fig. 18-2 also serves to show the relationships existing when R is *greater* than Z_0, for the same reflection coefficient.

If the termination is a pure resistance the maximum and minimum points always occur at exact multiples of 1/4 wavelength from the termination. A voltmeter (or ammeter) inserted in the line will measure variations similar to those shown in Fig. 18-2 as it is moved from the termination toward the source of power. The variations are called standing waves – "standing" because their positions on the line remain fixed, and "waves" because of the wavelike nature of the graphical plot of rms amplitudes. The variations appear to have half-sine-wave shape because a voltmeter or ammeter does not indicate the polarity of an alternating current but only shows absolute values.

VOLTAGE STANDING-WAVE RATIO: The ratio of maximum voltage to minimum voltage is called the voltage standing-wave ratio (VSWR) and is

$$\text{VSWR} = \frac{E_i + E_r}{E_i - E_r} = \frac{1 + k}{1 - k}.$$

The coefficient k is always considered to be positive when using this formula.

In the example illustrated by Fig. 18-2 the VSWR is 1.6/0.4, or 4 to 1. The SWR for current is the same.

The VSWR is a very useful quantity because nearly everything that needs to be known about the operating conditions on a line can be derived from it. Although direct measurement of VSWR would require going along the line with measuring instruments for a distance of at least a half wavelength, this would be inconvenient – and probably impracticable because of the difficulty of connecting measuring instruments without disturbing the operation of the line. Also, it is unnecessary, because E_i and E_r can be separated and measured at the input end of the line, and the equipment for this is simple. Once these two voltages are known the reflection coefficient, VSWR, and incident and reflected power can readily be calculated. The incident power is

$$P_i = \frac{E^2_i}{Z_o},$$

and the reflected power is

$$P_r = \frac{E^2_r}{Z_o}.$$

The power actually reaching the load is

$$P = P_i - P_r = \frac{E^2_i - E^2_r}{Z_o}$$

The relationship between VSWR and the ratio E_r/E_i, and also the ratio of reflected to incident or forward power, is shown in Fig. 18-3.

When the termination is a pure resistance, it can be shown from the equations given earlier that

$$\text{VSWR} = \frac{R}{Z_o} \text{ or } \frac{Z_o}{R}$$

This relationship is given in two forms because it is customary to express VSWR as a ratio greater than 1 to 1. The larger of the two quantities is therefore always placed in the numerator of the fraction.

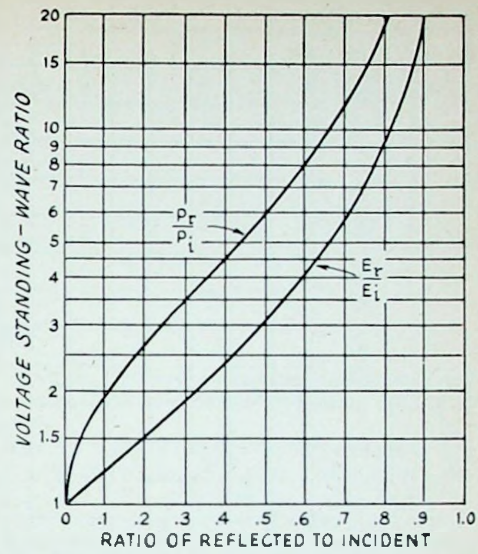


Fig. 18-3

INPUT IMPEDANCE: The impedance looking into the input end of the line from the source of power is the actual load for the source, when power is being delivered through the line to a termination. It can be inferred from Fig. 18-2 that the input impedance of a line operating with standing waves will vary with the line length (that is, with the distance in wavelengths) from the termination, because both E and I vary along the line. In the case shown, where R, the termination, is smaller than Z_o, the voltage is maximum and the current is minimum 1/4 wavelength (90 deg) from the termination; the input impedance, E/I, is therefore maximum at this point. At a distance of 1/2 wavelength (180 deg) the current is maximum and the voltage is minimum, so the input impedance is minimum and equal to R at this length. At intermediate line lengths the impedance ranges between these minimum and maximum values.

When the termination is purely resistive the input impedance is likewise purely resistive at the maximum- and minimum-impedance points, but *only* at these points. In between, the input impedance is complex. For $R < Z_o$ the type of reactance (but not the value) is the same as for a short-circuited line; for $R > Z_o$ it is the same type as for an open line. The value of the input reactance at any line length depends on the standing-wave ratio, and so does the value of the resistive component of the input impedance. Fig. 18-4 shows some typical cases. In these curves the input reactance (X_i) and input resistance (R_i) are normalized to the Z_o of the line, to make the curves universally applicable. Actual values in a given setup can be found if the normalized values shown are multiplied by Z_o.

IMPEDANCE TRANSFORMATION: Fig. 18-4 suggests that by appropriate choice of line length and Z_o a given value of load resistance can be transformed to a new value at the input end of the line. This is in fact the case, and in this respect a line section resembles the L network discussed in Chapter 15. In general, the transformed impedance will be complex, because the line length required for a given value of input resistance will not be such as to result in a purely resistive input impedance. The equivalent input circuit then will have one of the two forms shown in Fig. 18-5. The line reactance in these two circuits can be nullified by connecting an equal value of reactance of the opposite type in series, as indicated in the drawing.

To illustrate, assume that a terminating resistance of 10 ohms (typical of a low-impedance antenna system) is to be matched to a line having a Z_o of 50 ohms. If the same type of line is used for the impedance-transforming section or matching section, the normalized value, R_i/Z_o, is required to be 50/50 = 1. The VSWR in the matching section is 50/10 = 5 to 1, so from Fig. 18-4 (upper curves) the required length is either 67° or 113°. At either length the normalized reactance (lower curves) at the input of the matching section is 1.9, so the

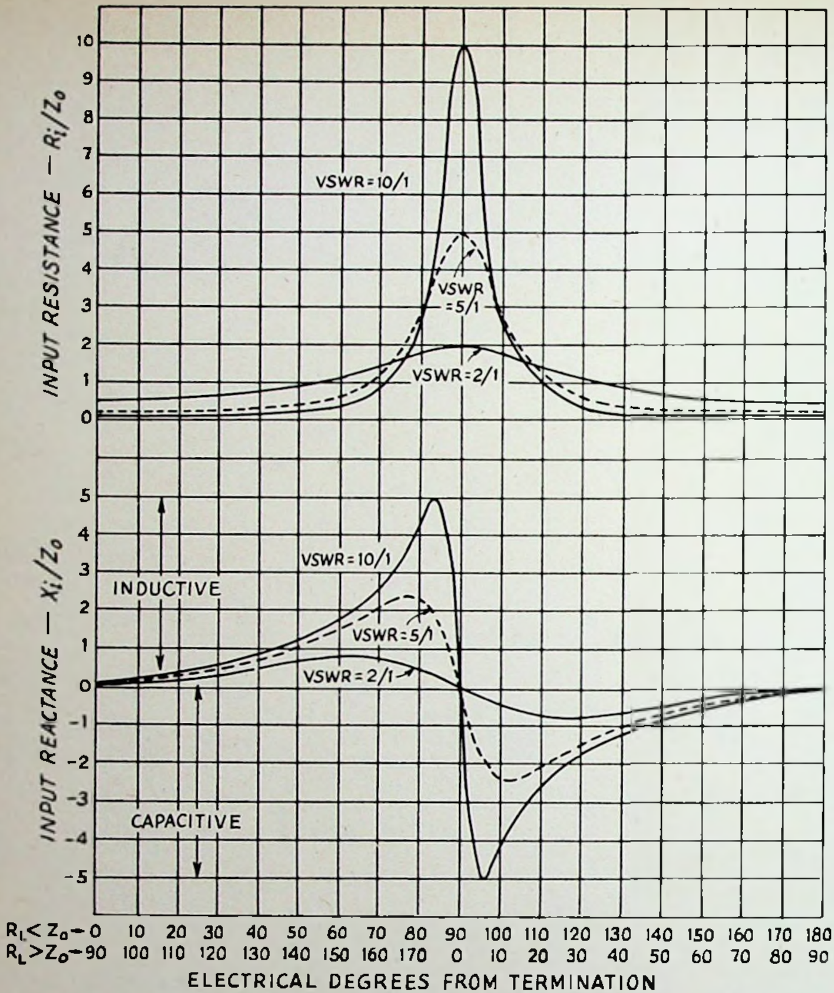


Fig. 18-4

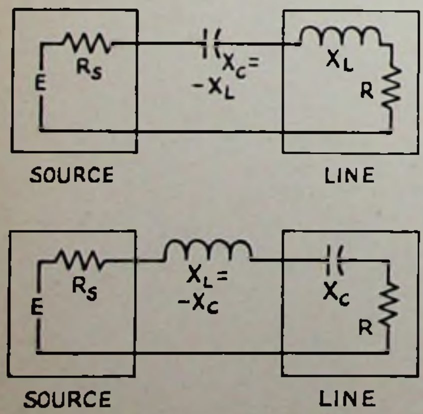


Fig. 18-5

actual value of reactance is 1.9 Z_o, or 95 ohms. If the section length is 67° the reactance is inductive; if 113° it is capacitive. In the inductive case it is necessary to add a series capacitive reactance of 95 ohms (upper drawing, Fig. 18-6) to make the input impedance a pure resistance. In the capacitive case a series inductive reactance of the same value (lower drawing, Fig. 18-6) is required. When this is done any length of 50-ohm line terminated by the matching section will be flat.

QUARTER- AND HALF-WAVE LINES: In the special case of the quarter-wavelength line (or any *odd* multiple of 1/4 wavelength) the input impedance is purely resistive if there is a pure-resistance termination, and no reactance cancellation is needed at the line's input end. In this case the formula for impedance transformation is simple:

$$R_i = \frac{Z^2_o}{R}$$

where R_i is the input resistance and R is the resistance of the termination (see Fig. 18-7). Ordinarily the termination and desired resistance have assigned values, in which case the required Z_o is

$$Z_o = \sqrt{R_i R}$$

When the line length is one-half wavelength or any integral multiple of a half wavelength, the input impedance is exactly the same as the impedance of the termination. This is true regardless of the type of termination – i.e., whether it is pure resistance or a complex impedance, so long as the line losses are negligible. This is because the impedance transformation that takes place in the quarter-wave section connected to the termination is inverted in the next quarter-wave section and so comes back to the original value. It is also true for any value of Z_o, so half-wave sections having different values of Z_o may be cascaded without changing the input impedance, provided *each half-wave section* has constant Z_o.

PARALLEL REACTANCE; MATCHING STUBS: When an impedance-transforming section is used between the termination and the remainder of a line in order to make the remainder operate as a flat line, a shunt reactance can replace the series reactance which cancels the input reactance of the matching section, Fig. 18-6. This requires finding the equivalent circuit of parallel resistance, R', and reactance, X', that can replace the series resistance, R, and series reactance, X, that will match the remainder of the line when X is canceled out. (See Chapter 9 for discussion of equivalent circuits.) The line circuit is then as shown in Fig. 18-8. The line length required for the normalized value of *series* input resistance to be equal to 1 will not be the length that will give an equivalent *parallel* value equal to 1. Neither will the values of X and X' be the same, although both will be of the same *type* (inductive or capacitive, as the case may be).

Fig. 18-4 can be used for converting from series to parallel. By moving 90 deg farther away from the termination than the actual point at which the equivalent values are to be found, the R/Z_o curves give the normalized *conductance* for the original point, and the X/Z_o curves give the normalized *susceptance*. For example, if a line 50 deg long (or 50 deg plus any integral multiple of 180 deg) is terminated in a resistance lower than Z_o and has a VSWR of 5 to 1, the input resistance as read from Fig. 18-4 is $0.4Z_o$ and the input reactance is $1.0Z_o$, near enough. Moving 90 deg to the right (to 50 + 90 = 140 deg) the conductance, as read from the resistance curve, is 0.35 and the susceptance as read from the reactance curve is 0.7. The values of conductance and susceptance can be converted into equivalent parallel resistance and reactance by dividing each into 1, giving R' = $2.86Z_o$ and X' = $1.4Z_o$. As mentioned above, the equivalent reactance is of the same type as the reactance found in the series case, not as given by the point on the curve from which the susceptance is obtained. If R is greater than Z_o the series impedance values are found in the right-hand half

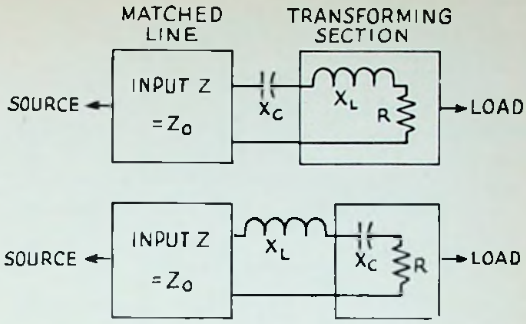


Fig. 18-6

of the chart, and in this case the conductance and susceptance are determined by moving 90 deg to the *left*.

Matching with a parallel reactance requires that the normalized conductance, rather than the resistance, be the same as that of the line to be matched; that is, a line length must be selected such that G = 1. The required susceptance is read off at the same point as G, and its reciprocal gives the equivalent parallel reactance, X', needed for cancelling the equivalent input reactance. However, the *actual* matching-section length is either 90 deg longer or 90 deg shorter than the length at which G = 1.

An example may help clarify this. Assume that a parallel reactance is to be used in the example previously discussed (R = 10, Z_o = 50, VSWR = 5, where the line length for series-reactance matching was determined to be 67° or 113°). At these line lengths G also equals 1, but for an equivalent resistance R' at a point 90 deg farther away. Thus

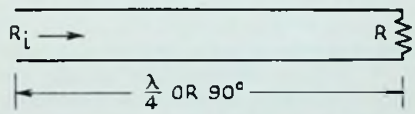


Fig. 18-7

the *actual* line length for matching with a parallel reactance is either 67 + 90 = 157 deg or 113 – 90 = 23 deg. In the 157° case the equivalent input reactance X' is capacitive because R $>$ Z_o and the matching-section length is greater than 90 deg but less than 180 deg. In the 23° case it is inductive because the matching-section length is less than 90 deg, R being less than Z_o. The normalized susceptance at either point is approximately 0.3, and the equivalent reactance X' therefore is 1/0.3 = $3.34Z_o$; that is, 3.34 × 50 = 167 ohms.

It is generally more convenient to use a shunt reactance for matching because the reactance can be connected directly across the main transmission line (at the proper point) without the necessity for cutting the line, which a series reactance would require. Further, the reactance can be a shorted or open section of the same type of transmission line (see Chapter 17 for determination of reactances). A section of line acting as a shunt reactance is called a matching stub.

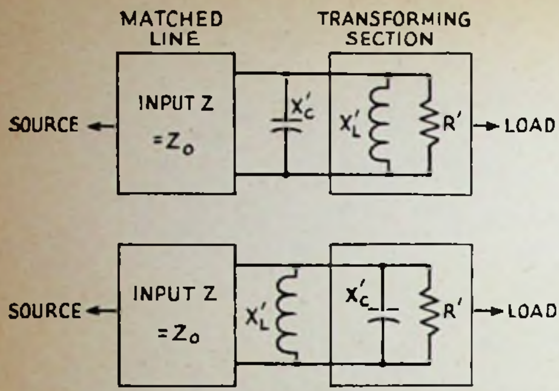


Fig. 18-8

Provided the termination is a pure resistance and the VSWR without matching is known, the required reactance and position of a matching stub can be determined from Fig. 18-4 by the method described. If necessary, values between the curves can be found by interpolation. From the standpoint of power lost in the line because of standing waves remaining after approximate matching, high accuracy is not too important, because it is relatively easy to reduce the VSWR below 2 to 1 with values that are only approximate. The initial measurement of VSWR can be made with any convenient length of line.

CANCELLATION OF REACTANCE: If the load or termination is a complex impedance its reactance may be "tuned out" by supplying an equal amount of reactance of the opposite type. This is shown in Fig. 18-9(A) for a load having capacitive reactance in series with its resistance. L, C, and R form a series-resonant circuit at the operating frequency, so the net impedance is simply equal to R. This provides a pure-resistance termination, and the methods described above may be used for determining the input impedance of a line of any length.

A resistive termination also will result if the series impedance of the load is converted into its equivalent parallel resistance and reactance as outlined in Chapter 9. In this case the compensating reactance required for making the load into a pure resistance is connected in *parallel* with the load, as shown at (B). The values of the equivalent-circuit elements differ from those of the series circuit, as explained earlier, and the load impedance calculations must be based on R' rather than R when this alternative is used.

Whether or not such reactance compensation is used, the impedance seen by the source of power is the *input* impedance of the line. Similar compensation may be employed at the input end to make the input impedance purely resistive, as already mentioned (Fig. 18-5), and the coupling or impedance-matching methods described in earlier chapters can be used for supplying power to a line in the same way as for a supplying power to a lumped-constant circuit.

FREQUENCY RESPONSE: An open- or short-circuited line is a sharply resonant tuned circuit because it simulates a high-Q lumped-constant circuit. That is, the loss resistance is very low, and as the frequency is moved away from the one at which the input impedance is purely resistive the reactance changes very rapidly. From inspection of Fig. 18-4 it is apparent that when the line is terminated in a resistance the slope of the reactance curve on either side of the quarter-wave point (this change in wavelength is equivalent to varying the applied frequency) depends on the VSWR – that is, upon the degree of mismatch. The higher the VSWR the more rapid the change in reactance for a given change in frequency near resonance. A line operating with a high VSWR thus is sharply tuned.

Conversely, the lower the VSWR the more tolerant the line becomes of changes in input frequency. When the line is matched and the VSWR is 1 to 1, there is no reactance in the input impedance at any applied frequency. Such a line has no selectivity at all. (This can be strictly true only if the load itself is purely resistive and constant in value, and the line losses are low.)

In practice, the line losses at hf will be low enough so that the line itself is substantially unselective. Practical loads, however, are rarely purely resistive over an appreciable frequency range. Thus any observed frequency discrimination is likely to be caused by variations in load impedance, causing increasing mismatch as the frequency is moved away from the one at which the termination is actually a resistance.

LOSSES IN PRACTICAL LINES: There is bound to be some power loss in an actual line, because the conductor resistance cannot be zero, and because there are dielectric losses in the insulating material used for keeping the line conductors at constant separation. In a two-wire line having spacers (rods installed at intervals along the line to keep the wires separated) the insulation is mostly air and the losses are usually negligible at hf if the spacers are made of low-loss material such as steatite or polyethylene.

There is a small radiation loss in two-wire lines because the conductors have to be separated a finite fraction of a wavelength and the external fields therefore cannot cancel completely at a distance. However, if the spacing is small – a few hundredths of a wavelength – this loss is inconsequential.

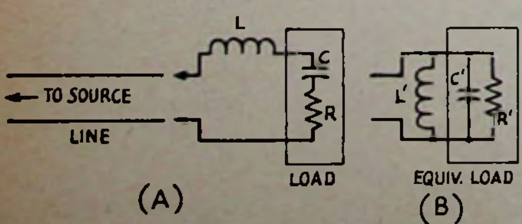


Fig. 18-9

Losses are higher in two-wire lines having a solid-dielectric web between the wires, although the fact that part of the electromagnetic field is in the air surrounding the insulation helps to reduce loss.

Coaxial lines with air insulation and internal spacers also have relatively low losses. The type in which a solid dielectric fills the space between the inner and outer conductors has the highest loss of all the types mentioned.

When a line is terminated in its characteristic impedance all the power flows in one direction, toward the termination. In this case the power lost in each unit of length will be a constant percentage of the power entering it. For example, if 1% of the power is lost in 10 feet of line, 99% of the original power enters the next 10 feet. Another 1% is lost in this section, so the power at the end of 20 feet is 0.99 × 0.99, and at the end of the third 10 feet it is 0.99 × 0.99 × 0.99, and so on. This relationship can be expressed in decibels per unit length. The unit length used in most specifications is 100 feet.

Losses in lines increase with frequency, just as they do in other circuit components. The reasons also are the same – higher conductor resistance and dielectric losses. The loss/frequency relationship varies with the type and construction of the line, and manufacturers' data should be consulted when commercially produced lines are used. As a rough estimate it can be assumed that the loss in decibels in solid-dielectric lines will be multiplied by about 1.25 each time the frequency is doubled.

LOSS CAUSED BY MISMATCHED TERMINATION: The loss mentioned in the preceding section is, as stated, based on the assumption that the power travels only in one direction in the line – i.e., that the termination matches the line Z_0. If there is a mismatch, the reflected power also contributes to the total loss, because power lost on the way back to the source is not available to reduce the net load on the source at the input end of the line.

Since the amount of power reflected depends on the reflection coefficient, which also sets the standing-wave ratio, the additional loss caused by reflection can be expressed in terms of the VSWR. Fig. 18-10 shows this. The additional loss in any particular case depends on what the total loss would have been had the line been perfectly matched by its termination. For example, if a line has a loss of 1.5 dB per hundred feet and is 200 feet long, its actual loss when matched is 3 dB. But if it is mismatched so that the VSWR is 7 to 1, the *additional* loss caused by the mismatch is 3 dB, making a total loss of 6 dB. Note, however, that if the *matched* loss had been 1 dB instead of 3, the additional loss caused by the same VSWR would have been 1.7 dB, approximately.

That is, the loss caused by standing waves depends on the *matched* loss in the line, and may be inconsequential if the matched loss is small. This is the case with air-insulated two-wire lines, which can operate at high VSWRs with low overall loss, for practically any lengths used in the average installation.

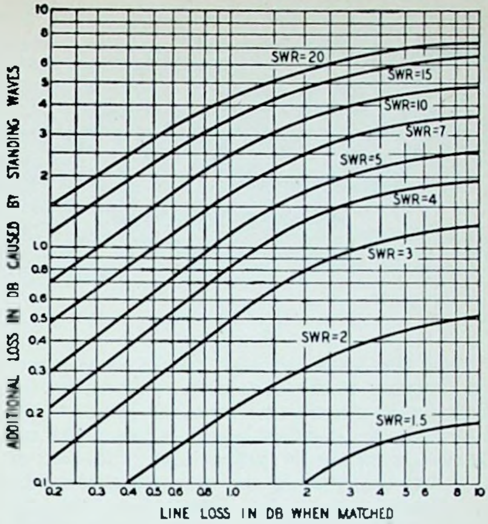


Fig. 18-10

LOSS CAUSED BY INPUT MISMATCH: The line losses so far discussed are real, in the sense that they are caused by actual dissipation in the line when power is being delivered to a load. There is another form of mismatch loss which can occur when the *input* impedance of the line does not have the proper value to represent an optimum load for the source of power. This loss is not "real" in the sense that any power is actually dissipated, but simply represents the difference between the power actually put into the line and the power that *could* have been put into it had the input impedance been properly matched to the source impedance (see Chapter 10). The use of a suitable impedance-matching circuit between the source and the line will eliminate this "loss."

If it is inconvenient or difficult to match the power source to the line (this might be the case when the source is a receiving antenna feeding incoming signals through a line to a receiver) a match can be brought about at the input end of the line by deliberately mismatching the line at the *output* (receiver) end, using the impedance transforming properties of the line described earlier in this chapter. While this increases the VSWR on the line, the added loss caused by standing waves often will be considerably less than the input-mismatch loss, so there is a net gain in power at the receiver. This procedure is frequently called "tuning the line."

OPERATING CHARACTERISTICS OF LOSSY LINES: The exact determination of voltages, currents, and impedances when the line loss is appreciable is not possible with simple algebra. If the total line loss is small – up to perhaps 1 dB – the relationships discussed for loss-free lines will provide sufficiently good approximations for practical work, since adjustments (e.g., coupling) can usually be made to result in optimum power transfer from a source to a load.

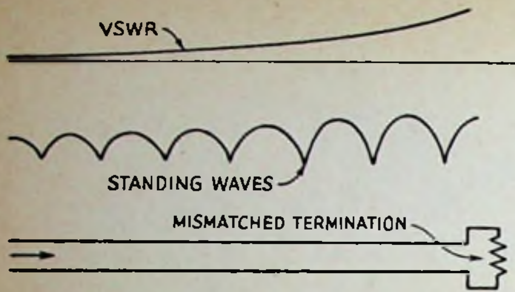


Fig. 18-11

When the losses are high, the incident power level is highest at the input end of the line, but the reflected power level is lowest at this point. At the termination, the incident power is lower, but the reflected power has its highest value. The result of this is that the standing-wave ratio varies along the line. Fig. 18-11 is an illustration. At the termination the VSWR is calculated as described earlier. However, at the input end of the line it is smaller because of the power lost in traveling back over the line.

Measurement of VSWR at the input end in such a case is always misleading, because the VSWR indicated is lower than the actual VSWR at the load. In extreme cases it may be nearly 1 to 1 at the input end even though it may be 10 to 1 or more at the termination. Such cases can occur at vhf and uhf when solid-dielectric coaxial cable is used in relatively long lengths.

RADIATION FROM LINES: Two-wire lines operating as described in this chapter will lose very little power by radiation; practically all of the loss will be in the resistance and dielectric of the line itself. However, radiation loss can be appreciable if the currents in the conductors are not equal and opposite at every point. Unbalance of this type can occur quite readily when, for example, a line is used for feeding power to an antenna system. The field set up by the antenna can cause currents in the line which flow in the *same* direction in both wires and thus their fields do not cancel at a distance. Unbalance thus causes radiation of power in the same way as from an antenna of dimensions similar to the line's and having the same unbalanced current.

With coaxial line this cannot occur *inside* the line; only the transmission-line currents can exist there. However, the *outside* of the outer conductor is just as susceptible to antenna fields as the two wires of a parallel-conductor line, so with coaxial line radiation can take place because of currents on the outside of the outer conductor.

Neither type of line has an advantage with respect to radiation of this type. The only remedy is to prevent such currents from appearing on the line. Each case must be analyzed to find the cause of the unwanted coupling between the antenna and line before a suitable remedy can be applied.

WAVELENGTH IN LINES: It is to be understood that the lengths used in the discussions in this chapter (and also Chapter 17) are electrical wavelengths, which include the effects of any departure of the velocity of energy moving along the conductors from the velocity of light, and of any incidental external loading. The actual *physical* length of a line may differ from the length calculated by the ordinary wavelength formula.

A principal cause of this departure is the presence of a dielectric between the line conductors. The effect of the dielectric is to reduce the velocity. If the entire electric field is in some other dielectric than air, the velocity is modified by the factor $1/\sqrt{k}$, where k is the dielectric constant of the insulating material. In air-insulated lines only a relatively small amount of dielectric is used, so the reduction in velocity is small; the velocity factor – the ratio of the actual velocity to that in air – is 0.95 or more in the average case. In solid-dielectric two-wire lines, where the field is partly in air, the velocity factor is about 0.82; in solid-dielectric coaxial lines it is 0.66. Thus the physical length for a given electrical length in wavelengths varies according to the type of line used.

The factors given above, or as given by the manufacturer for a particular type of line, will suffice in ordinary cases, but if high accuracy is required the line length must be adjusted to obtain the exact results described earlier. This refinement is seldom justified in practical cases.

Questions and Problems

18-1) What is meant by a standing wave?

18-2) What is the load placed on a source of power when the source is connected to a transmission line? Is it the same as the characteristic impedance of the line?

18-3) If a certain type of line has a rated loss of 0.25 dB per 100 ft, what is the loss in 150 feet of such line?

18-4) In what way does a line terminated in a resistance less than its characteristic impedance resemble a short-circuited line? Why?

18-5) What effects result in line operation from the presence of reflected voltages and currents?

18-6) If a 200-ohm resistance is to be matched to 50 ohms for terminating a line, by what method or methods could it be done?

18-7) On what does the frequency discrimination of a line terminated in a power-absorbing load depend?

18-8) If a quarter-wave impedance-matching transformer is to be used to match 20 ohms to 50 ohms, what Z_0 in the quarter-wave section is required?

18-9) What is the relationship between terminating resistance, line Z_0 and VSWR? Under what conditions is this relationship applicable? When is it not applicable?

18-10) The VSWR measured at the termination of a line is 5 to 1. A similar measurement at the input end of the line shows 3 to 1. What is the reason?

18-11) The input impedance of the line in Q.

18-10 is found to be reactive as well as resistive. Can the VSWR be improved by tuning out the reactance?

18-12) Suppose the line length in Q. 18-11 is doubled. Will the VSWR at the input end be reduced? What if the line length is halved? Is there an intermediate length at which the VSWR will be least?

18-13) What is the relationship between standing-wave ratio and radiation from a line?

18-14) Using Fig. 18-4, what would you estimate the input impedance of a line having a Z_0 of 75 ohms to be if the line is 3/8 wavelength long and is terminated in a resistance of 300 ohms?

18-15) If a line provides a poor load for a source of power, what should be done?

18-16) If the source of power used with the line of Q. 18-14 requires a 50-ohm load for optimum operation, what can be done to provide it?

18-17) What is the meaning of reflection coefficient? Is there any relationship between reflection coefficient and VSWR?

18-18) If the line in Q. 18-3 is operating with a VSWR of 6 to 1, what is the total loss in the 150-foot length?

Thermionic emission, space charge; cathodes; conduction in vacuum; dependence of plate current on plate voltage; power input and plate dissipation; plate resistance; control grid; amplification; characteristic curves, voltage amplification factor; transconductance; plate-current cutoff; grid current; zero-bias tubes; grid-cathode resistance and power; interelectrode capacitances; the screen grid; secondary emission; characteristics of screen-grid tubes; interelectrode capacitances of tetrodes and pentodes; screen dissipation; transit time; lead inductance.

Chapter 19

Vacuum Tubes

THE POWER in signals at a receiving point after propagation over long distances is of the order of micro-microwatts – much too weak to be usable without first being built up or amplified by a factor of millions. Amplification is based on unilateral conductivity which can be controlled by the signal. The same principle is also used for the generation of rf power.

The principal devices used for this purpose are vacuum tubes and semiconductors. Since vacuum-tube principles are somewhat simpler than those of semiconductors they are considered first; semiconductors are taken up in the following chapters.

THERMIONIC EMISSION; SPACE CHARGE: When a conductor is heated its free electrons acquire additional energy, and if the temperature is high enough (a few thousand degrees C for ordinary metals) electrons near the surface will pick up enough velocity to fly out of the conductor. This is called thermionic emission.

If the emitter is in a vacuum, the escaping electrons will travel away from the surface. Some will quickly lose energy and fall back to the emitter, but others will form a cloud of negative charges near the emitter. This accumulation is called the space charge. Its electric field repels electrons subsequently leaving the surface, driving them back to the emitter, so equilibrium is quickly established.

CATHODES: An emitter of electrons is called a cathode. It is one of the electrodes or elements that enter into the construction of vacuum tubes. The cathode is raised to the necessary temperature by electrical heating.

Thermionic emitters are various metals and metal compounds that can withstand high temperatures. Some are more efficient emitters than others. Pure metals are least effective; a mixture of thorium and tungsten is more so, and certain compounds called "rare earths" are most efficient of all. Efficiency is defined as the number of electrons (or electron current) emitted per unit of cathode area per watt of heating power.

There are two principal forms of cathode construction. The directly heated cathode is simply a wire; it is heated by sending enough current through it to develop the required heating power. If a rare-earth emitter is used it is coated on the surface of a wire chosen for mechanical strength and the ability to operate at high temperatures.

The second form is the indirectly heated cathode, in which a rare-earth emitter is coated on the surface of a metal tube. A heater of insulated wire is coiled or folded inside the tube, which then conducts heat to the actual emitting surface. There need be no electrical connection between the heater and cathode in this type of structure. Several common forms of construction are shown in Fig. 19-1. Directly heated types are at (A), (B), and (C). (D) and (E) are indirectly heated types.

CONDUCTION IN A VACUUM: The second important element in a vacuum tube is the anode or plate. Physically, the plate can have a variety of forms – a hollow metal cylinder surrounding the cathode, flat metal plates on either side of the cathode, a rectangular structure enclosing the cathode, and so on.

The plate ordinarily is given a positive charge by an external voltage source having its negative terminal connected to the cathode. The positive potential attracts electrons from the space charge, which after reaching the plate flow through the external voltage source back to the cathode. Thus there is a current flow – plate current – in the plate circuit. If the plate is given a negative charge the space-charge electrons are repelled by it and no current can flow.

If the plate has no charge at all with respect to the cathode, there may be a minute current flow because some of the electrons will be emitted with enough velocity to reach the plate even though it has no positive charge to attract them. This initial-velocity current (sometimes called contact-potential current because contact potential, a phenomenon associated with dissimilar metals, has a similar effect on current flow) is small enough to be neglected in all but a few applications.

A vacuum tube having a cathode and plate is called a diode or two-element tube.

DEPENDENCE OF PLATE CURRENT ON PLATE VOLTAGE: If the positive voltage on the plate is increased, starting from zero, it is found that the plate current increases with increasing positive plate voltage. This is shown in Fig. 19-2. With very low voltage the current rises slowly at first, but as the voltage is increased the *rate* at which the current rises per volt change in plate voltage increases. At some low value of plate voltage the curve begins to straighten out and becomes an essentially straight line, as shown. This does not continue indefinitely, however, and eventually a point is reached where a given increase in plate voltage produces a smaller increase in plate current than was the case in the straight-line portion. Ultimately, the curve bends toward the horizontal, indicating that a further increase in voltage has little or no effect on the plate current. This is called the **plate saturation** region, and the rounded portion (called the **knee**) of the curve marks its beginning.

Saturation occurs when all, or nearly all, of the electrons emitted by the cathode are attracted to the plate. Tubes ordinarily are designed with sufficient cathode emission so that plate saturation does not occur at any plate voltage used in normal operation.

POWER INPUT AND PLATE DISSIPATION: At any point on the curve the plate current at that point multiplied by the plate-to-cathode voltage that produces it is the power input to the plate. This power, like the power dissipated in a resistor, is converted into heat. The plate must dissipate this heat, and the amount of power lost in the plate that can be dissipated without damage to the tube is called the safe plate dissipation rating. This rating is important in tubes that handle appreciable amounts of power.

PLATE RESISTANCE: Dividing the plate voltage at any point on the curve by the plate current at that point gives a value of resistance which is sometimes called the "static" resistance of the plate-cathode circuit. The resistance so found is different for every value of plate voltage, and is not particularly useful in making circuit calculations.

However, the straight or linear portion of the curve does have a constant resistance in one sense. With any resistance obeying Ohm's Law the graph of current vs. voltage is a straight line. If the

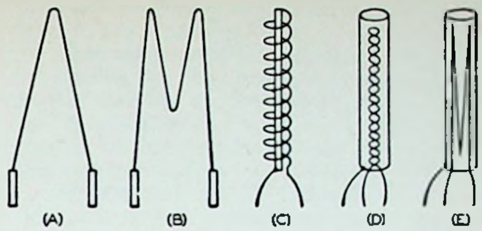


Fig. 19-1

straight part of Fig. 19-2 is extended to intersect the voltage axis as shown by the dashed line, the intersection at *A* becomes the "zero-voltage" point for such a resistance. Subtracting the voltage at *A* from the actual voltage applied to the plate gives the voltage which is effective in causing current to flow in such a resistance. Taking point *B* as an example,

$$r_p = \frac{B - A}{C}$$

C being the value of plate current. r_p is called the **dynamic plate resistance** of the diode, and its value is constant anywhere on the curve between points *D* and *E*. This dynamic resistance (usually just called "plate resistance," is the resistance that is effective when the plate current is *changing*, which is the condition under which vacuum tubes do useful work. The curve in Fig. 19-2 is somewhat idealized; in a practical tube no appreciable portion would be *exactly* straight. However, a segment of it, if short enough, will approximate a straight line, and the slope of the segment gives the value of plate resistance for that segment. The slope is found by dividing a *small* change in plate voltage by the corresponding change in plate current. That is,

$$r_p = \frac{\Delta E_b}{\Delta I_b}$$

where ΔE_b is the change in plate voltage and ΔI_b is the change in plate current caused by ΔE_b. Using this method, the plate resistance can be found for

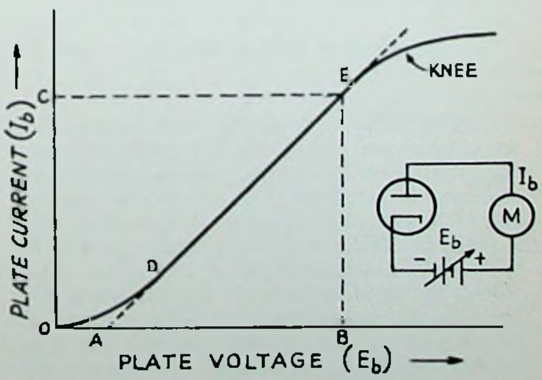


Fig. 19-2

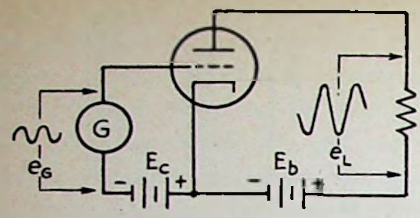


Fig. 19-3

any part of the curve. The plate resistance is high at low plate voltages (in the region from *0* to *D*), has a substantially constant value from *D* to *E*, and rises again when the knee of the saturation region is reached.

CONTROL GRID: The diode has unilateral conduction but no means of controlling it, aside from varying the plate voltage. While this method has its applications, it does not offer the possibility of obtaining a larger output signal from the tube than is applied to it. What is needed is a means for varying the plate current independently of the plate voltage, preferably one that takes little or no power from the signal source.

This can be done by inserting a control grid between the cathode and plate so that a signal applied to the control grid can vary the space charge, and thus vary the plate current, independently of the plate voltage. The grid is usually an open-work wire structure. If it is at the same potential as the cathode, the positive charge on the plate can attract electrons through the open spaces, so plate current flows. If the grid is given a negative charge (negative grid bias) the plate's field strength is reduced in the space-charge region and fewer of the space-charge electrons will "feel" the positive charge on the plate strongly enough to be attracted through the grid. As a result, the plate current is decreased. On the other hand, a *positive* charge on the grid enhances the field set up by the plate, so more of the space-charge electrons flow through the grid openings to the plate and the plate current is increased. For a given plate voltage, the change in plate current is approximately proportional to the change in grid voltage, so the plate current can be smoothly varied by an alternating voltage on the grid.

A three-element tube of this sort is called a **triode.**

AMPLIFICATION: In Fig. 19-3, assume that the plate-supply voltage, E_b, and grid-bias voltage, E_c, have been chosen so that some moderate value of plate current flows. The grid bias is negative, which is the normal mode of operation (the reasons for this will become more apparent in subsequent discussions). If a source, G, of ac voltage is connected between the bias source and the grid, the positive swing of the ac voltage will cancel part of the negative bias and the plate current therefore will increase. Conversely, the negative swing will add to the negative bias and *decrease* the plate current.

The plate current thus consists of a steady dc component on which is superimposed an ac component (see Chapter 6). The plate load, R_L (shown as a resistance although in practice it may be and impedence), will have a varying as well as a fixed component of plate current flowing through it. If the proper operating conditions are chosen, the varying component, e_L, is an enlarged or amplified reproduction of the ac grid voltage, e_G, as indicated in the drawing.

The determination of suitable operating conditions for amplification requires familiarity with the ratings and characteristics of the type of tube to be used, as does also the selection of a proper tube type to provide specified performance.

CHARACTERISTIC CURVES: Fig. 19-4 shows a set of triode characteristic curves known as the **plate family**. They are obtained by setting the grid potential or grid bias to selected fixed voltages and then varying the plate voltage to observe the corresponding changes in plate current. (Although the curves shown do not correspond to those of any particular tube type, they are representative of actual tubes.) Curves for different values of grid voltage are more or less parallel except near zero plate current, where the curvature resembles that shown for the diode in Fig. 19-2. The plate saturation region does not appear in this graph because the saturation current would be far above the safe operating range of plate current.

The grid-voltage curve marked $E_c = 0$ shows how the plate current would follow the plate voltage if the grid and cathode were at the same potential. For a given plate voltage, the plate current is always smaller than the zero-bias current when the grid bias is negative, and is always larger than the zero-bias current when the grid bias is positive.

The line at the left marked $E_c = E_b$ is the diode line; it shows what the plate current will be if the same positive voltage is applied to both the grid and the plate. The tube is not ordinarily operated in the region to the left of this line because when the grid becomes more positive than the plate most of the electron current flows to the grid and is not available for the plate circuit.

The dashed curve marked "9W" in Fig. 19-4 locates the values of plate current for which there is a loss of 9 watts in the plate of the tube, for the range of plate voltages on the graph. If 9 watts is assumed to be the rated plate dissipation for the hypothetical triode shown, any operating point (fixed plate and grid voltages which form the center of operation for the tube) would have to be chosen to lie on or below this curve.

VOLTAGE AMPLIFICATION FACTOR: It is useful to know the relative effects of grid voltage and plate voltage on the plate current. The "static" values such as are read directly from the curves are (like the static plate resistance in Fig. 19-2) of little use in circuit calculations. A more meaningful relationship can be found by taking the effect of small *changes* in plate and grid voltage. The **voltage amplification factor** is the ratio of a small change in plate voltage to the small change in grid voltage that will make the plate current change by the same amount. That is,

$$\mu = \frac{\Delta E_b}{\Delta E_c},$$

where μ is the amplification factor.

The small triangle *ABC* in Fig. 19-4 illustrates this. The curve for a grid bias of −4 volts has been partially drawn so that the effect of a 1-volt change in grid voltage, *AB*, for a constant plate voltage of 250, can be found. It is approximately 5 mA from the graph. To make the same 5-mA change in plate current with a fixed grid bias of −3 volts the plate voltage must be reduced by approximately 30 volts, from *B* to *C*. In this example ΔE_b is 30 and ΔE_c is 1, so the amplification factor, μ, is 30, for E_b = 250 volts and E_c = −3 volts. For other voltages μ could be different, although for the parallel straight parts of the curves it will remain the same.

In practical triodes the amplification factor will range from a low of about 3 to a high of 100 or more, depending on the application for which the tube is intended.

TRANSCONDUCTANCE: For many purposes a more useful quantity than μ is one expressing the relationship between a change in grid voltage and the change it causes in the plate current at a given plate voltage. This is called the grid-plate transconductance, or simply transconductance. It is given by

$$g_m = \frac{\Delta I_b}{\Delta E_c},$$

where g_m is the transconductance. In Fig. 19-4, with 250 volts on the plate, a change in grid bias of 1 volt, from *A* to *B*, causes a change of 5 mA in the plate current, so ΔI_b = 5, ΔE_c = 1, and g_m = 5/1 or 5 mA per volt. Since I/E is a conductance, this can be expressed as 0.005/1 = 0.005 mho or (more usually) as 5000 micromhos (μmho). Values of transconductance for the majority of triodes range between about 1500 and 5000 μmho.

The same triangle, *ABC*, can be used to find the plate resistance as described earlier. In this case ΔE_b is 30 V and ΔI_b is 5 mA, so r_p = 30/0.005, or 6000 ohms.

The three quantities, μ, g_m and r_p, are obviously not independent, since they are all derived from the same triangle. The relationship between them is

$$g_m = \frac{\mu}{r_p}.$$

The equation can be transposed to find any of the three when the other two are known. Since all three vary with the choice of grid voltage and plate voltage, these must be specified.

PLATE-CURRENT CUTOFF: The curves in Fig. 19-4 could be carried down to the zero-current line, although the exact point at which the plate current becomes zero, for a given plate voltage, becomes difficult to determine from curves because of the bending at the bottom. The negative grid bias that causes the plate current to be zero, or substantially zero, is called the cutoff bias for that plate voltage. It is an important quantity in certain types of tube operation.

As an approximation, the cutoff bias is found by dividing the plate voltage by the amplification factor. Thus for the tube illustrated in Fig. 19-4 the cutoff bias for E_b = 500 volts would be 500/30, or −16.7 volts. Actually, more bias than this would be needed for complete plate-current cutoff; because of the lower-end curvature there is still some plate current flowing even with a bias of −18 volts. However, the approximation is useful as a guide to the order of magnitude of grid bias required.

GRID CURRENT: When the grid is negative with respect to the cathode it repels electrons and no current flows to it. (At zero grid voltage there can be a small initial-velocity current, as described earlier for the diode, but ordinarily it need not be considered in circuit performance.) However, when the grid is positive it will attract electrons and a grid current will flow.

If the positive grid voltage is small with respect to the plate voltage the grid current is small because most of the electrons go through the grid to the higher-potential plate. As the grid is made increasingly more positive, a larger proportion of the total current goes to the grid; this proportion rises rapidly as the grid voltage approaches the diode line, and eventually it becomes larger than the plate current.

A tube having characteristics such as those shown in Fig. 19-4 (it has a medium value of μ) normally would be operated in such a way that the grid is always negative. This means that the grid takes no power from the signal source, since there will be no grid current at any voltage, fixed or instantaneous, that may be applied between the cathode and grid.

ZERO-BIAS TUBES: For some purposes a tube that has a very high amplification factor is desirable; tubes of this sort usually are intended to be operated with zero or only slightly negative bias. The grid has relatively small openings so that

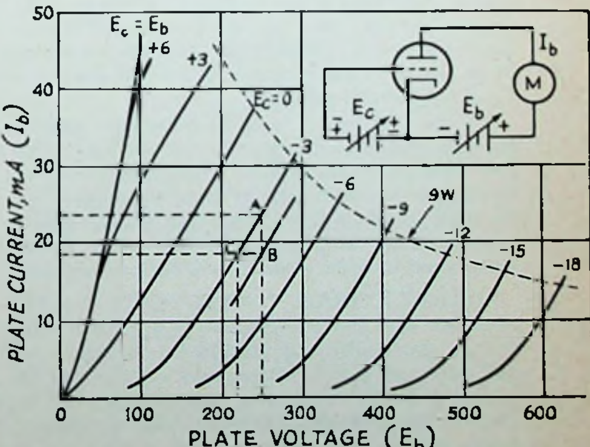


Fig. 19-4

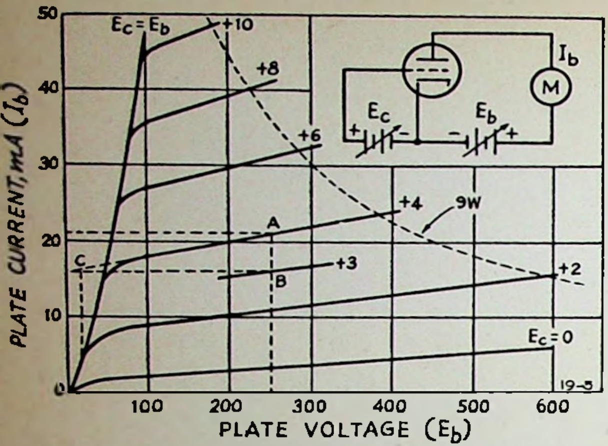


Fig. 19-5

it tends to shield the space charge from the field set up by the plate potential, thus reducing the plate current at zero bias to a level which holds the plate dissipation (at the normal operating plate voltage) within the safe plate-dissipation rating.

The plate family of curves for such a tube is constructed in the same way as in Fig. 19-4. However, because of the high μ the curves rise much more slowly with increasing plate voltage, as shown in Fig. 19-5 (which again is for a hypothetical tube), and if the plate currents are to fall in the same region as in Fig. 19-4 the grid voltage must be positive.

The triangle *ABC* again shows the relationship between g_m, μ, and r_p, but in this case the base of the triangle is so large that terminating it on the actual curve (as was done in Fig. 19-4) would include a section with too much bending for accurate results. Consequently the slope is extended by the straight line *AC* (this was also done in Fig. 19-2, where circumstances were similar). The transconductance is about the same, 5 mA per volt, but the plate resistance is much higher; ΔE_b (*BC*) is about 230 V, resulting in an approximate r_p of 230/0.005 = 46,000 ohms. Since a 230-volt change in plate voltage causes the same change in plate current as a 1-volt change in E_c, μ is approximately 230.

If the grid bias is zero, an ac voltage applied between grid and cathode will cause grid current to flow throughout the positive swing, but there will be no grid current during the negative swing. Also, the plate current will be cut off during much of the negative swing of grid voltage. This causes the ac component of plate current to be highly distorted.

GRID-CATHODE RESISTANCE AND POWER: The grid and cathode of a vacuum tube form a diode which, depending on the values of grid bias and signal voltage, may or may not be operating in the conducting region shown in Fig. 19-2. If the instantaneous sum of the alternating signal voltage and steady dc bias at the grid is negative with respect to the cathode, no grid current flows. If the instantaneous sum is positive, grid current does flow.

The general shape of this relationship is shown in Fig. 19-6. The curve is not the same as that for a simple diode because of the presence of the electric field from the plate; the value of the dynamic resistance is different at every value of grid voltage, and also at every value of plate voltage. This complicated relationship between voltage and current means that the grid-current waveshape has little resemblance to the voltage waveshape that produces it; i.e., the current waveshape is highly distorted. This must be taken into account in any application in which grid current flows.

The power lost in the grid-cathode circuit *at any instant* is equal to the grid voltage multiplied by the grid current at that instant, and appears as heat in the grid. Since the power is varying throughout the ac cycle when a signal is present at the grid, the loss has to be averaged over the cycle to determine the power loss that is observed as heat. Because of the nonlinear relationship between current and voltage and the fact that the relationship is different in every tube type, the grid power cannot be expressed by a simple formula. Graphical methods must be used to find the exact values.

INTERELECTRODE CAPACITANCES: The electrodes of a tube are metallic structures close to and insulated from one another, and so interelectrode capacitances exist between them. There are three such capacitances in a triode, as indicated in Fig. 19-7; although they are shown external to the tube in this drawing they exist *inside* the tube and cannot be changed or modified in any way in a given tube structure. The capacitance between the grid and plate is shown as C_{gp}; C_{gk} is the capacitance between grid and cathode, and C_{pk} is the capacitance between plate and cathode.

The capacitances are small, since the tube elements are not large. In small triodes intended for receiving applications the average is 1.5 to 2 pF for C_{gp} and C_{gk}, and about 0.5 to 2 pF for C_{pk}. The lower values of C_{pk} are associated with high-μ tubes. Capacitances are considerably larger in transmitting-type triodes because of the larger elements used in such tubes.

These interelectrode capacitances are small enough to be of little concern at low frequencies (in most of the audio range) but their reactances become comparable with normal circuit reactances

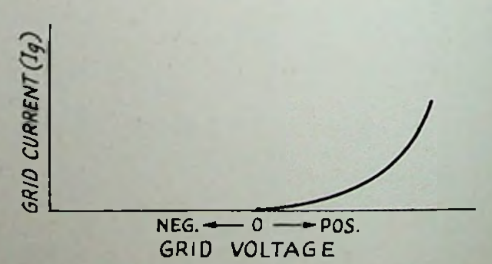


Fig. 19-6

at radio frequencies. This has a considerable influence on amplifier operation, a subject that will be dealt with in a later chapter.

THE SCREEN GRID: Additional control elements, usually in the form of grids, can be incorporated in the tube structure to perform special functions. The most universal of these is a grid inserted between the plate and the control grid. It is called a screen grid or accelerator grid, the former term referring to its function as a screen or shield between the control grid and plate to reduce the grid-plate capacitance. Where shielding is not a primary function, as in some power tubes designed for audio-frequency use, the second name is sometimes used, but "screen" is in general use for all cases.

The screen is operated at a positive potential with respect to the cathode, and its circuit is arranged so that the impedance between the screen and cathode is very low at the signal frequency. (This is usually accomplished by bypassing with a capacitance of very low reactance.) Since the screen prevents most of the field set up by the charge on the plate from penetrating into the space-charge region, the electrons are attracted almost wholly by the positive potential on the screen. Ordinarily the screen has a relatively low amplification factor – of the order of 3 – so that the electron current is large even with fairly high values of negative grid bias. However, the screen itself has relatively little area, and by proper tube construction it is possible to make nearly all of the electron current flow through its open spaces and travel on to the plate.

The result of this is that quite large plate currents flow, as compared with an equivalent triode, and the plate current is almost independent of the plate voltage, even when the plate voltage is low, because the electrons acquire sufficient velocity, in traveling to and through the positively charged screen, to be captured by the field of the plate.

As the screen is the fourth element in such a tube, the tube is called a tetrode.

SECONDARY EMISSION: Electrons striking the plate at relatively high velocity are capable of causing other electrons to "splash" from the metal surface, a phenomenon called secondary emission. In the triode these secondary electrons cause no trouble because they are immediately attracted back to the plate, that element being the one having the highest positive potential in the tube.

In the screen-grid tube, however, many of the secondary electrons will be attracted by the positive charge on the screen. Thus a reverse current can flow between the plate and the screen. This is undesirable, but it can be overcome if a third grid, called the suppressor grid, is inserted between the screen and plate. The suppressor, which has a large proportion of open area so it does not greatly interfere with electron flow, is usually connected to the cathode; that is, it is at zero potential with respect to the other tube elements. Its function is to shield the plate region from the field of the screen grid so that secondary electrons will not be

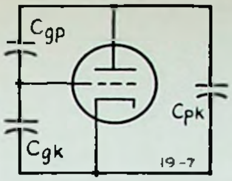


Fig. 19-7

attracted to the screen but will fall back on the plate.

Depending on the design of the suppressor, the tube may be either a pentode (five distinct elements, three of them grids) or a tetrode. In the latter, the suppressor takes the form of rods or plates which are in the vicinity of the plate but not actually in the electron stream. In either type the secondary emission is reduced to the point where the operation of the tube is not affected by it. Although there are differences between the two types, the characteristics are quite similar.

CHARACTERISTICS OF SCREEN-GRID TUBES: Fig. 19-8 is typical of the plate family of a power-amplifier type tetrode or pentode. Note that the curves at the higher plate voltages (E_b greater than E_{G2}) are almost horizontal, indicating that in this region the plate voltage has very little influence on the plate current. The plate-current change, *AB*, about the selected operating point, *A*, is about 5 mA per volt, so the transconductance, 5000 µmho, is in the normal range.

Because the curve shows appreciable bending at the operating point it is not possible to get a

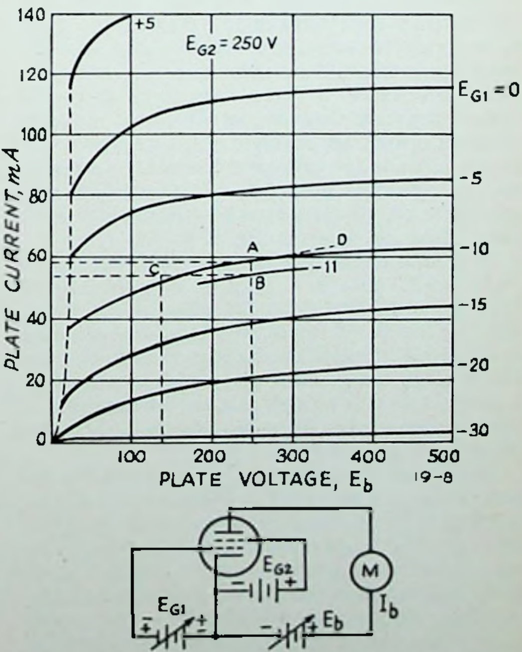


Fig. 19-8

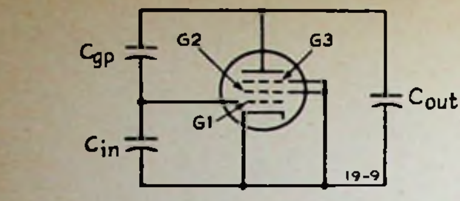


Fig. 19-9

reasonably accurate determination of the plate resistance by using the curve itself as one side of the triangle *ABC*, as was done in Fig. 19-4, nor is there a straight-enough section to allow using the method of Fig. 19-5. However, a *tangent* to the curve at *A* will give the slope of a small segment around *A* (the tangent is a straight line touching the curve, without intersecting, at just one point). Using the tangent, *CD*, gives the needed point *C* for finding the plate resistance and amplification factor. The voltage change is 250 – 140 = 110 V, and since the corresponding current change is 5 mA, r_p is 110/0.005 = 22,000 ohms. The amplification factor is 110/1 = 110.

The characteristic curves of small receiving-type pentodes used as rf amplifiers are as a general rule "flatter" – that is, more horizontal – over a wide range of plate voltage than those shown in Fig. 19-8. In these, the control grid (G_1) is more thoroughly shielded from the plate's field. The plate resistance of such types is typically 0.1 to 1 megohm, although transconductances are generally 3000 μmho or more.

INTERELECTRODE CAPACITANCES OF TETRODES AND PENTODES: Because the structure of a tetrode or pentode involves more than the three elements of the triode, there are obviously more capacitances between the various electrodes than in the three-element case. However, the extra elements usually are operated at the same potential as the cathode for the signal frequency (although the dc operating potentials of course differ) so only three capacitances need be considered in most cases. These are shown in Fig. 19-9.

The capacitance between grid and plate, C_{gp}, is much smaller than in a triode because the control grid, G_1, is shielded from the plate by the screen grid, G_2, and (if used) G_3. C_{gp} in pentodes designed for rf amplification in receivers may be as low as a few thousandths of a picofarad, although in some cases it is as high as a few hundredths. In tetrode power tubes for audio use the capacitance usually is of the order of 0.5 pF.

The input capacitance, C_{in}, is principally the sum of the grid-to-cathode and grid-to-screen capacitances, with small contributions from the other elements. The output capacitance is principally the capacitance from plate to screen and G_3, again with small additions from other elements. Both C_{in} and C_{out} are generally specified to be the capacitance from control grid and plate, respectively, to all other elements connected together and grounded.

SCREEN DISSIPATION: Although the major part of the electron current from the cathode in a tetrode or pentode goes to the plate, some electrons go to the screen grid, giving rise to a screen current in the circuit formed by the screen and cathode. In normal operation this current does no useful work in an external load, and so the power it represents is entirely used in heating the screen. The power is equal to $E_{G2} \times I_{G2}$, and while it is much smaller than the plate current, the screen has much less metal in it than the plate, and so for a given rise in temperature the power lost in the screen – the screen dissipation – must be held to a considerably lower value than the plate dissipation.

The ratio of safe screen dissipation to safe plate dissipation varies with the tube construction, and tube data must be consulted to find the safe values in a particular case.

TRANSIT TIME: Throughout the foregoing discussion it has been tacitly assumed that no time delay occurs when electrons are accelerated by an electric field. At frequencies up through the vhf region, at least, the field generated by voltage applied to one tube element is "instantaneously" felt at another – that is, the time interval is short enough to be neglected – considering the spacings between elements in vacuum tubes. However, electrons accelerated by the field do not move nearly as rapidly as the fields themselves, and the time required for an electron to move from one element to another can have a marked effect on the tube operation. The time delay or transit time amounts to a difference in phase between the applied voltage and the electron current it causes. As the phase difference depends on the ratio of the transit time to the time of one cycle, it increases with frequency.

Whenever there is a phase difference (other than 90 or 180 deg) between current and voltage there must be a resistive component in the impedance of the circuit. This means that the phase difference causes a power loss in the circuit. For example, the grid-cathode circuit of a tube operating with negative bias is not a simple capacitance (C_{gk}) but exhibits a finite value of shunt resistance which absorbs power from the signal source. It varies with the tube construction, operating conditions, and the frequency. Transit time begins to cause observable effects when the frequency is of the order of 10 MHz, and at 100 MHz typical shunt resistance values are from approximately 1000 to 20,000 ohms. In general the resistance is inversely proportional to (or the conductance is directly proportional to) the square of the frequency.

LEAD INDUCTANCE: A connection that goes from a tube element to its external terminal, although usually short in terms of ordinary length measure, has a finite value of inductance. This internal inductance is small enough to have no appreciable effect on the overall circuit performance in most of the hf range (although the inductance of an *external* lead from the tube terminal to the circuit must be small so that the *total* lead inductance is negligible). However, the internal inductance begins to make itself felt, even with a minimum of external lead length, at about the same frequency at which transit time becomes a factor; i.e., in the upper hf and all higher-frequency ranges. The effect of lead length on circuit performance has already been considered in Chapter 17.

In vacuum-tube circuits the inductance of the cathode lead causes the most difficulty, because this inductance usually is common to the input and output circuits. It gives rise to feedback (this is discussed in a later chapter) which introduces a phase shift into the input circuit and has much the same effect as transit time. That is, a resistance component is introduced into the input circuit. The effect of cathode-lead inductance increases with frequency.

Questions and Problems

19-1) What is secondary emission?

19-2) How is power used in heating the plate measured?

19-3) In the circuit of Fig. 19-3, if E_b is 350 volts, R_L is 10,000 ohms, and E_c is chosen so that the plate current is 15 mA, how much power is dissipated in the plate in the absence of an ac signal on the grid?

19-4) Why is the grid-plate capacitance of a triode higher than that of a tetrode or pentode?

19-5) Define transconductance. How is it related to plate resistance and amplification factor?

19-6) Under what conditions does current flow in the grid-cathode circuit of a tube?

19-7) Compare the operating characteristics of triodes having low, medium, and high amplification factors.

19-8) Describe the relative influence of grid voltage, screen voltage, and plate voltage on plate current in a screen-grid tube. How does the plate resistance compare with that of a triode having the same transconductance?

19-9) What is meant by space charge? How does it enter into vacuum-tube operation?

19-10) What is meant by the diode line? What is its significance?

Conduction and Rectification in Semiconductors

WHILE THERE are similarities between ordinary conductors and the materials known as semiconductors, a semiconductor when appropriately processed during manufacture is not simply a partial conductor as the name might seem to imply. It is capable of behaving either as an ordinary conductor or as a rectifier, but the details of operation differ considerably from those of simple conductors and vacuum tubes.

SEMICONDUCTORS: Semiconductors are crystalline materials in which there is always a small number of free electrons, but far fewer of them than in an ordinary conductor. These electrons have become detached from their atoms by having been given sufficient heat energy to break free. The number of free electrons increases with the temperature, consequently the resistivity of the material decreases when its temperature is increased. The resistivity of the pure material is very much higher than that of a good conductor, but not nearly so high as that of a good insulator (which for practical purposes may be considered to be infinitely high).

The most useful semiconductor materials are the elements germanium and silicon.

CONDUCTION IN A SEMICONDUCTOR: Each thermally generated free electron leaves its atom with one electron less than it should have, and the atom therefore has a net positive charge that equals the missing electron's negative charge. This is shown in a highly simplified way in Fig. 20-1. As there are just as many free electrons as there are positively charged atoms, the material as a whole is electrically neutral. If a free electron moving through the space between atoms happens to pass close to one that has lost one of its electrons, it will be attracted by the positive charge and will drop into the spot or hole formerly occupied by the atom's missing electron. This is called recombination.

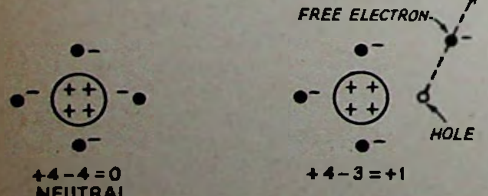


Fig. 20-1

The process of thermal generation and recombination is continuous. Each thermally freed electron leaves a hole behind, a hole that is soon filled by another thermally freed electron which in its turn has left a hole behind. Thus in a pure semiconductor the *positions* of the holes move about. Although only the electrons actually move, it is convenient to think of the holes as moving too; the overall effect is the same either way.

When an electric field is set up inside a semiconductor by an external voltage source, the free electrons drift toward the positively charged terminal. Similarly, the holes drift toward the negatively charged terminal (because electrons are repelled from that end, although it is again convenient to think of the holes as being *attracted* to the negative terminal). There is therefore an electric current in the semiconductor consisting of electrons moving one way and holes moving the opposite way. The electrons and holes are called current carriers.

Since the semiconductor structure is uniform in any direction, current can flow in a manner dictated solely by the electric field; that is, in any direction. In this respect a semiconductor is the same as any ordinary conductor,

DOPING: The conductivity can be greatly increased by adding very small amounts of other elements to the pure (intrinsic) material. The process of doing this is called doping. Doping adds impurity atoms to the intrinsic material in such a way that they fit into the original crystalline pattern, but in doing so they increase either the number of free electrons or the number of holes (but not both). This reduces the resistivity to a relatively low value.

If the doped semiconductor has an excess of free electrons it is called n-type material; if it has an excess of holes it is called p-type. Despite these excesses of one type or the other of carriers, the material remains electrically neutral. The exact way in which this comes about is beyond the scope of this treatment, but can be understood on the basis that the impurity atoms are themselves electrically neutral and therefore do not upset the neutrality of the intrinsic semiconductor.

In n-type material the excess electrons are free to move about while the corresponding positive charges or holes associated with atoms are bound in fixed places in the atomic structure. Similarly, in p-type material the negative charges are fixed in location and the positively charged holes are free to move. Holes in p material are called majority carriers. In n material the majority carriers are electrons.

Doped material is capable of conducting equally well in any direction, and does not differ from intrinsic material in this respect. The resistivity of the doped semiconductor depends on the amount of impurity added; light doping results in relatively high resistivity and heavy doping results in a relatively low value.

THE pn JUNCTION: If a single piece of semiconductor has been doped to be p type in one region while the immediately adjoining region is doped to be n type, the crossover between the n and p types is called a pn junction. Fig. 20-2(A) indicates such a junction. The line between the two types of material must not be thought of as a simple contact; a junction must be formed by opposite types of doping without disturbing the crystalline structure of the intrinsic material.

When a junction is formed some of the electrons from the n side migrate or diffuse across the junction, where they immediately recombine with holes on the p side. These electrons have left behind an excess of bound holes close to the junction on the n side, so a small positive charge remains on the n side of the junction. Similarly, holes from the p side diffuse across the junction, leaving a small negative charge on the p side near the junction. Equilibrium is soon reached because the bound positive charge on the n side repels further hole diffusion, and the bound negative charge on the p side discourages further electron diffusion. Thus there is a small space charge across the junction, and since most of the holes and electrons in this region have recombined there are no carriers to disturb the space charge. This is known as the depletion region.

Electrons which cross a junction from the n side are called minority carriers when they are in the p side, and holes that diffuse into the n side are minority carriers in the n material.

CONDUCTION AND NONCONDUCTION: If a small voltage is applied to a junction as shown in Fig. 20-3(A), positive side to the p material and negative side to the n material (forward bias) electrons in the n material will be repelled from the negative terminal and holes in the p material will be repelled from the positive terminal. The majority carriers on both sides therefore move toward the junction. If the applied voltage is sufficient to

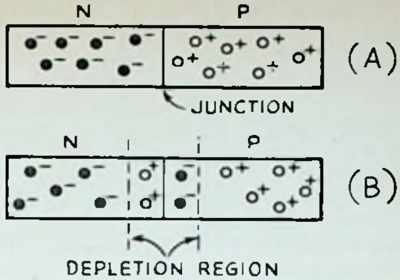


Fig. 20-2

overcome the space charge, some electrons will be propelled across the junction into the p material. (They are said to be injected into the p material.) Once there, they quickly recombine with free holes. Holes from the p side likewise will be injected into the n region, where they recombine with free electrons. This process is continuous so long as forward bias is applied.

Since holes and electrons are continuously moving in opposite directions through the semiconductor a current is flowing through it and the external circuit. (It will be realized that the whole process is equivalent to a stream of electrons moving in *one* direction through the semiconductor because, as mentioned earlier, only the electrons actually move. The hole concept is not used in the external circuit, and only electrons are shown moving there in the figure.)

The voltage needed for overcoming the space charge and thus giving the carriers sufficient velocity to be injected into the opposite type of material is about 0.15 to 0.2 volt if the semiconductor is germanium, and about 0.6 volt if it is silicon. If the voltage is less than this a small current flows across the junction, but it is very much less than the current that flows when the space charge has been overcome. When the forward bias just reaches the space-charge voltage there is a rapid increase in current. The device is said to turn on when this value of forward bias is reached.

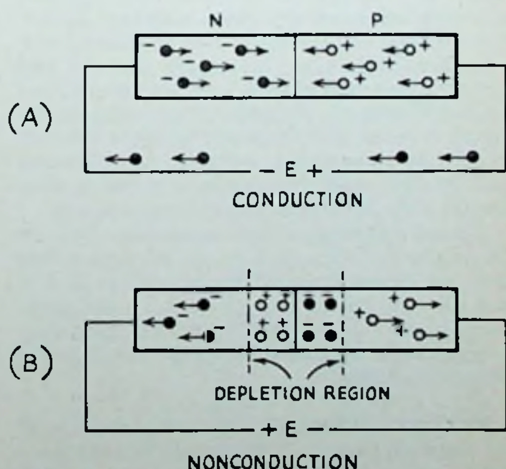


Fig. 20-3

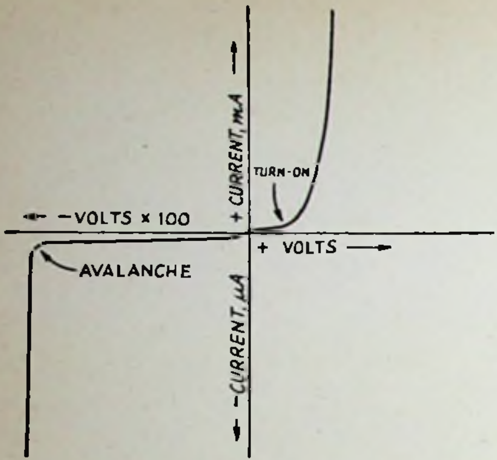


Fig. 20-4

Except for the small region near the junction where recombination is continually going on, the moving holes and electrons are *majority* carriers. However, it is essential for injection to occur, and recombination to take place, before there can be any conduction.

If the polarity of the applied voltage is reversed (reverse bias) as in Fig. 20-3(B), the direction of movement of the holes and electrons also reverses. Neither can now cross the junction; in effect, the depletion region has been widened. Since there is now a gap which the carriers cannot cross, no current flows in the circuit, and the current is said to be turned off.

Thus the pn junction permits conduction in only one direction, and therefore is a rectifier. Since there are only two elements, it is a diode.

STORAGE AND LIFETIME: When forward bias on a diode is suddenly reduced to zero some minority carriers are left on both sides of the junction. In terms of fractions of microseconds, a considerable length of time may elapse before these carriers recombine with majority carriers. The minority carriers are said to be stored under these conditions, and the length of time that elapses before most of them have been recombined (all but 37 percent, according to the usual specification) is called their lifetime. The stored minority carriers permit current to continue flowing during their lifetime, and thus conduction ceases some time *after* the diode is no longer forward-biased.

This effect becomes important at radio frequencies, where a change of state can occur each half cycle. In diodes intended to be used at rf it is necessary to reduce storage time to the point where it will not deteriorate circuit performance. Storage can be controlled by suitable processing of the semiconductor.

REVERSE CURRENT: Thermal generation of free electrons goes on continuously in a semiconductor, and if a diode is reverse-biased so that normal majority-carrier current cannot flow a certain number of electrons generated thermally on the p side (where they are minority carriers) near the junction will be attracted by the positive space charge in the n region and will cross over the junction. Similarly, holes generated thermally on the n side will be attracted across the junction by the negative space charge on the p side. Thus a small minority-carrier current flows through the circuit. This current is variously known as leakage current, reverse saturation current, or simply reverse current. It flows in the opposite direction to the majority-carrier current that flows with forward bias.

Being caused by thermal effects, leakage-current amplitude is essentially constant, at a given temperature, no matter what the reverse bias may be. It doubles, approximately, for each 10 deg C (18 deg F) rise in temperature (slightly less than this in silicon and slightly more in germanium). The leakage current at a *fixed* temperature is of the order of 100 or more times smaller in silicon than in germanium. This makes silicon much the preferable material in the many applications of semiconductor devices in which leakage current is an important factor.

AVALANCHE: As the reverse bias on a pn junction is increased, the current at first is the leakage current just described. This current remains practically constant up to a certain reverse voltage known as the avalanche voltage. At this point a further small rise in reverse voltage will cause the reverse current to become abnormally large (avalanche breakdown) unless there is sufficient resistance in the external circuit to limit it.

The reason for avalanche breakdown is found in the minority carriers that constitute the leakage current. Increasing the reverse voltage increases their velocity, and when the velocity becomes high, speeded-up carriers colliding with atoms splash out more carriers. These in turn acquire sufficient velocity to cause further ionization by collision, and the process can continue until the current reaches exceptionally large values with little or no change in reverse bias.

Avalanche breakdown will cause the destruction of a diode if not controlled. However, with suitable control by the external circuit it can be turned to advantage, since the voltage across the diode will be substantially constant over a wide range of currents. This feature is used in many semiconductor applications, particularly in voltage-regulator circuits, with diodes especially made for the purpose (Zener diodes).

The voltage at which avalanche occurs depends on the doping. Light doping (high resistivity) increases it; heavy doping (low resistivity) decreases it.

DIODE CHARACTERISTICS: The behavior of current as the applied voltage is varied can be shown by the characteristic curve of the semiconductor diode. A typical curve is shown in Fig. 20-4. In this drawing positive current (increasing vertically upward from zero) means current flow by majority carriers when the diode is forward biased. Reverse current caused by minority carriers is drawn below the horizontal axis and increases downward; it flows when the diode is reverse biased. (Note the change in scale for forward and reverse voltages and currents.) The positive curve bends sharply upward when the space charge has been overcome; as mentioned earlier, this occurs when the forward bias is about 0.15 volt for germanium and about 0.6 volt for silicon. The sharp downward bend in reverse current occurs at the avalanche voltage.

The forward resistance (dynamic) of the diode is found from the slope of the curve (see Chapter 19) on the right-hand side of the drawing. The slope is small (high resistance) at first. When the bend is reached the resistance decreases rapidly and becomes quite low in the region where the current is appreciable. Resistances in this region of 100 ohms or less are common, depending on the type of diode.

Actual values in such a curve vary with the semiconductor material and the construction and processing of the diode. A wide range of performance characteristics is available, depending on the purpose for which the diode is intended. The current at a particular voltage depends on temperature; it is larger in both the forward and reverse directions when the temperature of the diode is increased.

DIODE CAPACITANCE: In a pn junction the p and n materials are conductors on either side of the depletion layer, which is practically nonconducting when there is no bias or reverse bias. Thus the assembly is equivalent to a capacitor, the capacitance being dependent on the area and separation of the conducting plates. If reverse bias is applied the depletion layer widens; hence increasing the reverse voltage decreases the capacitance. (This effect is used in the varactor, a voltage sensitive capacitor used in some radio communication circuits.)

Because the depletion layer is extremely thin the diode is, in effect, shunted by a capacitance that, at radio frequencies, has a reactance that is too low to be ignored in comparison with the diode's resistance. This capacitance limits the frequency at which a given diode can be used effectively. The capacitance can be reduced by making the junction area as small as possible; this is done in the point-contact diode, in which n-type material (usually germanium) has a wire point resting on its surface. During manufacture a tiny spot of p-type is formed under the point, resulting in a diode having a capacitance of about 1 pF. Although the amount of current that can be carried safely by such a junction is relatively small (about 50 mA) it is sufficient for many applications, including radio-frequency circuits. In the latter the rectified current is usually only a few mA at most.

RECTIFIER RATINGS: There are two classifications of diodes, based on size and type of application. The very low power rectifying devices used in much radio-communication work and in computer service usually are called "diodes." In larger sizes, suitable mostly for low-frequency power-supply applications, they are called "rectifiers." With diodes, generally speaking, operating voltages and currents are far below what the device can handle safely. In the application of rectifiers the safe ratings become important.

Ratings can be explained with the aid of Fig. 20-5, which shows a basic rectifier circuit. Alternating current is changed into direct current by rectifying the positive half-cycles (in this diagram) of ac. (The p side is the anode and the n side is the cathode.) The rectified current charges a capacitor, C, across which the load, represented by resistor R_L, is connected. Energy is stored in the capacitance whenever the voltage applied to the rectifier is increasing in the positive direction and concurrently exceeds the voltage already at the capacitor terminals. When the applied voltage is decreasing from the positive maximum, and also during the entire negative half cycle, the capacitor discharges into the load resistance. The average voltage across the load therefore depends on the ratio of capacitance to load resistance. This ratio is generally made large enough so that the instantaneous variations in voltage across the load (called **ripple voltage**) are quite small compared with the average dc voltage.

When C is large compared to R_L (that is, when the time constant of the CR_L circuit is of the order of 10 times the ac period) the average output voltage, E_{DC}, will be very nearly equal to the peak ac voltage. Thus $E_{DC} \approx 1.4 E_{AC}$, where E_{AC} is the rms value of voltage applied to the rectifier (assuming the applied ac voltage to be substantially sinusoidal). Considering just the series circuit formed by the ac voltage source, the rectifier, and

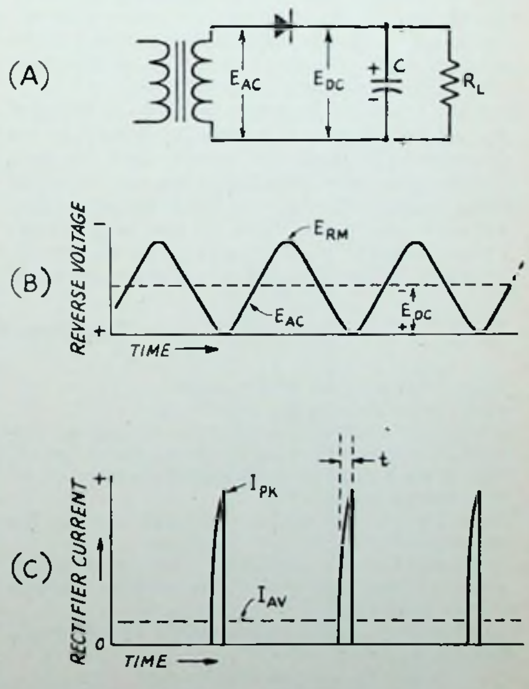


Fig. 20-5

C, and using the positive dc terminal of the capacitor as the reference, it can be seen that at the peak of the *negative* half cycle the voltage across the diode is equal to the sum of the voltage stored in C and the voltage at the negative peak, as shown in Fig. 20-5(B). This sum, E_{RM}, is a reverse voltage applied directly to the rectifier, and is called the peak reverse voltage (PRV). In practice, the peak reverse voltage applied to the rectifier must be kept safely below the voltage at which avalanche occurs. Single rectifiers are available with various PRV ratings up to about 1000 volts. If the required PRV rating is higher, rectifiers can be connected in series so that the reverse voltage applied to every rectifier in the "string" will not exceed its rating.

The forward current flowing into the capacitor and load under the same conditions is illustrated at (C). Since the output dc voltage is very nearly equal to the peak positive ac voltage when C is large compared to R_L, current flows through the rectifier only during a small time interval, as indicated by the time, t, between the vertical dashes.

Also, all the energy used by the load must go through the rectifier, so the sum of all the instantaneous values in one current pulse, divided by the time of one cycle, must equal the average load current. Thus the shorter the time interval during which current flows through the rectifier the higher the peak current must be, for a given average current. As there is a pulse for each ac cycle, in the particular circuit shown in Fig. 20-5(A), its peak value is called the repetitive peak current.

Rectifiers are limited in current-carrying capacity by the rise in junction temperature caused by power lost in the rectifier. The average power dissipated in the rectifier is the average of the product of the instantaneous voltage drop and the instantaneous current. The voltage drop is small – about 0.6 to 1 volt in silicon rectifiers, so the rated current can be relatively large for a given power dissipation. Rectifiers are usually rated for both average and peak repetitive current, the latter rating being of the order of five times the average rating in most cases. There are also surge-current ratings, usually; surge current is an abnormally large current that lasts for only a few ac cycles, at most, under transient conditions such as may occur when the power is turned on or off.

The rectifier circuit used as an illustration is only one of several, although it is perhaps more commonly used than most others, either in the basic form shown here or in a series arrangement called a voltage doubler. In general, the relationship between PRV and applied ac voltage, and between peak repetitive current and average current, will depend on the exact form of the rectifier circuit used, and also on the filter circuit which smooths the ripple in output voltage.

HEAT DISSIPATION: The rise in temperature in a semiconductor that occurs because of heat generated internally during operation can be a serious limitation. As stated earlier, both the forward and reverse currents will be larger, for a given applied voltage, when the temperature rises. Forward current presents no problems in rectifier applications since it is limited by the circuit resistance. However, the reverse current is practically independent of everything except temperature, and because the resistance of the diode in the reverse or back direction is high, a relatively small reverse current can cause a relatively large amount of heat to be generated. The resulting rise in temperature in turn increases the reverse current, a process which becomes cumulative. If allowed to go unchecked, it will eventually result in breakdown and destruction of the device through thermal runaway.

If the power to be dissipated is appreciable – of the order of a half watt or greater – it is usually necessary to provide some means for carrying heat away. This is true of all semiconductor devices. Cooling by simple air circulation is seldom effective enough. The larger rectifiers (and transistors) have metal cases to which one of the elements is connected for maximum heat transfer to the case. The case itself is then equipped with or mounted on a heat sink, a metal structure having good heat conductivity and a large surface for both radiation and air convection of the generated heat. With an adequate heat sink making good thermal contact it is possible to maintain an equilibrium temperature in the semiconductor that will permit continuous operation at normal ratings.

Questions and Problems

20-1) What is meant by avalanche?

20-2) What is a minority carrier?

20-3) If a rectifier circuit such as that in Fig. 20-5(A) is operating with an rms ac voltage of 75, what is the maximum PRV that should be expected in normal operation?

20-4) What is the effect of storage time on the operation of a rectifier?

20-5) What is the difference between p and n material? How does each differ from an intrinsic semiconductor?

20-6) When is a heat sink necessary in operating a semiconductor device?

20-7) Define forward and reverse bias.

20-8) Why is the resistivity of a semiconductor dependent on temperature? Which material shows the greater temperature dependence, silicon or germanium?

20-9) Other things being equal, which material has the greater leakage current in a semiconductor device, germanium or silicon?

20-10) Why is leakage current so important in a diode?

20-11) Describe how conduction can take place across a pn junction.

20-12) What is meant by the depletion region of a junction? On what does it depend?

20-13) How is it possible for a pn junction to cause rectification of an ac voltage?

20-14) What is space charge? About what value does it have?

Chapter 21

The junction field-effect transistor; junction FET characteristics; amplification; the insulated-gate FET; interelement capacitances; the dual-gate MOSFET; temperature effects in FETs.

Field-Effect Transistors

THE transistor is a semiconductor device so constructed that a signal applied to one pair of terminals will appear at another pair in amplified form. There are two principal classes of transistors, the field-effect type and the bipolar type.

In several ways the field-effect transistor resembles the vacuum tube in both principles and operation, although the transistor's operating voltages are very much smaller. Because the vacuum tube has already been discussed (Chapter 19) as well as the fact that the field-effect type is easier to understand than the bipolar type, discussion of the bipolar transistor is deferred until the following chapter.

THE JUNCTION FIELD-EFFECT TRANSISTOR: A piece of doped semiconductor material of the general form shown in cross section in Fig. 21-1, having a central section with one type of doping and the two outer sections the opposite type, is called a field-effect transistor (FET). Since the outer sections form junctions with the inner section, it is a junction-type FET. In Fig. 21-1 the inner section is n-type material and the outer sections are p-type, but the opposite arrangement also can be used – p inner section and n outer sections.

The inner section is called the channel and the outer sections, usually connected together internally, form the gate. Direct connections are made to the ends of the channel, one being designated the source and the other the drain. When a voltage is applied between the drain and source a current (drain current) will flow through the channel, its amplitude depending on the channel resistance. The polarity of the drain-source voltage is such that the majority carriers (electrons in the case of the n channel shown) flow from the source to the drain.

In the absence of bias voltage between the gate and the channel the usual depletion regions are formed at each junction. If reverse bias is now applied between the gate and source the total depletion region will widen, thereby reducing the conducting cross-section of the channel. This increases the channel resistance. The change in resistance is a function of the amplitude of the reverse bias on the gate, and if the bias is made large enough the depletion region will extend across the channel. The channel resistance then becomes very high. The bias voltage that reduces the drain current practically to zero is called the pinch-off voltage.

If *forward* bias sufficient to overcome the junction space charge is used, the gate loses its control over the channel resistance. The device then acts like an ordinary diode in the conducting region. Consequently, this type of transistor is always operated with reverse bias. (Actually, diode conduction is only partial so long as forward bias is below the junction space-charge voltage – about 0.6 volt for silicon – but even this forward-bias region is avoided in most amplifier applications.)

Over the range of gate-bias voltages for which diode conduction does not occur, the gate controls the drain current in much the same way that the grid controls plate current in a negative-grid vacuum tube. Also, since there is very little current flow across the gate-channel junction when the bias is below the space-charge voltage, the input resistance of the FET is very high. The current that does cross the junction is the leakage current described in Chapter 20, and with a silicon

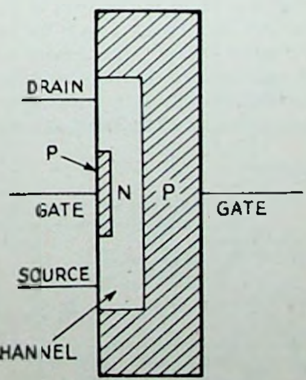


Fig. 21-1

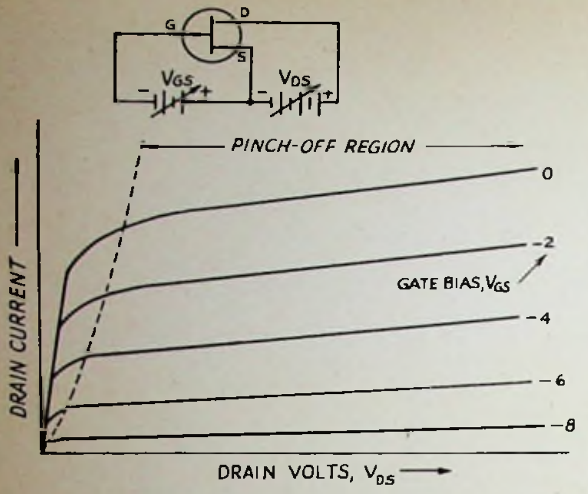


Fig. 21-2

semiconductor this is small enough so that at normal operating temperatures the input resistance is of the order of 5 to 10 megohms. (Because of its low leakage as well as other characteristics, silicon is always used as the basic semiconductor in a field-effect transistor.) However, the input resistance drops to a low value when appreciable diode current flows across the gate-channel junction.

JUNCTION FET CHARACTERISTICS: A set of curves showing how the drain current of a junction FET varies as the drain-to-source voltage is increased from zero is shown in Fig. 21-2. Curves for a number of values of fixed reverse bias on the gate are given. This set of curves resembles the plate family of a vacuum tube, and the general shape of the curves is similar to that of a pentode.

It can be seen from the small diagram that the gate-to-channel junction is reverse biased by the gate-to-drain voltage as well as by the gate-to-source voltage, since these voltages are in series. Consequently, if the drain voltage is increased, the bias between gate and source being fixed, the depletion region widens in the part of the channel between the gate and drain. This is the reason why the curves have a "knee": the depletion effect of the drain-to-gate voltage begins to predominate at this point and a further increase in drain voltage has relatively little effect on the drain current. The section to the right of the knee is called the pinch-off region. The transistor normally is operated in this region in amplifier applications. If the drain voltage is increased indefinitely the depletion region will extend all the way to the drain and a breakdown point will be reached. At the breakdown voltage the drain current increases abruptly and transistor action ceases. In operating the transistor the drain voltage, fixed or instantaneous, must be kept below the breakdown voltage.

Because the drain voltage has very little effect on the drain current in the pinch-off region, the output resistance (corresponding to plate resistance in a vacuum tube) is high relative to the resistance that would be suitable for a load. Its value depends on the gate and drain voltages and the drain current, and is measured in the same way as tube plate resistance; i.e., by the ratio of a small change in drain voltage to the corresponding small change in drain current, the gate bias being fixed. Values vary from a few thousand to a few tens of thousands of ohms when measured at the operating points that would be used in practice.

The transconductance of a field-effect transistor has the same significance that it has with vacuum tubes. That is, it is the ratio of a small change in drain current to the small change in gate voltage that produces the current change, the drain voltage being constant. The symbol for it is g_{fs} (the subscript f means forward – that is, from the input side to the output side – and s means that the source terminal is the common point between the input and output circuits). Transconductances in small FETs vary from about 1500 to 10,000 micromhos (1.5 to 10 millimhos) depending on the operating point at which they are measured. In general, the transconductance increases with increasing drain current, and is approximately zero when the gate voltage is near pinch-off. The drain current is determined almost entirely by the gate voltage and relatively little by the drain voltage, when operation is in the pinch-off region, so the transconductance is practically constant, for a given gate voltage, when the drain voltage is varied throughout this region.

AMPLIFICATION: Amplification with a field-effect transistor is very similar to amplification with a negitive-grid vacuum tube, and particularly with a pentode-type tube. A simplified amplifier circuit is given in Fig. 21-3, which may be compared with Fig. 19-3, Chapter 19. The fixed or operating gate bias is applied between the source and gate through a signal source here represented by a generator, G. The signal source must provide a dc path for the bias voltage, V_{GS}; if it does not, other means must be used to apply the bias to the gate without unduly loading the signal source. The ac signal voltage, e_G, causes variations in drain current which in turn cause amplified voltage variations to appear across R_L. This ac voltage, which is superimposed on the dc voltage drop in R_L (this drop is caused by the dc drain current obtained from the drain supply, V_{DS}) is an enlarged replica of the ac voltage applied to the gate.

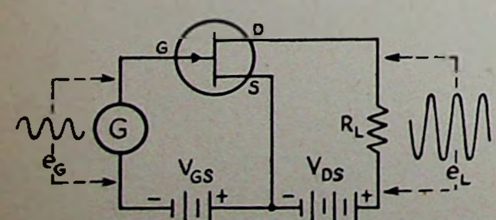


Fig. 21-3

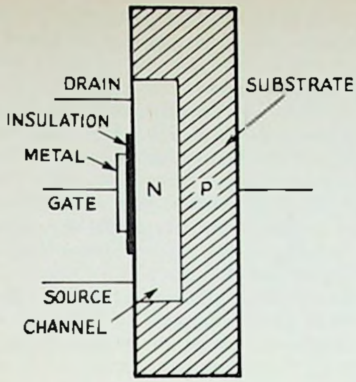


Fig. 21-4

THE INSULATED-GATE FET: In the insulated-gate field-effect transistor or MOSFET (the letters MOS stand for metal-oxide-semiconductor to distinguish this type from the junction FET) the gate is a metal electrode separated from the channel by an extremely thin layer of an oxide of silicon, a form of glass. The cross-section is shown in Fig. 21-4.

The channel is formed in a lightly doped piece of the opposite-type semiconductor called the substrate. An n-channel MOSFET is shown; the channel can be either n or p type, but in the latter case the substrate would be n type. The junction between the substrate and channel has the normal characteristics of a pn junction of comparable construction and processing. Under the usual operating conditions the drain is biased so that it attracts majority carriers in the channel, just as in the junction FET, and the substrate is connected to the source. Thus the channel-substrate junction is reverse biased by the drain-to-source voltage.

The gate in the MOSFET controls the resistance of the channel by inducing a charge in it on the side of the insulating layer opposite the metal electrode (the gate, insulating layer, and channel form a small capacitor). If the channel is heavily doped so its resistivity is low, as is done in the junction type just discussed, a relatively large majority-carrier current will flow between the source and drain when the gate bias is zero with respect to the source. If the gate is now negatively biased, a positive charge will be induced in the n channel near the gate. This causes a depletion region to be formed under the gate and the channel resistance increases. When the negative bias is made large enough, pinch-off voltage is reached and the depletion region extends across the channel. On the other hand, if the gate bias is made *positive* with respect to the source, more electrons are forced into the channel and the resistance is lowered.

Positive bias does not cause current to flow between the gate and channel, unlike the junction FET; the glass insulation prevents conduction. This feature of the MOSFET is a principal difference between it and the junction type. In this respect the MOSFET also differs from a vacuum tube; in the tube, grid current will flow whenever the grid becomes positive with respect to the cathode. The gate-to-channel resistance of a MOSFET with either bias polarity is extremely high – of the order of a million megohms – and is chiefly the result of surface leakage across the insulation from the gate to the drain and source.

The type of MOS transistor just described is known as a depletion-type FET because the gate bias which sets the operating point is normally such as to decrease the drain current below its zero-bias value. In another version, the enhancement type, the channel has relatively high resistivity and little or no current flows when the gate and source are at the same potential. With this type the bias for normal conduction is always in the region that *increases* the drain current. The enhancement type, too, can be made with either an n or a p channel. The enhancement FET is used principally in switching operations, the depletion type being favored for communications work.

The characteristic curves of a depletion-type MOSFET are quite similar to those of the junction FET, except that both polarities of gate voltage can be used. Fig. 21-5 is typical of the relationship between drain voltage, drain current, and gate voltage. The output resistance and transconductance are about the same as for the junction type. The input resistance, however, is much higher, as already mentioned.

The dashed line in Fig. 21-5 is a constant-dissipation locus, in this case for a rated dissipation of 150 mW. This value is representative of a small-signal MOSFET intended for use in low-level rf amplifier applications.

INTERELECTRODE CAPACITANCES: The capacitances in a field-effect transistor are rather complex, consisting of distributed capacitances inside the transistor itself as well as direct capacitances between leads and electrodes outside the semiconductor but part of the complete assembly inside the transistor case. For practical purposes they may be divided into three groups, the input

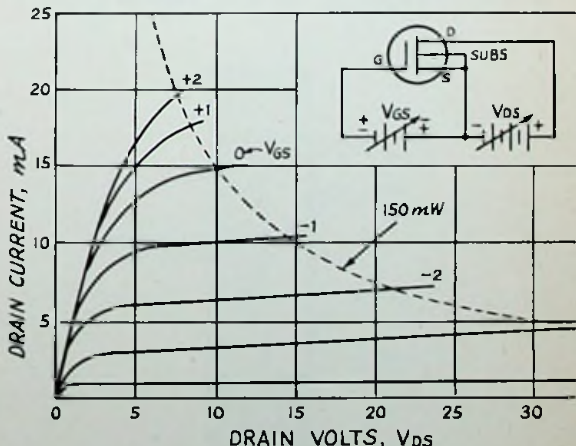


Fig. 21-5

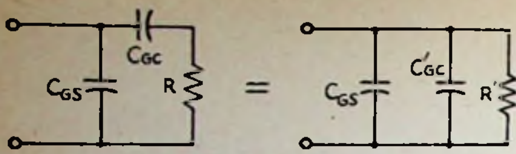


Fig. 21-6

capacitance (gate to source), the output capacitance (drain to source), and the reverse transfer capacitance (drain to gate). These correspond approximately to the grid-to-cathode, plate-to-cathode, and plate-to-grid capacitances of a triode vacuum tube.

The input circuit of the FET, despite the extremely low gate-to-channel leakage, is not a simple capacitance but is an impedance having the simplified form shown in Fig. 21-6. In this circuit C_{GS} is the capacitance between the leads, and is constant with frequency. The capacitance between the gate and channel, C_{GC}, is in series with the resistance, R, of the channel section between C_{GC} and the source. The impedance of this series circuit, $C_{GC}R$, varies with frequency in the way described in an earlier chapter. In addition, the values vary with the voltages applied to the transistor. The series circuit can be represented by the equivalent parallel circuit shown at the right, it being understood that the equivalent parallel capacitance, C'_{GC}, and resistance, R', are dependent on frequency and voltage. At frequencies below the vhf range the reactance of C_{GC} is quite high compared with R, so the equivalent circuit at these frequencies is essentially a simple capacitance, R' being very large under such conditions. The input capacitance at low radio frequencies is 5 to 10 pF, approximately. At vhf an input admittance is usually specified, including the effect of resistance along with capacitance.

The gate-to-drain capacitance is generally in the neighborhood of 0.1 to 0.2 pF in a MOSFET, and the output capacitance is 1 to 2 pF.

THE DUAL-GATE MOSFET: MOS transistors also are made with two gates when an additional control element is needed for special purposes. In transistors of this type either gate can be used to control the channel resistance, or both can be used in combination.

The internal arrangement of the dual-gate MOSFET is as shown in Fig. 21-7. It is useful to think of it as two single-gate transistors in series, with the section of the channel between the two gates acting as the drain for the first unit (controlled by Gate No. 1) and as the source for the second unit (controlled by Gate No. 2). Pinch-off voltage on either gate pinches off the entire channel, so either gate can control the drain current. The signal to be amplified is usually applied to Gate No. 1. Gate No. 2 can then be used to set the drain current, and thereby the transconductance, of Gate No. 1. An adjustable dc bias can be used on Gate No. 2 for gain control, as an example, or an ac voltage can be applied to it to vary the transconductance of Gate No. 1 at an ac rate for mixing two signals or modulating one by the other.

Fig. 21-8 illustrates the way in which the drain current is varied by the bias voltage on Gate No. 2.

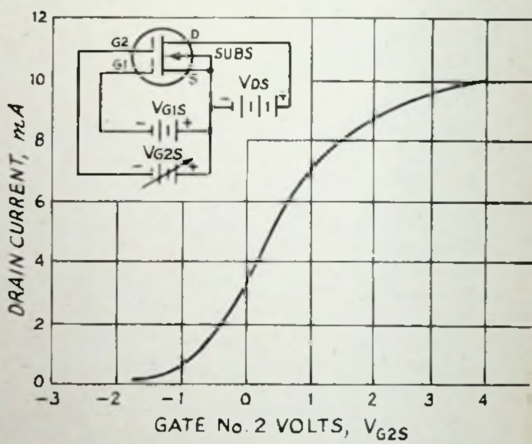


Fig. 21-8

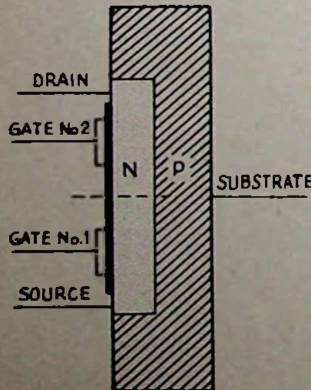


Fig. 21-7

Actual drain-current values depend on the bias on Gate No. 1; in this case, Gate No. 1 is biased so that the drain current is 10 mA when the bias on Gate No. 2 is 4 volts positive with respect to the source terminal.

The shape of the Gate No. 1 transconductance curve is similar to Fig. 21-8. The transconductance has a value of zero at zero drain current and reaches, typically, about 10 millimhos with a drain current of 10 mA.

Grounding Gate No. 2 for the ac signal on Gate No. 1 helps reduce the capacitance between Gate No. 1 and the drain, a useful feature in rf amplification. (Gate No. 2 can be grounded for ac by connecting a bypass capacitor between it and the source, if the dc potential on the source differs from that on Gate No. 2.) The capacitance between Gate No. 1 and the drain can be reduced to as little as 0.01 or 0.02 pF when this is done. The input and output capacitances are approximately the same as for the single-gate MOSFET.

TEMPERATURE EFFECTS IN FETS: In the junction FET the usual temperature considerations apply to reverse current across the junction. Increasing the junction temperature will decrease the input resistance. However, the FET is usually operated far enough below dissipation ratings so that the input resistance remains high, if the transistor is operated with the recommended drain voltage and current and has reasonable provisions for carrying off the heat generated at the junction.

The insulation between the metal gate and the channel in the MOSFET prevents reverse current flow. Although the surface leakage mentioned earlier is affected to some extent by temperature, the change in input resistance is not serious.

In both types thermal generation of carriers takes place in the channel. A rise in temperature therefore will decrease the channel resistance, which causes the drain current to rise somewhat for a fixed gate bias. In terms of percentage, the rise in drain current is greatest when the operating drain current is small; the percentage rise decreases with larger values of drain current. The current change can be made negligible by obtaining the gate bias from a resistor connected in series with the source. The principle is similar to the case of the emitter resistance used with a bipolar transistor as described in the following chapter.

Questions and Problems

21-1) Describe the construction of an insulated-gate-type field-effect transistor.

21-2) In what way does the junction type differ from the insulated-gate type?

21-3) What part do minority carriers play in the operation of field-effect transistors?

21-4) Define pinch-off voltage.

21-5) What is meant by the pinch-off region?

21-6) What happens when a junction FET is forward biased?

21-7) Compare the operation of FETs with that of vacuum tubes.

21-8) What is the substrate?

21-9) Define gate, drain, source, channel.

21-10) What is the purpose of the second gate in a dual-gate MOSFET?

21-11) What is the difference between n-channel and p-channel FETs?

21-12) What is the difference between depletion-type and enhancement-type MOSFETs?

21-13) How do the gate voltages affect the drain current in a dual-gate MOSFET?

21-14) Why is the input impedance of an FET different from its dc input resistance?

21-15) Is the input impedance constant with frequency?

21-16) Is the operation of a field-effect transistor subject to variations with temperature?

The bipolar transistor; conduction in a bipolar transistor; control of current; amplification; bipolar transistor characteristics; transfer characteristic; collector family; leakage current; bias stabilization; interelectrode capacitances.

Bipolar Transistors

THE OPERATING principles of the bipolar transistor differ in several respects from those of the field-effect type described in the preceding chapter. In the FET only majority carriers are used, but in the bipolar transistor amplification depends on both majority and minority carriers. Because of this the transistor's operating characteristics do not duplicate those of the FET.

THE BIPOLAR TRANSISTOR: The bipolar transistor consists of two pn junctions or diodes back to back, the "back" being common to both diodes. That is, the transistor is a piece of semiconductor material that has been doped in one of the two ways shown in Fig. 22-1. The intermediate region, which is very thin (less than 0.001 inch) is doped oppositely to the outer sections; thus the structure may be doped in either npn or pnp sequence. The three regions are called the emitter, base, and collector, as shown in the drawing. Although both the emitter and collector have the same type of doping, the emitter is doped much more heavily than the collector and thus has much lower resistivity.

If there is no bias on these junctions, diffusion and recombination will occur and space charges will form on both sides of each, as described in Chapter 20. Also, if one junction alone is biased, the rectifying action will be essentially the same as for any diode of the same construction.

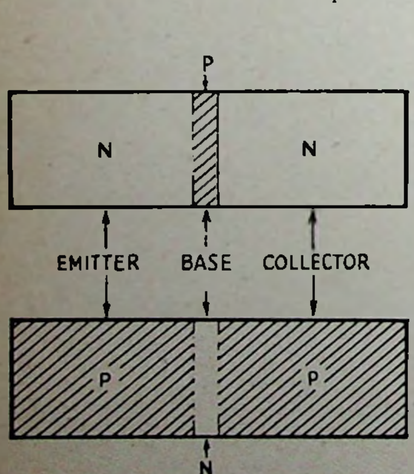


Fig. 22-1

CONDUCTION IN A BIPOLAR TRANSISTOR: When both junctions are biased, the behavior can best be described by using one of the two types shown in Fig. 22-1 as an example. If the doping sequence is npn, as in Fig. 22-2, forward bias on the emitter-base diode will cause electrons from the emitter to be injected into the base, where they become minority carriers. Because the emitter region is heavily doped a great many electrons are injected into the base, and most of them do not immediately find holes with which to recombine, since the base is extremely thin and less heavily doped. Now if the collector is *reverse* biased, the positive space charge on the collector side of the base-collector junction is reinforced by the battery potential, and injected electrons which have not had a chance to recombine in the base are attracted across the junction into the collector. Once there they are majority carriers again, and are attracted to the positive collector terminal. Thus a current (collector current) flows through the collector and external circuit.

From the above description it can be seen that the current flowing through the emitter (emitter current) consists of two components. One is a small diode current flowing through the base lead (the base current). The second, and much the larger, is the collector current which, depending on the transistor processing, can be as much as 99% or even more of the total current leaving the emitter. The proportion that flows out through the base lead is the remainder, and amounts to no more than a few percent, at most, of the total. The ratio of collector current to emitter current is designated by the symbol α.

If holes are substituted for electrons, the above description also applies to the pnp transistor. Both types operate on the same principle, but in the one case the majority carriers in the emitter and collector are electrons and in the other they are holes. Transistors of either type are called bipolar because both majority and minority carriers take part in conduction.

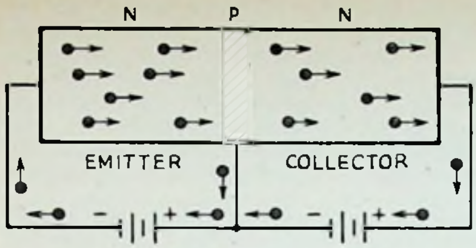


Fig. 22-2

CONTROL OF CURRENT: If the emitter-base diode is *reverse* biased, or if the forward bias is insufficient to overcome the diode space charge, little or no current leaves the emitter. When this is the case the collector current is also negligible. (This statement neglects leakage current, which will be considered later.) However, if the forward bias is increased so that the emitter-base diode begins conducting, the emitter current increases rapidly and, because more carriers cross the emitter-base junction, the collector current also increases. The amplitude of the collector current depends directly on the amplitude of the emitter current so long as α is constant. This is essentially the case over a considerable operating range of currents.

The ratio of collector current to base current is called the current gain of the transistor and is designated by the symbol β. Values of β may range from 20 or so to a few hundred, depending on the construction and processing of the transistor.

Since, by definition,

$$\alpha = \frac{I_C}{I_E}$$

where I_C and I_E are the collector and emitter currents, respectively, and

$$I_B = I_E - I_C,$$

where I_B is the base current, substitution gives the relationship between β and α as

$$\beta = \frac{\alpha}{1 - \alpha}.$$

Current gain is usually designated h_{FE} in transistor data sheets when dc values are used. If the current gain is for ac signals the designation is h_{fe}.

The emitter-base forward-bias voltage at which appreciable collector current begins to flow is that which just overcomes the space charge associated with the emitter-base diode. As explained in Chapter 20, this depends on the semiconductor material, and is about 0.15 volt for germanium and 0.6 volt for silicon, the figures being the same as for ordinary diodes.

If the bias is below this the transistor is said to be cut off or turned off. It turns on when the bias is at the point where space charge is overcome and the collector current starts to rise.

AMPLIFICATION: Amplification is possible in the bipolar transistor because a small base current, supplied at a low emitter-base forward voltage, causes a large emitter-collector current to flow at a considerably higher value of collector voltage. That is, a small amount of power supplied to the base-emitter diode is responsible for controlling a large amount of power in the collector-emitter diode.

Fig. 22-3 is a simplified amplifier circuit in which the emitter is common to both the base and collector circuits (an npn transistor is shown). The base is forward-biased with respect to the emitter by V_{BE}, thereby establishing a base-bias current. The collector is reverse biased by the supply voltage, V_{CC}. These designations are ones commonly used. The signal source is an ac generator, G, having an internal resistance R_S. The load resistance for the transistor is R_L. An alternating current i_B, supplied by the signal source, flows through the base-emitter diode, being superimposed on the bias current caused by V_{BE}. A much larger alternating current i_C flows in the collector and R_L. If R_L is small compared with the internal collector resistance (this is usually the case) i_C is approximately equal to i_B multiplied by the current gain, β. It is superimposed on the dc collector current (the value of which is established by the base bias) taken from the collector supply, V_{CC}. This current, i_C, develops an ac voltage e_L (equal to $i_C \times R_L$) across R_L.

The circuit of Fig. 22-3 (common-emitter circuit) corresponds to a simple triode amplifier circuit such as is shown in Fig. 19-3, Chapter 19, and makes clear the functional similarity of the emitter, base and collector to the cathode, grid, and plate, respectively, of a vacuum tube. Although any of the three electrodes in the transistor can be used as the reference or "common" point of a circuit, the common-emitter arrangement has a number of features that make it the most desirable one for many applications. The transistor characteristics given in manufacturers' data usually are based on this circuit.

BIPOLAR TRANSISTOR CHARACTERISTICS: A curve of collector current as a function of base current, for a fixed value of collector voltage, is shown in Fig. 22-4. This curve is typical of a low-power germanium transistor (about 250 mW safe dissipation with a reasonably good heat sink). The curve is linear over a large part of its range, indicating that the current gain is constant in this region. As given by the slope of the curve, the

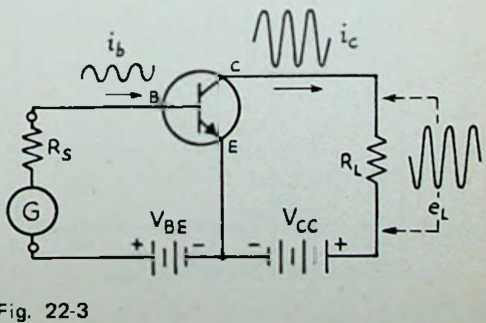


Fig. 22-3

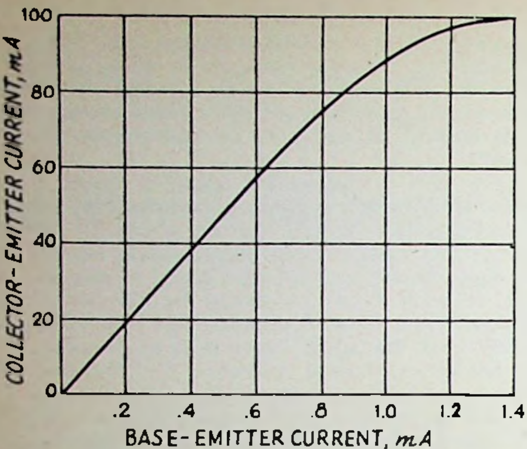


Fig. 22-4

current gain is about 100 in this case; that is, each 100 μA change in base current causes the collector current to change 10 mA.

At the upper end the curve bends away from the straight portion, resulting in a reduction in the collector-current change for the same change in base current. The current gain decreases in this region. Eventually, if carried far enough, the curve would become horizontal, at which point the current gain would be zero and the collector current would reach its maximum possible value.

Exact values on a curve of this type depend on the temperature of the transistor; currents are higher when the temperature is higher. This, in fact, is true of all bipolar transistor curves, and must not be forgotten in establishing operating conditions for transistor amplifiers.

TRANSFER CHARACTERISTIC: Since the base current depends on the voltage between the emitter and base once the space-charge voltage has been overcome, it is possible to show by means of a characteristic curve how the collector current varies as a function of the emitter-base *voltage*. Fig. 22-5 is such a curve (called a transfer characteristic) for a low-power silicon transistor. The current change per unit voltage change is equivalent to a conductance (I/E), and the slope of the curve gives the transconductance of the transistor at some specified value of collector-to-emitter voltage. It is thus similar to the transconductance of a vacuum tube or FET.

Such a transconductance curve for a small silicon transistor is shown in Fig. 22-6. It starts out at zero at the point where conduction just begins, and then rises almost linearly through the normal operating-current range. The transconductance is expressed in millimhos, and in general is much higher in a bipolar transistor than in a tube or field-effect transistor.

Varying the transconductance by varying the base bias offers a convenient means of varying the amplifier gain.

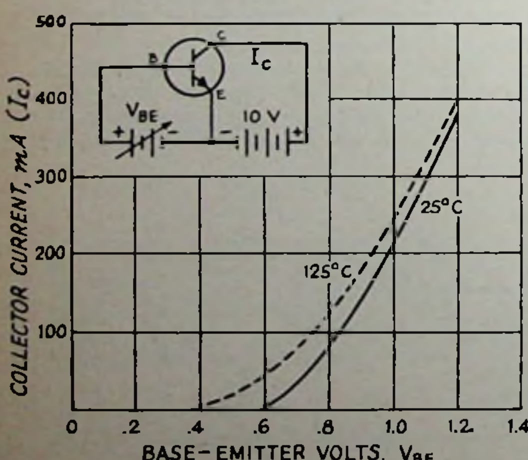


Fig. 22-5

Fig. 22-6

COLLECTOR FAMILY: The set of curves in Fig. 22-7 is the collector family, similar to the plate family for the vacuum tube (see Chapter 19). It shows the collector current as a function of collector voltage for various values of base bias current. These curves are for a medium-power silicon transistor. The fact that the curves are almost horizontal (the slope is small) indicates that the internal collector resistance (output resistance) is quite high.

Another characteristic of interest in circuit design is the relationship between base voltage and current at some selected value of collector-to-emitter voltage. This relationship determines the current and voltage levels, and thus the power level, at which an ac signal must be supplied to the base (input) circuit of an amplifier. Such a curve is shown in Fig. 22-8 for the transistor whose collector family is given in Fig. 22-7. The slope of this curve, at a given point, is the dynamic input resistance (for small signals) at that point.

A bipolar transistor always requires input power, since the ac output depends on ac variations in the base *current*. These cannot occur without a finite value of ac base *voltage* superimposed on the dc base bias. If the transistor of Fig. 22-8 is operating with 10 volts on the collector the rate of base-current change with an operating point set at 0.8 volt on the base will be about 1 mA for each 0.07 volt change. This corresponds to a resistance of 70 ohms. The resistance would be different at other operating points, because the slope of the curve is not constant.

LEAKAGE CURRENT: The leakage current or reverse current (see Chapter 20) across the reverse-biased collector-to-base junction is a major consideration in the operation of bipolar transistors. If the base connection in FIg. 22-2 is opened so that no current can flow out of the base, the reverse current crosses the collector-to-base junction and then continues as a forward current across the base-emitter junction. It is therefore practically indistinguishable from the normal base-emitter current. When the base circuit is completed as in Fig. 22-3, the *total* base current is the sum of the normal base current and the part of the reverse current that flows across the base-emitter junction. (Some of the reverse current will be diverted through the external base circuit, depending on the relationship between the resistance of the base-emitter diode and the resistance of the external base circuit, since the two resistances are in parallel as seen by the reverse current; thus the proportion that adds to the normal base current depends on the circuit conditions.)

This total base current is amplified in the same way as the normal base current; that is, it causes a collector current equal to β times the base current. The presence of leakage current therefore causes the collector current to increase, and since the reverse current depends on temperature, the collector current will rise, when the temperature rises, even though there is no change in the applied voltages.

Leakage currents in a transistor are actually more complex than the description above would indicate, but the overall result is the same no matter how the current originates: The collector and base currents both increase with an increase in temperature, with the former increasing in proportion to the current gain multiplied by the increase in base current. Since the temperature depends on the power dissipation in the transistor, and the power dissipation increases when the collector current rises, the process is cumulative. Eventually, it can lead to the destruction of the transistor through thermal runaway, if allowed to go unchecked.

As stated in Chapter 20, the reverse current with silicon is far less than with germanium, so a silicon transistor can operate normally over a considerably larger temperature range than a germanium transistor can.

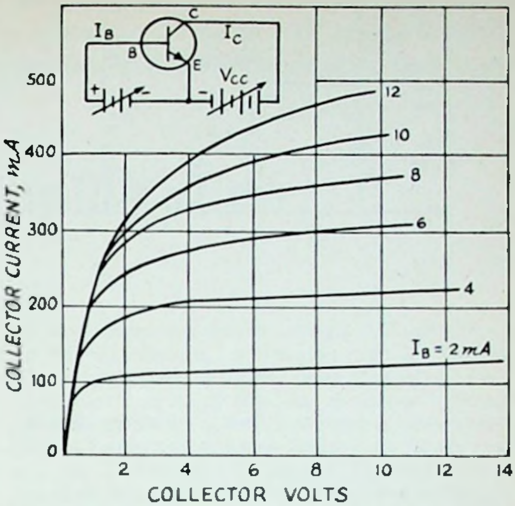


Fig. 22-7

BIAS STABILIZATION: In practical common-emitter circuits it is usually inconvenient to furnish separate voltage sources having low internal resistances (such as separate batteries) for biasing the base and collector. Neither is it always possible to use a signal source having very low internal resistance.

Base bias is usually obtained from the collector supply voltage. This can be done by connecting a resistor from the collector supply to the base to establish the proper bias current, but this method offers no protection against changes in base current because of collector-base leakage.

The necessary stabilization of emitter current is ordinarily obtained by using a voltage divider to supply the base voltage, together with an automatic means for maintaining the collector current constant, or as nearly constant as possible, despite temperature changes. A bias circuit commonly used for this purpose is shown in Fig. 22-9.

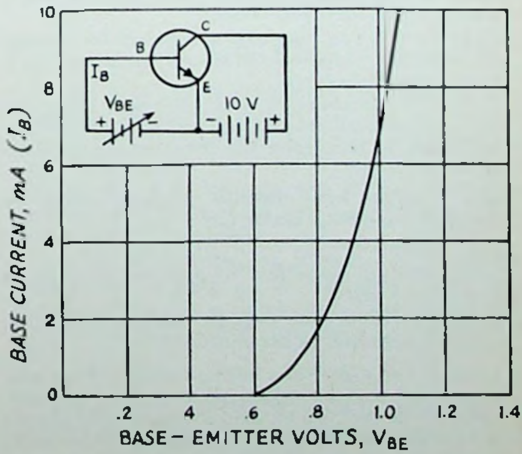


Fig. 22-8

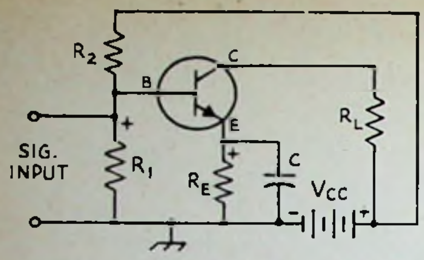


Fig. 22-9

In Fig. 22-9 R_E is a resistor connected from the emitter to the negative or ground side of the collector supply (npn transistor shown; the battery polarity would be reversed for a pnp transistor). The collector current therefore flows through R_E and causes the emitter to be at a positive potential with respect to ground. This potential is equal to $R_E I_C$, where I_C is the collector current (actually, the current is equal to the collector current plus the base current, but the latter, being small in comparison to I_C, can be neglected). This drop reverse-biases the base-emitter diode through R_1. However, R_1 is part of a voltage divider R_1R_2 across V_{CC}, and the voltage drop in R_1 is positive with respect to ground. Since R_E and R_1 are in series between the emitter and base and the voltage drops in the two resistors are opposite in polarity, the net voltage between the base and emitter is the voltage difference. The resistances are selected so that the net voltage forward-biases the base-emitter diode at the value required for setting the desired value of operating collector current.

R_1 is usually large compared with the internal base-emitter resistance, so nearly all the leakage current flows through R_E. An increase in leakage current will cause a much larger change in collector current, and this in turn increases the reverse bias developed in R_E. The forward bias consequently is reduced, and the collector current is restored to practically its original value.

Capacitor C across R_E is a bypass for signal currents that otherwise would flow through the resistor. Without C, the ac voltage drop in R_E would similarly oppose ac changes in collector current and would thereby reduce the amplification.

INTERELECTRODE CAPACITANCES: Capacitances exist between the base and emitter, base and collector, and emitter and collector in a bipolar transistor. Those inside the transistor structure itself are not simple reactances but are associated with the internal resistances along which they are distributed. The internal capacitances depend on voltages and currents and therefore are not constant if these quantities change, as they do during amplification. There are also capacitances of essentially purely reactive nature between parts of the transistor assembly, but outside the semiconductor itself.

Average values of both input and output capacitances (common-emitter circuit) are of the order of 1 pF for small-signal silicon transistors manufactured for high-frequency use. The collector-to-base capacitance (reverse transfer capacitance) is generally less than 1 pF. In power transistors the capacitances are considerably higher, and the output capacitance of the large units may run 100 pF or more.

At low frequencies the reactances (or admittances) are high enough to be of no consequence, but they must be taken into account in transistor amplifiers operating at hf and vhf.

Questions and Problems

22-1) Name the types of carriers that participate in conduction in a bipolar transistor.

22-2) Why is the collector-base junction of a bipolar transistor reverse-biased?

22-3) How much base bias is required for cutting off collector current in a silicon transistor? In a germanium transistor?

22-4) Will the collector pass current if there is a short circuit between the base and emitter? What is the effect of increasing the collector-emitter voltage in this case?

22-5) Explain how thermal runaway could be initiated in a bipolar transistor.

22-6) Define α and β.

22-7) What is the effect of collector current on transconductance?

22-8) How does the internal collector resistance compare with that of the emitter? Why?

22-9) If two transistors having similar ratings, one germanium and one silicon, are biased for the same collector current at 25 deg C, would you expect the currents to be smaller, larger, or the same at 75 deg C? Assume that the bias voltages do not change with temperature.

22-10) If a change of 2 mA in base current causes a change of 130 mA in collector current, what is β? What is α?

22-11) Describe the differences between pnp and npn transistors.

22-12) Why is the collector current larger than the base current in a bipolar transistor?

22-13) What is leakage current? How does it affect transistor operation?

22-14) What is the effect of collector current on current gain?

22-15) Where do minority carriers predominate in a bipolar transistor?

22-16) What is the function of the emitter resistor in a bias circuit?

22-17) What are the majority carriers in an npn transistor? In a pnp transistor?

22-18) Describe how amplification can take place with a bipolar transistor.

22-19) How do the interelectrode capacitances of a bipolar transistor differ from those of a vacuum tube?

22-20) How do the transconductances of FETs and vacuum tubes compare with those of typical bipolar transistors?

Chapter 23

Linear amplification; equivalent circuit; voltage, current and power amplification; small-signal amplification; operating constants; use of characteristic curves; selection of load resistance; distortion; signal source.

Fundamentals of Amplification

THERE ARE several ways of operating amplifiers, depending on the purpose for which the amplifier is to be used. They are all extensions of the basic method, in which the output current flows throughout the entire cycle of the input ac signal voltage. Amplifiers operating in this way are called Class A amplifiers. This and the following chapters discuss some of the elementary principles of Class-A amplification.

LINEAR AMPLIFICATION: If the waveshape of the ac signal undergoes no change during amplification, the amplifier is one of a class called linear amplifiers. Linearity requires that the transfer function from, for example, signal input voltage to signal output current be a straight line when step-by-step values are plotted to a suitable scale as shown in Fig. 23-1. The straightness of the line indicates the required direct proportionality between the input and output quantities plotted.

In electronic amplifying devices there are always dc components in the input and output circuits. The dc in the output circuit is a steady current and voltage by means of which the power required for amplification is supplied. The input dc component, usually called the control-electrode bias, establishes the operating point in conjunction with the dc output component. The signal or ac components can take on any desired amplitudes, when superimposed on the dc, that may be possible within the imposed requirements such as linearity (see Chapter 6). Fig. 23-1 shows hypothetical dc components plotted along the axes of the graph.

For the ac components the operating point, A, establishes a set of zero axes; these are also shown in Fig. 23-1. For linear amplification the operating point may be chosen anywhere along the straight line (or anywhere in the straight portion of a line that is curved over part of its length) so long as all values of instantaneous voltage lie entirely on the straight portion.

When a signal voltage varying with time as shown in the figure is applied, the output current varies correspondingly. The instant at which the signal begins is denoted by B. In the case shown, both voltage and current have ac peak values of 2. The *dc* values swing between 3 and 7, since zero on the ac axes corresponds to 5 on the dc axes.

EQUIVALENT CIRCUIT: A very general and much simplified equivalent circuit of an amplifier operating linearly can be drawn as shown in Fig. 23-2. The circuit ignores the dc bias and power-supply voltages; aside from furnishing the necessary steady voltages and power the dc components do not enter into the amplifier operation. It also ignores the screen voltage for vacuum-tube tetrodes and pentodes, since the screen, aside from setting the tube characteristics and operating point, does not ordinarily take part in the actual amplification of a signal. It is assumed that the entire circuit, including the signal source, G, which has an internal resistance R_S, is linear in the sense that the alternating currents are strictly proportional to the ac voltages. That is, the input and output circuits follow Ohm's Law for ac in the same way as in the steady-state ac circuits described in earlier chapters.

The ac output current, here labeled i_2, is caused by the ac input voltage, e_1, as shown in Fig. 23-1. This current flows through the internal resistance (output resistance) r_o of the amplifying device, and also through the load resistance, R_L. The voltage drop in the complete output circuit is

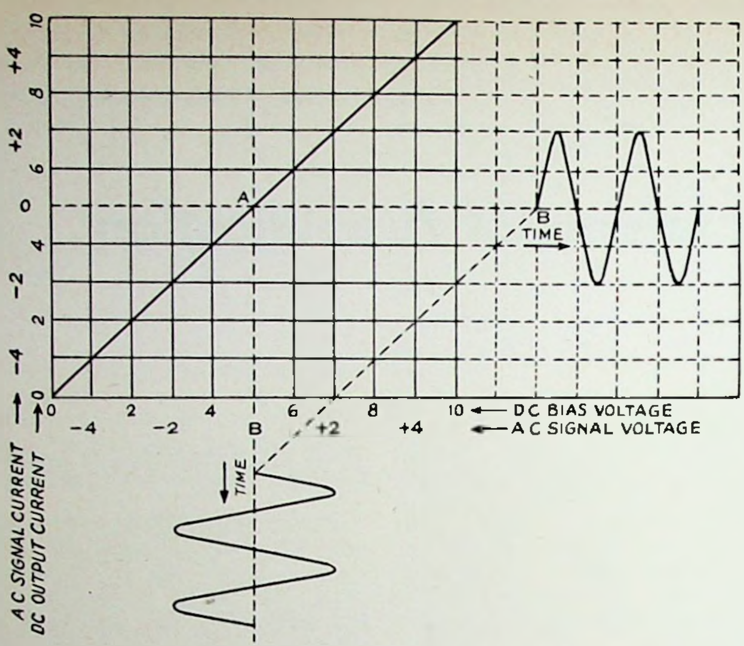
$$i_2 (r_o + R_L).$$

This drop can be considered to be caused by a separate source of voltage, e_2, in the output circuit, with the understanding that its amplitude is not independent but is proportional to the amplitude of e_1 because the output *current* is directly proportional to e_1, once the operating point and R_L are chosen.

Of the total voltage e_2 generated in the output circuit only a part, e_o, is developed in the load resistance, R_L; e_o is called the output voltage, and together with i_2 constitutes the useful output of the amplifier.

VOLTAGE, CURRENT AND POWER AMPLIFICATION: The ratio of output to input, whether

Fig. 23-1



voltage, current or power, measures the amplification or gain of the system. The three ratios are defined as follows;

$$\text{Voltage gain} = \frac{e_o}{e_1}$$

$$\text{Current gain} = \frac{i_2}{i_1}$$

$$\text{Power gain} = \frac{e_o i_2}{e_1 i_1}$$

At low frequencies (up to the frequencies considered to be in the radio spectrum) the input resistance, r_i, of amplifying devices such as the negative-grid vacuum tube, MOSFET, and junction FET when operated with reverse bias on the gate, can be considered to be so high that practically no current flows in the input circuit. Since the input power is negligible under such conditions the power amplification is extremely high. However, this is not true in the rf range (see preceding chapters for discussion of device input impedances) nor is it true of the bipolar transistor at any frequency. In the general case, therefore, it has to be assumed that there will be a finite value of input resistance or impedance.

When signal power is consumed in the input circuit the power gain is frequently expressed in decibels rather than as a simple ratio. It is generally so expressed in amplifying-device data sheets for those devices that require input power.

SMALL-SIGNAL AMPLIFICATION: It has been shown in the chapters discussing amplifying devices that the "constants" (frequently called parameters, especially in connection with semiconductors) such as voltage amplification factor, current amplification factor, and input and output resistances, are not strictly invariable but depend on the operating point. This is because the characteristic curves are not perfectly straight lines. If the curvature is appreciable, the parameters take on different values at each point on the curve because their values at that point depend on the slope. Because applying an ac signal voltage to an amplifier is equivalent to shifting the operating point back and forth at the signal frequency, the value of (for instance) the transconductance will vary with the instantaneous signal voltage.

If the voltage and current excursions are confined to an essentially straight section of the transfer characteristic the parameters are constant, for practical purposes. A signal that keeps the excursions within such limits can be considered to be *small*, and the behavior of the amplifier can be calculated by using the equivalent circuit of Fig. 23-2. For example, if the transconductance and signal input voltage are known the output current is

$$i_2 = g_m e_1,$$

and the output voltage is

$$e_o = i_2 R_L.$$

The voltage and power amplification can be found from the equations given previously.

OPERATING CONSTANTS: The value of transconductance in the above relationship is not the same as that given by the data sheets for the amplifying device used. Such data ordinarily are given for the case where the load resistance is zero, in which case the entire resistance of the output circuit is simply r_o, the internal output resistance of the amplifying device. Adding R_L in order to obtain useful output is equivalent to increasing r_o by the same amount. This decreases the transconductance to an *operating* value

$$g'_m = g_m \frac{r_o}{r_o + R_L}$$

where g_m is obtained from the data sheets or curves for the chosen operating conditions. The value g'_m should be used for calculating gain.

In the formula for g'_m the factor $r_o/(r_o+R_L)$ is a ratio which approaches 1 if R_L is small with respect to r_o. If R_L is *very much* smaller than r_o, the values of g'_m and g_m are close to the same; for example, if r_o is 10 times as large as R_L the ratio is 10/11, and basing the gain calculation on the published value of g_m rather than on g'_m will cause an error of about 9 percent. The variations in actual, as contrasted with published (the published values are averages), values of parameters are often much greater than this. The transconductance correction therefore can be ignored with amplifying devices, such as small vacuum-tube pentodes, which have values of r_o which are high compared with the values of R_L ordinarily used with them. This is also true of many types of transistors.

The voltage amplification factor, μ, of an amplifying device is usually published only for vacuum-tube triodes. If it is specified, the equation below gives the voltage gain:

$$\text{V.G.} = \frac{e_o}{e_1} = \mu \frac{R_L}{r_o + R_L}$$

μ and r_o being the amplification factor and plate resistance, respectively, at the selected operating point. With such tubes R_L is usually two or more times r_o for small-signal amplification. This improves the linearity of the amplifier, because R_L, being larger than r_o, has more control over the plate current than r_o does. The total resistance, $R_L + r_o$, thus becomes more linear because R_L, which is predominant, is inherently linear.

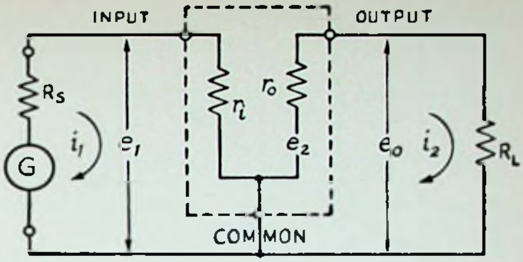


Fig. 23-2

USE OF CHARACTERISTIC CURVES: Characteristic curves relating current to voltage in the output circuit for various values of control-electrode voltage or current (the plate family or collector family) also can be used for calculating amplification. The advantage of using the curves is that they show the actual performance of the amplifying device, and it is not necessary to assume perfect linearity – an inherent assumption when using the equivalent circuit.

Fig. 23-3 is an example of their use. The supply voltage, E_D, for the MOSFET amplifier shown in the small circuit drawing is to be 30 V and the load resistance, R_L, is to be 1600 ohms. A load line with a slope equal to the load resistance, 1600 ohms, is drawn on the graph from 30 V on the supply-voltage axis to the drain current axis. The proper slope is easily found by dividing the supply voltage by the load resistance; i.e.,

$$\frac{30 \text{ V}}{1600 \text{ ohms}} = 18.75 \text{ mA},$$

so the load line intersects the I_D axis at 18.75 mA.

The point where the load line intersects a gate-voltage (V_{GS}) curve gives the actual drain current and the voltage, V_{DS}, between the drain and source at that particular value of gate voltage using a 1600-ohm load. If a series of values of I_D

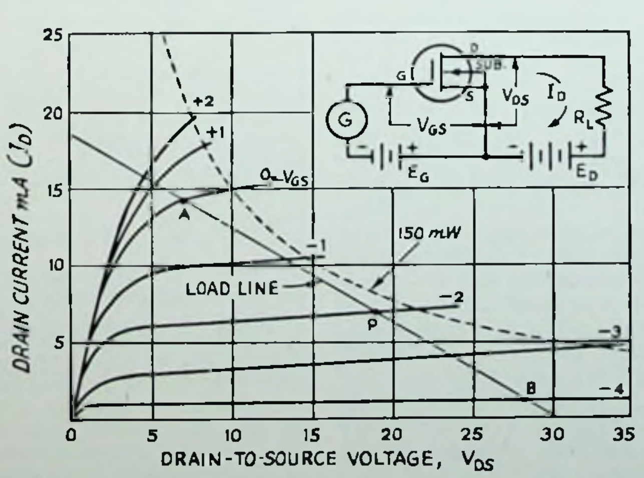


Fig. 23-3

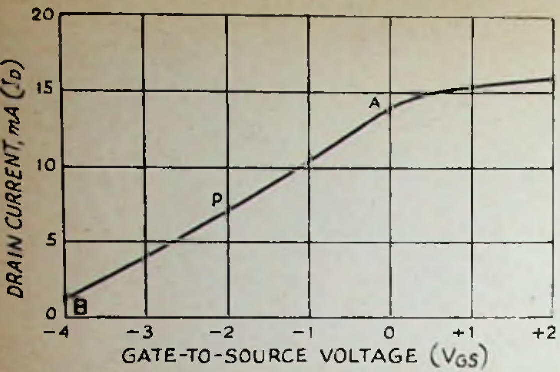


Fig. 23-4

and V_{GS} are used for plotting a transfer characteristic, as in Fig. 23-4, the departure of the amplification from true linearity is shown by the way the curve deviates from a straight line.

The transfer characteristic in Fig. 23-4 is reasonably straight between the points marked *A* and *B*, but the slope changes rapidly to the right of *A*. These points mark the useful limits of the curve if good linearity is to be obtained. The operating point, *P*, is placed in the center of the reasonably straight portion, and in this case *P* represents a gate-to-source bias, E_G, of −2 V. If the signal voltage varies between plus and minus 2 V – that is, if V_{GS} swings between 0 and −4 volts – the amplified output will be a reasonably good replica of the input signal waveform.

The three points can be transferred to Fig. 23-3 as shown. Then at the operating point, *P*, the drain current, I_D, will be 7 mA and V_{DS} will be 19 V. At the positive peak of the signal swing, *A*, where $V_{GS} = 0$, V_{DS} is 7 volts, and at the negative peak, *B*, where $V_{GS} = -4$ V, $V_{DS} = 28$ V. Thus the total (peak-to-peak) swing in drain voltage is 28 − 7 = 21 V, and since a peak-to-peak signal input swing of 4 V is required for producing it, the voltage gain is

$$\text{V.G.} = \frac{21}{4} = 5.25.$$

SELECTION OF LOAD RESISTANCE: In this illustration the load resistance, R_L, was specified initially, but other values could be used. (An important restriction is that the load line must lie below the dashed curve representing the safe power dissipation, 150 mW in this case). If the load resistance is higher, the slope of the load line will be less – that is, the intersection with the I_D axis will be at a lower current – and the voltage gain will be greater. However, it may also be found that the reasonably straight portion of the transfer characteristic is shortened, and if so the input signal amplitude must be reduced to confine the operation to the linear part of the curve; thus the ac output voltage available may be less even though the voltage gain is higher.

The selection of an appropriate load resistance also depends on the supply voltage available. With $E_D = 15$ V a considerably lower resistance (steeper slope to the load line) obviously would be called for if the amplifier is to handle the same peak-to-peak gate signal voltage satisfactorily.

DISTORTION: Although the transfer characteristic of the amplifier may be linear to a satisfactory degree, as in the example above, it is never perfectly so. As a result, the output waveform is never an *exact* reproduction of the input waveform. That is, there is a certain amount of distortion even under the best of conditions.

Any repetitive waveform that is not a pure sine wave can be resolved into a fundamental frequency and a series of harmonics, as described in Chapter 6. Thus when an amplifier distorts the waveform of the signal it causes new frequencies to be present in the output that were not present in the input. In many applications, such as in audio amplifiers, this is an undesirable condition.

Distortion is usually expressed, in amplifying-device published operating conditions, as a percentage of the total output. The distortion rating represents the sum of all the harmonics present in the output, unless a particular harmonic is specified. Distortion percentages are of the order of 5 to 10 percent for an amplifier operating as described above.

While distortion can be minimized by selection of the most suitable amplifying device and operating conditions, this generally restricts the amplification to rather small signals. Even so, there may still be more distortion than is desirable. An effective way of reducing distortion (it can never be completely eliminated) in such cases is the application of negative feedback in the amplifier, as described in Chapter 24.

SIGNAL SOURCE: In the above discussion the effect of the internal resistance R_S of the signal source was not considered. It can be neglected if the input resistance, r_i, is very high compared with R_S. In such cases e_l is simply equal to the voltage generated by G. However, if current does flow R_S becomes an important factor in the overall performance of the circuit. This is especially so with bipolar transistors. These exhibit low input impedance because current always flows in their input circuits when they are forward biased, as they must be for amplification to take place.

When signal current flows in the input circuit the output of the amplifier is usually greatest when the power taken from the signal source has its maximum possible value. This occurs when r_i is matched to R_S. Any of the applicable impedance-matching methods discussed in earlier chapters may be used for this purpose.

BIPOLAR TRANSISTORS: The simple equivalent circuit of Fig. 23-2 is most useful with amplifying devices in which the ac output current has a linear relationship with the signal input voltage. The vacuum tube and the FET, when handling small signals, are examples. The bipolar transistor, however, does not have this characteristic; the ac output current is a linear function of the signal input *current*. The relationship between the input voltage and output current in this type of transistor is not linear (see Fig. 22-5, Chapter 22, as an illustration); consequently it would be necessary, in using Fig. 23-2, to take account of the fact that the input resistance, r_i, varies with the input voltage in a rather complex way.

Although equivalent circuits can be devised to handle this, it is easier to calculate the approximate gain by using β or h_{fe}, the latter usually being given in the published characteristics. By making the reasonable assumption that the internal collector resistance will be much higher than the load resistance that would be optimum for the particular transistor, the ac output voltage is

$$e_o = h_{fe} i_1 R_L .$$

For example, if h_{fe} is 50, R_L is 2000 ohms, and the ac rms input current is 35 μA (35×10^{-6}A) the output voltage is

$$e_o = 50 \times 35 \times 10^{-6} \times 2000 = 3.5 \text{ V rms}.$$

In order to calculate the voltage gain it is necessary to know the voltage that the signal source must generate. Since the input resistance of the transistor is low (of the order of 500 ohms, at a typical operating point, in a transistor that would be used as a voltage amplifier) and varies with the base current, it is necessary that the signal source have high-enough resistance so that the *overall* input circuit will be substantially linear. If the source resistance is 10 or more times the transistor input resistance this condition will be met fairly well. Thus if the input resistance of the transistor itself averages 500 ohms and the source resistance is 5000 ohms, an input current of 35 μA will require a signal source that can generate a voltage of

$$35 \times 10^{-6} \times 5500 = 0.192 \text{ V}.$$

Then the voltage gain from the signal source to R_L is equal to e_o divided by the voltage generated in the source:

$$\frac{3.5}{0.192} = 18.3.$$

It is sometimes more useful to express the gain in terms of the ratio of the output power to the signal input power. The total power generated in the signal source is $i_1{}^2R$, where R is the total resistance of the input circuit – in this case, 5000 + 500 = 5500 ohms. Then the total power delivered by the signal source is

$$\begin{aligned} i_1{}^2R &= (35 \times 10^{-6})^2 \times 5500 \\ &= 6.76 \times 10^{-6} \text{ W}. \end{aligned}$$

The output power is $i_2{}^2R_L$, and since $i_2 = h_{fe}i_1$, the current is $50 \times 35 \times 10^{-6}$ A, or 1.75×10^{-3} A, so

$$\begin{aligned} i_2{}^2R_L &= (1.75 \times 10^{-3})^2 \times 2000 \\ &= 6.12 \times 10^{-3} \text{ W}. \end{aligned}$$

The power gain is then

$$\frac{6.12 \times 10^{-3}}{6.76 \times 10^{-3}} = 905,$$

which is 29.6 dB.

Note that for good linearity it may be necessary to sacrifice a major part of the power available from the signal source. When the internal resistance of the source does not meet the above requirement, resistance can be added between the source and the base of the transistor. If linearity is not a consideration, as may be the case in some types of rf amplifiers, more of the power available from the source can be utilized in supplying the power actually consumed in the base, and the overall power gain of the amplifier will be correspondingly greater.

Questions and Problems

23-1) What is the purpose of establishing an operating point for an amplifier?

23-2) What is meant by linearity in an amplifier of the type considered in this chapter?

23-3) What effect does nonlinearity have on the output waveform of an amplifier? To what does it correspond?

23-4) Draw a simple equivalent circuit of an amplifier showing the signal source and load resistance. State the functions of the elements of the equivalent circuit.

23-5) Under what conditions is the equivalent circuit useful?

23-6) What is a Class A amplifier?

23-7) What value of load resistance would be required for a voltage gain of 10 from the transistor whose characteristic curves are shown in Fig. 23-3? Use a supply voltage of 30 V. What would the drain current and voltage be at the operating point?

23-8) Using the operating conditions determined in Q. 23-7, how much output voltage could be obtained without undue waveform distortion?

23-9) How is the output voltage from an amplifier calculated?

23-10) Define power amplification, voltage amplification, and current amplification. Are they the same in a given amplifier?

23-11) If the rated transconductance of an amplifying device is 5000 micromhos and the internal output resistance is 20,000 ohms, determine the voltage gain when the load resistance is 5000 ohms. What would it be with a 25,000-ohm load? With a 100,000-ohm load?

23-12) Why does a large value of load resistance, compared with internal resistance, improve linearity?

23-13) What part of the voltage generated by an amplifier in its output circuit provides usable amplification of the signal?

23-14) Why is it desirable to have a large resistance in series with the signal source and the base of a bipolar transistor in the common-emitter circuit?

23-15) A bipolar transistor having a load resistance of 1500 ohms has a gain of 24 dB from signal source to load, in the common-emitter circuit. If h_{fe} is 35 and the signal source has a series resistance of 6800 ohms, determine the transistor input resistance. (Hint, assume some convenient value of output power, such as 1 watt, and work back to the input circuit.)

Input-output phase relationships; voltage between electrodes; input impedance; feedback; gain with negative feedback; distortion with negative feedback; other effects of negative feedback; applications; dc stabilization; bias circuit design; oscillators.

Feedback

ALTHOUGH THE characteristics of amplifying devices themselves do not change when the devices are in operation, the *apparent* characteristics can be greatly modified by the circuits in which the devices are used. The cause lies in the fact that the currents and voltages in the output and input circuits are affected by the way they react on each other through the amplifying device, and also because of the way the circuits themselves modify them.

INPUT-OUTPUT PHASE RELATIONSHIPS: Referring back to Fig. 23-3, Chapter 23, note that when the drain current has its maximum value at *A*, the voltage V_{DS} between the drain and source reaches its *lowest* value, and when I_D is minimum at *B*, V_{DS} reaches its *maximum* value. This is because the voltage drop in R_L is greatest when I_D is highest, thus leaving the least voltage on the drain, and vice versa. In terms of the input and output *ac* voltages measured with respect to the common source electrode, this relationship is shown in Fig. 24-1. The letters correspond to those used in Fig. 23-3. The output voltage is 180 deg out of phase with the input voltage.

The same phase relationship is true of the common-cathode vacuum-tube amplifier and the common-emitter bipolar transistor amplifier. The tube amplifier is shown at (A) in Fig. 24-2. When the ac grid voltage, e_g, changes in the positive direction with respect to the cathode, the plate current increases and the instantaneous voltage e_p between plate and cathode decreases. The npn bipolar transistor at (B) operates in the same way; a positive increase in base voltage, e_b, is accompanied by an increase in collector current and a negative swing (decrease) in collector voltage, e_c. With the pnp bipolar transistor at (C) a positive increase in base voltage increases the bias in the *reverse* direction, causing the collector current to decrease; thus the collector-to-emitter voltage, e_c, becomes larger – i.e., more negative. Although the polarities of the voltages are reversed, the collector voltage remains 180 deg out of phase.

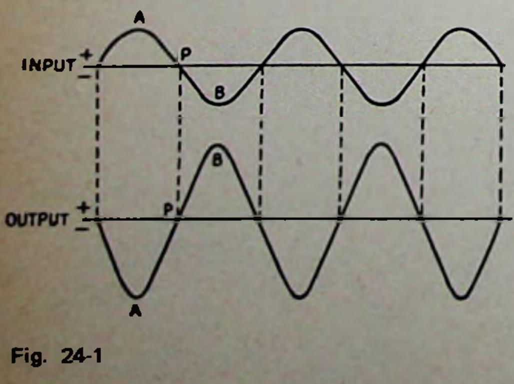


Fig. 24-1

VOLTAGE BETWEEN ELECTRODES: A consequence of this phase relationship is that the instantaneous voltage between the control electrode and the output electrode is the *sum* of the instantaneous voltages at those electrodes. This can be demonstrated as shown in Fig. 24-3, a skeleton version of Fig. 23-2, Chapter 23. If e_1, the ac control-electrode voltage, is instantaneously positive, as indicated, the output-electrode voltage is negative at that same instant. Going around the circuit, these two voltages are in series across the two electrodes, and the polarities in this series case are such that the voltages are in phase. They therefore add arithmetically between the electrodes: $e = e_1 + e_o$, using the notation for input and output voltage shown on the figure.

If the actual voltage amplification e_o/e_1 is some ratio A, then $e_o = Ae_1$, and the voltage between the output and control electrodes is

$$e = e_1 + Ae_1 = e_1 (1 + A).$$

The larger the actual voltage amplification, the greater the ac voltage between the control and output electrodes – grid and plate in the case of the vacuum tube, gate, and drain in the FET, and base and collector in the bipolar transistor.

INPUT IMPEDANCE: The voltage between the input and output electrodes is significant because the two electrodes are never completely isolated from each other in an amplifying device. There is always at least a capacitance between them, and in the bipolar transistor there is collector-to-base leakage resistance in addition. Because of this coupling the conditions existing in the output circuit exert a considerable influence on the input characteristics of the amplifier.

Fig. 24-4 is the equivalent circuit of a vacuum tube operated in the negative-grid region with the grid-plate capacitance, C_{gp}, added. The tube is used for this discussion because it represents possibly the simplest case, but the remarks also apply in a general way to transistors. Since the grid is always negative with respect to the cathode, r_i is practically infinite and can be ignored. (This is not true of the bipolar transistor, but its low input resistance does not change the *principle* under discussion.)

The voltage e_2 generated in the plate circuit is directly proportional to e_1, as explained in Chapter 23, and is in phase with it from the series standpoint discussed above. With C_{gp} present the current through r_o, the plate resistance of the tube, consists of two components. One, i_2, flows through R_L and the other, i_c, flows through C_{gp} and the signal source. Because of i_c the signal source sees a finite impedance rather than an infinite resistance.

Analysis of the circuit will show that if R_L is a pure resistance the input impedance of the amplifier is a pure capacitive reactance; that is, i_c leads e_1 by 90 deg. The equivalent input capacitance is

$$C = C_{gp}\,(1 + A),$$

where $A = e_o/e_1$ and is the actual voltage gain of the amplifier. This capacitance acts in parallel with the normal input capacitance of the tube, C_{gk}, (see Chapter 19) and the total input capacitance is therefore the sum of the two capacitances in parallel.

When the load is a complex impedance and thus contains reactance along with resistance, the input impedance of the amplifier is modified still further. Although the additional input capacitance caused by feedback remains essentially dependent on the voltage amplification, there is now a resistive component in parallel with it. (The values are often expressed in terms of conductance and susceptance, which can easily be converted into equivalent parallel resistance and reactance by taking the reciprocals, as described in earlier chapters.)

The equivalent parallel component of input resistance can go through wide variations, depending on the kind of reactance and the reactance/resistance ratio in the load impedance. In general, the input resistance component is *negative* if the load impedance is inductive, and is *positive* if the load reactance is capacitive. Negative resistance means that the phase of the current through C_{gp} is such that e_i is *increased* over its initial value, while positive resistance acts like ordinary resistance and places a heavier load on the signal source.

The effect of the load on input impedance is usually small enough to be neglected at frequencies in the lower part of the audio range. This is

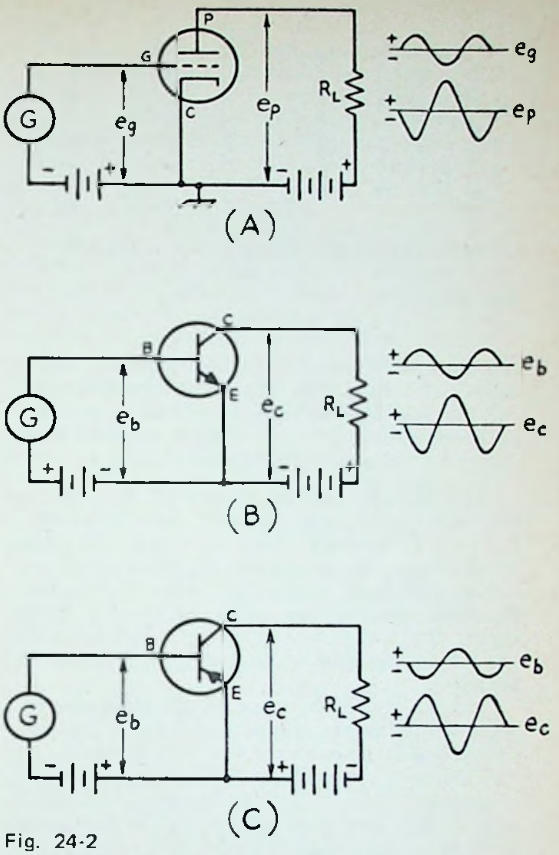


Fig. 24-2

because the reactance of the capacitance between the output and input electrodes is extremely high in this frequency range, with most amplifying devices. The current through the capacitance is therefore too small to have appreciable effect. However, because the reactance decreases with frequency this is no longer true above the audio range. An exception to this can be made with receiving-type vacuum-tube pentodes, which have very low grid-plate capacitance. (The normal input capacitance of these tubes, however, is not necessarily negligible.)

The discussion above has neglected various resistances that may be present, such as the collector-base leakage resistance and base-emitter resistance in bipolar transistors, and the channel

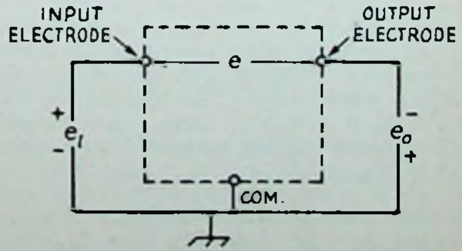


Fig. 24-3

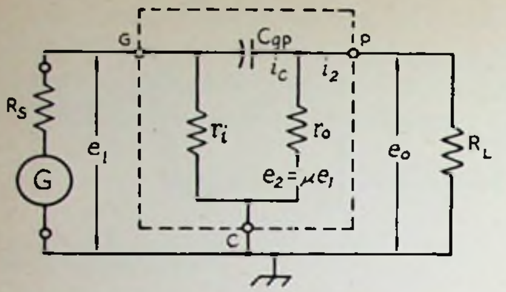


Fig. 24-4

resistance in the input circuit of field-effect transistors (see Chapter 21). These add further complications, and lead to equivalent circuits that are more complex than Fig. 24-4. More advanced texts should be consulted for information on these.

FEEDBACK: Current flow from the output electrode to the input electrode, described above, is a form of feedback. There are two general types of feedback. If the phase of the current (or voltage) fed back to the input circuit is such that the input signal voltage, e_1, is reinforced or made larger, the feedback is positive. If the fed-back current (or voltage) *opposes* e_1, the feedback is negative.

Positive feedback as described above occurs when amplifiers are operating with loads that are, or can readily become, reactive. This is the case in radio-frequency amplifiers using tuned input and output circuits. If the amount of energy fed back is large enough the circuit will generate its own signal and will not require an external signal source at all. A circuit operating in this way is an oscillator. In a circuit intended for amplifying external signals oscillation is an undesirable condition, and steps must be taken to prevent it. These methods will be discussed later. However, when the amplifier is intended for generating its own signal positive feedback provides the means by which continuous oscillation can be maintained.

The general case of an amplifier with feedback is shown in simplified form in Fig. 24-5. A signal e_1 applied directly to the input terminals of the amplifying device will cause an output voltage e_o to be developed across the load resistance. Some fraction, β, of the output signal is fed back to the input circuit. The feedback voltage therefore is βe_o. Since e_o is equal to the input voltage e_1 multiplied by the voltage gain A, the feedback voltage also is equal to $e_1\beta A$.

This voltage either adds to or subtracts from the voltage e from the signal source, so the voltage e_1 actually operating at the input terminals is not the same as e. However, the gain, A, of the amplifying device itself is not affected by the feedback, so to obtain the same output voltage with feedback as without, e_1 must stay constant. But if e_1 is to remain constant, e must change in such a way as to make the algebraic sum of βe_o and e also constant and equal to e_1. This means that the *circuit* gain, e_o/e, must differ from the *device* gain, e_o/e_1.

If the feedback is positive, e_o/e (the gain with feedback) is greater than e_o/e_1, the gain without feedback, because part of the input voltage, e_1, is now being supplied by feedback. If βe_o *subtracts* from e – i.e., the feedback is negative – e must be increased in order to preserve the same value of e_1; thus the gain with negative feedback is reduced as compared with the gain without feedback.

GAIN WITH NEGATIVE FEEDBACK: When the feedback voltage e_o – that is, $e_1\beta A$ – is 180 deg out of phase with e,

$$e_1 = e - e_1\beta A,$$

and therefore

$$e = e_1 + e_1\beta A = e_1(1 + \beta A).$$

As explained above, the gain with feedback is

$$\frac{e_o}{e} = \frac{e_o}{e_1(1 + \beta A)};$$

also, since $e_o/e_1 = A$,

$$\frac{e_o}{e} = \frac{A}{1+\beta A}$$

For example, if A = 50 and $\beta = 0.2$, the gain with feedback is

$$\frac{A}{1+\beta A} = \frac{50}{1 + (0.2 \times 50)} = \frac{50}{11}$$
$$= 4.55.$$

Whether or not the feedback voltage is exactly 180 deg out of phase with the input voltage depends on the circuit and the signal frequency. If the phase is not 180 deg the resultant input voltage, e_1, is not a simple sum as above but must be found by using the phasor method. Over a range of frequencies the phase of the feedback voltage may shift because of stray reactances. At the extremes of the useful frequency range of a given circuit the phase may shift sufficiently so that the feedback becomes positive instead of negative. This can lead to oscillation.

DISTORTION WITH NEGATIVE FEEDBACK: If other values of A are substituted in the above example while retaining the same feedback ratio ($\beta = 0.2$) it will be found that the gain with feedback changes relatively little. For instance, if A = 100 the gain is 4.77, while if A = 25 the gain is 4.17. Thus in this example a variation of 4 to 1 in gain without feedback is reduced to a variation of 1.15 to 1 with feedback.

Negative feedback, therefore, tends to make the circuit performance less dependent on the characteristics of the amplifying device. Since nonlinearity in a transfer characteristic such as is shown in Fig. 23-4 is caused primarily by nonuniform gain in the device itself, negative feedback has the effect of making the characteristic more linear. In turn, this reduces distortion. In a general way, the reduction in the amplitudes of distortion components in the output signal is the same as the reduction in amplification. In other words, if the amplification with feedback is 10 percent of its value without feedback, the amplitudes of the harmonics in the output signal will also be 10 percent of their amplitudes without feedback, for the same fundamental-frequency output voltage, e_o. Thus an original distortion of, say, 10 percent, will be reduced to 1 percent.

OTHER EFFECTS OF NEGATIVE FEEDBACK: At least three other effects occur when negative feedback is used. One is a reduction in the apparent internal output impedance of the amplifier. The value of A will change if R_L is changed, but the change in *overall* gain will be minimized by the feedback, as described in the preceding section. This means that e_o tends to remain constant when R_L is varied – exactly the same result that would be obtained by reducing the value of r_o in comparison to any value of R_L that reasonably might be used.

Second, the gain of the amplifier stays more constant over a wider range of input frequencies. The reason is the same as for the reduction in output impedance, because the principal factor affecting frequency response is the change in the effective load on the amplifier as the input frequency is varied. In amplifiers operating at audio frequencies R_L is shunted by stray reactances which become important at the high-frequency end of the frequency range, and there are usually coupling capacitors in the circuit which similarly are responsible for a loss in amplification at the low end. (With transformer coupling the transformer action is impaired at the low end because there will be some low frequency at which the primary inductance becomes insufficient – see Chapter 11.) Thus the amplification normally decreases at both the low- and high-frequency ends of the audio range. At radio frequencies the selectivity of the tuned circuits causes the amplification to decrease as the input frequency is moved away from the resonant frequency of the circuit. Negative feedback causes the amplifier to be more tolerant of these changes, thereby flattening and widening the frequency-response curve.

The third effect is that negative feedback increases the input impedance of the amplifier when the feedback voltage is in series with the signal source, as in Fig. 24-5. If the input circuit of the amplifying device has a finite value of resistance, r_i, the application of the signal voltage e_1 will cause a current i_1 to flow. When negative feedback is applied there is no change in i_1 if e_1 is unchanged. However, the resistance as seen by the signal source is changed because the source must supply a voltage

$$e = e_1(1 + \beta A)$$

to make the same current, i_1, flow. The input resistance with feedback is therefore

$$\frac{e_1(1 + \beta A)}{i_1},$$

which is equivalent to multiplying r_i by $(1 + \beta A)$.

It is possible to apply the feedback voltage in parallel with the signal source instead of in series with it. With parallel feedback the input impedance is decreased instead of increased. The reason is that the signal source sees not only the device input resistance, r_i, directly, in this case, but also sees the impedance of the feedback circuit in parallel with it. Thus the total input resistance is lowered. The reduction depends on the circuit used, and each case must be analyzed separately.

APPLICATIONS: This discussion of negative feedback has been confined to the simplest case, one stage of amplification with a minimum of circuit detail. Actual feedback circuits can take many forms, and feedback over more than one stage is not only feasible but quite common. When the feedback is applied over more than one stage, account must be taken of the fact that there can be a phase reversal in each one; this determines the choice of proper feedback-voltage phase in the initial circuit design. Details of this nature are beyond our scope, although they can be recognized in an actual circuit if the principles outlined above are kept in mind.

One simple circuit is worth mentioning since it is widely used. It is shown in simplified form in Fig. 24-6 for three basic types of amplifying devices (the circuit for junction-type FETs would be the same as for the MOSFET). Resistance R is common to the input and output circuits, so the ac current flowing through it is in series with the amplifying device and the load resistor, R_L. The ac voltage developed in it by the output current is applied between the common electrode (cathode, source, or emitter) and the control electrode. Instantaneous ac signal (*not* dc) polarities are indicated, and it can be seen that this is a series-type negative-feedback circuit because the voltage drop in R is out of phase with the voltage from the signal source, G, and the two voltages are of opposite polarity at the actual input terminals of the amplifying device. The battery voltages establish the operating point and do not take part in the ac operation since their internal resistance is assumed to be negligible.

In all three circuits the signal input current is low enough to be considered negligible in comparison with the output current, at frequencies which do not modify the input impedance because of coupling between the input and output electrodes of the type shown in Fig. 24-4. Under such conditions the amplified signal voltage developed

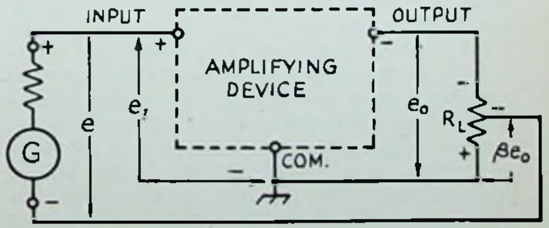


Fig. 24-5

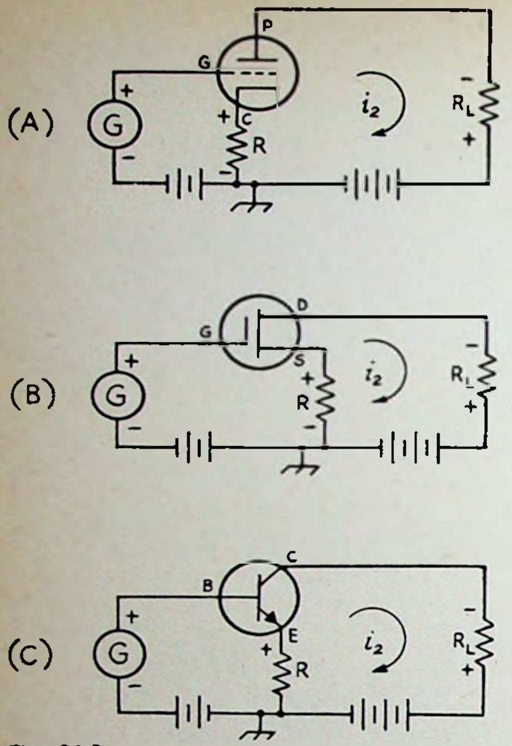


Fig. 24-6

in R is equal to Ri_2, and the voltage developed in R_L is R_Li_2. The total voltage between the output and common electrodes is the sum of these two voltages, and the two resistances form a voltage divider across the output circuit. Thus the fraction of output voltage fed back is

$$\beta = \frac{R}{R + R_L}.$$

The fraction β can be used as described above in calculating the gain with feedback and the reduction in internal output resistance, provided the actual amplification, A, without feedback is known. Similarly, in the bipolar circuit the increase in input resistance can be found if the input resistance without feedback is known.

D C STABILIZATION: The circuits shown in Fig. 24-6 have the additional advantage that they tend to stabilize the dc operating conditions. As explained in Chapter 22 (Fig. 22-9) the common resistor, there designated R_E, carries the direct current in the output circuit, and the polarity of the voltage drop in it is such as to increase the bias in the *reverse* direction when the current increases, thus reducing the current. Conversely, if for some reason the current decreases, the reverse bias is lowered, thus permitting the current to rise. The net result is that the current – plate current in a vacuum tube, drain current in the FET, and collector current in the bipolar transistor – tends to stabilize itself in spite of fluctuations that might arise because of temperature effects or small changes in the electrode voltages. This is a form of negative feedback for direct current.

BIAS CIRCUIT DESIGN: Because of the stabilizing effect, and also because operating bias for the control electrode can be obtained quite simply if the resistance is properly selected, the method shown in Fig. 24-6 is universally used for biasing small-signal (Class A) amplifiers. The dc voltage actually operating between the output and common electrodes is less than the supply voltage when this method is used, because the voltage which biases the control electrode ultimately comes from the supply. However, this is a minor consideration when the control-electrode voltage is small compared with the supply voltage, as is often the case.

The value of resistance needed for a specified bias voltage for a vacuum-tube or FET amplifier is

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

where E is the dc bias voltage required for a particular value of direct current, I, in the output circuit. The voltage and current can be selected from the device characteristic curves as described previously, or may be taken from published operating conditions.

In the case of the bipolar transistor, which must be forward-biased for operation of the type described in this chapter, a bias voltage divider must be added to the circuit as shown in Fig. 22-9, Chapter 22. Using the notation of Fig. 22-9, R_E is generally selected to develop a few volts bias at the selected value of collector current; the voltage is not especially critical, but the larger it is the better the stabilization. (However, if it is an appreciable percentage of the supply voltage, V_{CC}, a suitable compromise must be made between stabilization and loss of voltage.) Once the drop in R_E is selected, the voltage divider R_1R_2 should be designed so that the voltage drop in R_1 is about 0.6 volt larger than the drop in R_E, for a silicon transistor, or about 0.15 volt for germanium. The base-emitter current usually can be neglected in these calculations because it is small compared with the collector current; also, since the circuit is self-stabilizing the values are not especially critical. However, experimental adjustment of the value of R_E will allow setting the collector current as precisely as may be necessary, once R_1R_2 has been designed.

OSCILLATORS: As stated earlier, an amplifier can be made to generate its own self-sustaining signal input if sufficient positive feedback is used in the circuit. The requirements for oscillation are that the feedback voltage be of the proper amplitude and phase, and that the power amplification be great enough to supply all losses that may occur in the input circuit.

It is not necessary for an external signal to be present to cause oscillations to begin. Any random signal, such as the noise voltage always present in electrical circuits, can be amplified to obtain a feedback voltage which then reinforces the original voltage at the input. This in turn is reamplified to produce a larger signal which again reinforces the input, and so on. The process continues until the output of the amplifier is unable to increase any further. This equilibrium point may be determined either by saturation in the amplifier, in which case no further increase in output and input is possible, or by using some external means for limiting the signal build-up to a desired level. The oscillation frequency is always that at which the amplification is greatest.

Many types of oscillator circuits are possible. They may be classified in different ways, depending on the circuit type, the type of waveform generated, and so on. A useful method is to divide oscillators into types that generate sine waves and types that do not. Sine-wave oscillators for radio frequencies are especially useful in communication. They are described briefly in Chapter 26.

Questions and Problems

All questions refer to amplifier circuits in which the cathode, source, or emitter is the common electrode.

24-1) If an amplifier has a purely resistive load, what is the phase relationship, with respect to the common electrode, of the ac input and output voltages?

24-2) What is the phase relationship between ac input voltage and ac output current?

24-3) What is positive feedback?

24-4) What is the effect of feedback on the input capacitance of an amplifier?

24-5) On what does the frequency generated by an oscillator depend?

24-6) What is negative feedback? What are its advantages and disadvantages?

24-7) What resistance is required in the source lead if the operating point indicated in Fig. 23-3, Chapter 23, is to be established by the source-resistor bias method?

24-8) If the resistor in Q. 24-7 is used to bias the transistor, what value of drain supply voltage is required?

24-9) How does a bias resistor in the common-electrode lead help stabilize the direct current in the output circuit?

24-10) What is meant by negative resistance?

24-11) An amplifier has a voltage gain, A, of 30 without feedback. If negative feedback is used with $\beta = 0.1$, what is the gain with feedback?

24-12) What effect does negative feedback have on the frequency response of an amplifier?

24-13) A vacuum-tube amplifier has a gain of 12 without feedback when the plate load resistance is 50,000 ohms. If 1000 ohms of this is used as an unbypassed cathode resistance, what gain is actually realized in the circuit? Would this change if the cathode resistor is bypassed by a capacitive reactance of 50 ohms?

24-14) What effect do circuit constants, intended or stray, have on the phase of feedback voltage?

24-15) The interelectrode capacitances in a vacuum tube are C_{gk} = 2 pF, C_{gp} = 1.5 pF. The gain is 20 and the load is a pure resistance. What is the input capacitance?

24-16) What effect does a reactive load have on the input impedance of an amplifier?

24-17) How is the ac voltage between the control and output electrodes of an amplifier affected by the voltage gain?

24-18) In a simple amplifier circuit with no external provision for feedback, can feedback nevertheless exist? If so, what type is it and what is the cause?

Follower circuits; equivalent circuit; effect of feedback; output resistance; grounded-grid, -gate, or -base circuits; feedthrough; input resistance; resistance of signal source; multistage amplifiers; transformer coupling; capacitive coupling; undesired feedback.

Amplifier Circuits

THE DISCUSSION of amplification in the preceding chapters has been confined to a group of circuits comprising the common-cathode circuit for the vacuum tube, the common-source circuit for the field-effect transistor, and the common-emitter circuit for the bipolar transistor. However, any of the three basic electrodes present in electronic amplifying devices can be used as the meeting point between the input and output sections of an amplifier. This meeting or common point is usually, but not always, held at ground potential.

The terms "common" and "grounded" can be used interchangeably if the common point in the circuit is actually at ground potential.

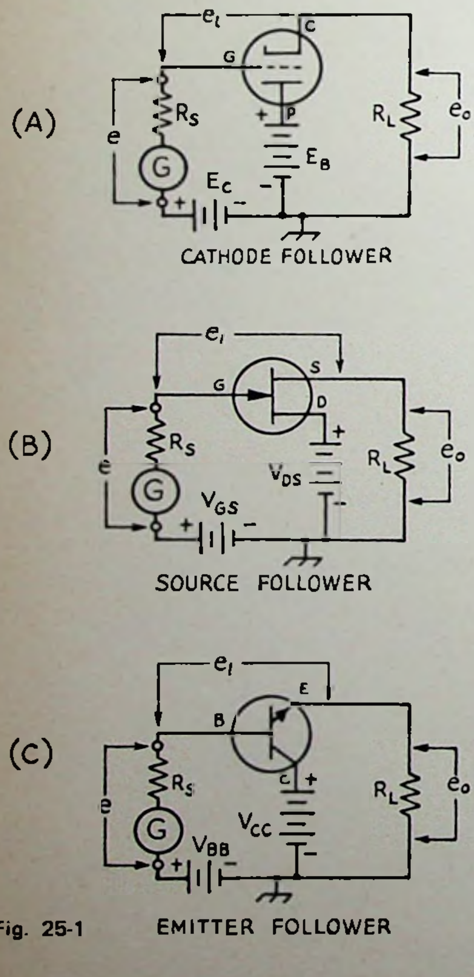


Fig. 25-1

FOLLOWER CIRCUITS: One such group of circuits is shown in Fig. 25-1. In this case the output electrode – plate, drain, or collector – is the meeting point for the input and output sections and is at ground potential for signal voltages in both sections. The polarities indicated are those for the dc operating voltages. The batteries (or other dc power sources) are assumed to have zero internal impedance, and therefore can be eliminated so far as signal currents are concerned. The polarities of the supplies for the bipolar transistor are for npn types; polarities would be reversed for pnp types. Also, the junction-type FET can be substituted for the MOSFET shown in the FET circuit and the discussion to follow will still apply if the gate is not driven into the conduction region by the signal input voltage.

Taking the vacuum-tube triode (upper circuit) as representative of the three circuits, the direct current in the output circuit flows through R_L, just as in the circuits discussed earlier. Thus the dc voltage drop in R_L is proportional to the plate current. This drop is applied as negative bias to the grid, since the cathode is positive with respect to ground.

With the values of R_L that would be used for normal operating conditions, so much bias would be developed that the plate current would be very small, and would differ considerably from the value required for establishing the selected operating point. It is necessary, therefore, to supply additional bias, E_C, of the *opposite* polarity so the actual bias between grid and cathode, the difference between E_C and the drop in R_L, will have the proper value. In practical circuits this can be accomplished in various ways without using a separate battery; for instance, a high value of resistance can be connected from the grid of a vacuum tube or the gate of an FET to a tap on R_L to provide the proper voltage between grid and cathode or gate and source. With bipolar transistors the biasing method shown in Fig. 22-9, Chapter 22, can be adapted to the same purpose.

EQUIVALENT CIRCUIT: By rearrangement of Fig. 23-2, Chapter 23, an equivalent circuit for this type of operation can be drawn as shown in Fig. 25-2. The letters G, C, and P identify the grid, cathode, and plate, respectively, of the tube. Corresponding designations can be substituted for the semiconductor devices. The other symbols are the same as those used in the preceding chapter.

The polarities in Fig. 25-2 are instantaneous ac signal polarities and are normal for the amplifying device; i.e., when the ac signal at the grid swings in the positive direction the instantaneous voltage at the plate swings in the negative direction. At this time the polarity at the cathode is positive with respect to ground, and is above ground by the ac voltage drop in R_L. Going around the grid-cathode-ground path in series, e_1 and e_o are in phase with reference to ground, and the voltage e from the signal source is the sum of the two. Thus all of the output voltage developed in R_L is fed back to the input circuit. The feedback is negative with β equal to 1.

This means that the voltage gain of the circuit is always less than 1. As the actual amplification, A, in the tube itself is unchanged by the circuit, the voltage gain is

$$\text{V.G.} = \frac{A}{1 + \beta A}$$

and since $\beta = 1$,

$$\text{V.G.} = \frac{A}{1 + A}$$

The larger the value of A, the more nearly the voltage gain of the circuit approaches 1.

EFFECT OF FEEDBACK: The effect on the input resistance of this circuit is as described for negative feedback in Chapter 24, and since $\beta = 1$ the input resistance of the amplifier is equal to its grounded-cathode value, r_i, multiplied by $(1 + \beta A)$; that is, by $(1 + A)$.

The large amount of feedback also reduces the effect of the grid-cathode capacitance, C_{gk}, of the tube (and the corresponding capacitances in the other amplifying devices). C_{gk} is not shown in Fig. 25-2, but is between the grid and cathode terminals. The current which flows through it is caused by e_1. However, for a given value of e_1 the source of signal must supply a voltage e, and since $e = e_1 + e_o = e_1(1 + A)$, a voltage $e_1(1 + A)$ is required for causing the same capacitive current. This is equivalent to reducing the effective input capacitance of the amplifier by the factor $1/(1 + A)$, or approximately $1/A$ times its value in the grounded-cathode circuit.

However, the grid-to-plate capacitance in this circuit is directly across the signal source. Thus the total input capacitance is the sum of the two capacitances. If A is at all large, the total is approximately the grid-plate capacitance alone.

OUTPUT RESISTANCE: The internal output resistance, r_o, of the amplifying device is reduced in approximately the ratio $1/g_m$, g_m being the transconductance. For example, if the transconductance is 4000 μmho (0.004mho) the output resistance is

$$r_o = \frac{1}{0.004} = 250 \text{ ohms.}$$

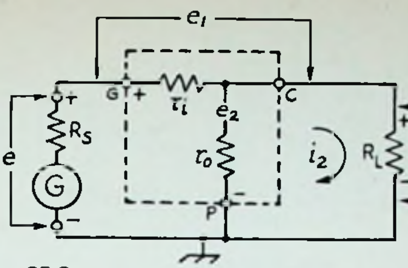


Fig. 25-2

The low output resistance combined with the high input resistance and reduction in input capacitance makes this circuit very useful as a step-down impedance transformer. It is capable of operating over a wide band of frequencies in rf applications.

Because of the position of R_L in these circuits the amplified ac voltage, e_o, is in phase with the signal input voltage, e_1, with respect to ground. That is, a positive swing in e_1 increases the current in R_L, and thus makes the voltage at the cathode, source, or emitter more positive. The circuits are called followers for this reason.

Because of the very large amount of feedback, much larger input signals can be handled without distortion than is possible in the circuits discussed earlier. The grid cannot be driven positive with respect to the cathode so long as the operation remains in the unsaturated part of the plate current vs. grid voltage curve.

GROUNDED-GRID, -GATE, OR -BASE CIRCUITS: A third group of circuits is shown in Fig. 25-3. The common element is the control electrode – grid, gate, or base. The polarities are dc potentials with respect to ground. The batteries (or source of voltage supply) again are assumed to have zero internal impedance, so no ac voltage is developed across them when a signal is being amplified. The load resistance, R_L, is connected between the output electrode and ac ground.

The vacuum-tube triode can again be used to describe the operation of this circuit, with the understanding that the remarks also apply to the transistor circuits when the corresponding electrodes are substituted for the grid, plate, and cathode.

Fig. 25-4 is the equivalent circuit of Fig. 23-2, Chapter 23, rearranged so that the grid of the triode is the common or ground point. In the output circuit an amplified ac voltage e_2 acts to cause the output current to flow through the internal resistance, r_o, as it did in Fig. 23-2 when the cathode was the common point. However, in this case the cathode is above ground by the input voltage, e_1. Thus the input and output voltages are in series, going around the circuit through R_L and the signal source.

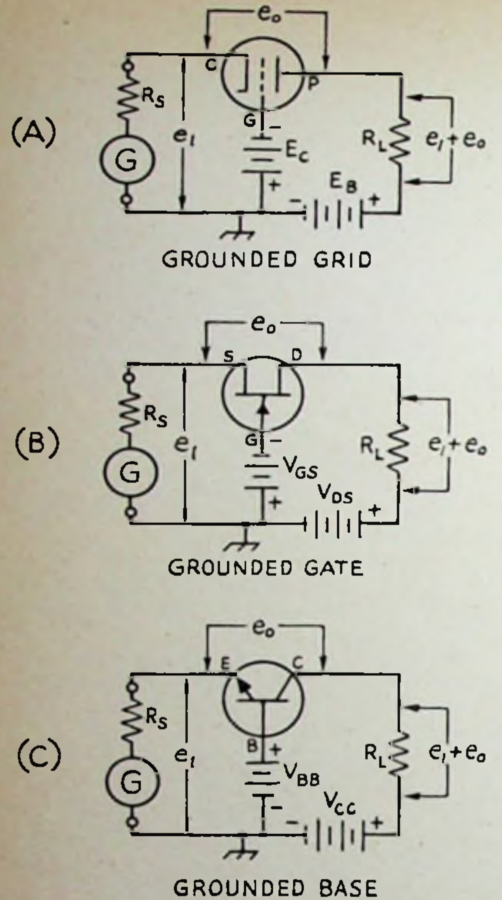


Fig. 25-3

FEEDTHROUGH: If the ac signal-input voltage e_1 swings positive with respect to ground at some instant, as indicated by the plus sign on the drawing, the grid must become more negative with respect to the cathode at that instant. This makes the ac voltage at the plate positive with respect to the cathode at the same instant, so e_1 and e_o are in phase. Thus the voltage across R_L is the *sum* of e_1 and e_o, and the input and output voltages both contribute to the power in the load. If r_i is very large compared with the signal-source resistance, R_S, as is usually the case with negative-grid tubes and FETs, all of the output current, i_2, can be assumed to flow through the signal source on its way from the plate to the cathode. Then the power supplied by G to the load is $e_1 i_2$, and the total power in the load is

$$P_o = (e_1 + e_o)\, i_2 .$$

The fraction of the load power that is supplied by the signal source is

$$\frac{e_1 i_2}{(e_1 + e_o)\, i_2} = \frac{e_1}{e_1 + e_o} ,$$

and since $e_o = Ae_1$, the ratio is

$$\frac{e_1}{e_1 + Ae_1} = \frac{1}{1 + A}$$

For example, if the voltage amplification, A (which equals e_o/e_1) is 12, the fraction of the total load power supplied by the signal source is

$$\frac{1}{1 + 12} = \frac{1}{13} = 0.077\text{, or } 7.7\%.$$

This power, called fed-through power, is supplied by the signal source whether or not there is any power loss in the input circuit of the amplifying device itself. If the input resistance, r_i, has a finite value – as it does in bipolar transistors, in particular – the power lost in r_i must be added to the fed-through power in order to determine the total power that the signal source must supply. Note that the fed-through power is not dissipated in the amplifier but is a useful contribution to the total output power in the load, R_L.

INPUT RESISTANCE: Because the signal source must supply appreciable power in this circuit, the input impedance of the amplifier is relatively low. The input voltage, e_1, required for a given output voltage is not affected because e_1 is applied directly between the cathode and grid. However, e_1 is now helping to supply the load current, i_2, in addition to the amplifier input current, i_1 (if any). The input resistance is therefore equal to

$$R_{in} = \frac{e_1}{i_1 + i_2} .$$

If i_1 is small in comparison with i_2 the approximate input resistance can be determined from R_L and the share of the output power supplied by the signal source. Since e_1 and e_o are in series and the contribution of each to the total power in R_L is in proportion to its fraction of the total voltage, the input resistance is

$$R_{in} = R_L \frac{1}{1 + A} .$$

If the load resistance for the amplifier used as an example above is, say, 2500 ohms, the input resistance is

$$R_{in} = 2500 \times 0.077 = 192 \text{ ohms}.$$

When r_i is too small to be neglected, it must be combined in parallel with the input resistance determined above, thus causing a further reduction in the resistance seen by the signal source.

RESISTANCE OF SIGNAL SOURCE: No mention has been made of the signal-source resistance, R_S, in this discussion nor in Chapter 24, but its significance has been pointed out in Chapter 23. In the negative-feedback circuits of Fig. 25-1, as well as those in Chapter 24, R_S becomes less of a factor because of the increased input resistance of the amplifier circuit. However, it cannot be neglected entirely, especially with bipolar transistors. Neither can it be neglected with other types of amplifying devices when operating at radio frequencies, where the input impedance is lowered for the reasons discussed in earlier chapters. When current flows in the input circuit the voltage e_1 actually operating at the input terminals of the device will always be less than the voltage generated in the signal source because of the voltage drop in R_S. If the maximum possible signal power is to flow the amplifier's input resistance must be matched to R_S.

In the circuits shown in Fig. 25-3 the output current, i_2, flows through R_S and the voltage drop in R_S cannot be neglected. Because the input impedance of the amplifier is low, R_S is also required to be low; otherwise most of the power available from the signal source will be dissipated in the source itself. The voltage drop in R_S, being caused by the *output* current, is a species of negative feedback since its polarity is such as to reduce the amplitude of e_1, for a given voltage

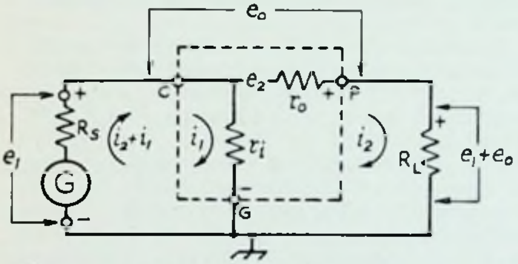


Fig. 25-4

generated in the signal source. The feedback fraction is

$$\beta = \frac{R_S}{R_S + R_L}$$

(this assumes that i_1 is small compared with i_2, which is usually the case).

In the discussion of negative feedback in Chapter 24 the total signal voltage, e, used in the formulas was also the *terminal* voltage of the signal source. The effect of current flow through R_S therefore did not need to be considered, and the assumption that the generated and terminal voltages were the same was justifiable because even in the case of the bipolar transistor the input current is very small, and the voltage drop in R_S therefore negligible. In the equivalent circuit of Fig. 25-4 the terminal voltage of the signal source, designated e_1, includes the effect on the source voltage of the drop in R_S, so the drop did not need to be included in calculating the input impedance of the amplifier. The formula for β in this circuit applies only when the voltage generated *internally* in the signal source is used in the calculations. Under these circumstances the voltage drop in R_S has the same effect as any other negative-feedback voltage in decreasing the sensitivity of the amplifier to changes in load.

MULTISTAGE AMPLIFIERS: More often than not, a single amplifier stage is incapable of amplifying the signal input voltage or power to the

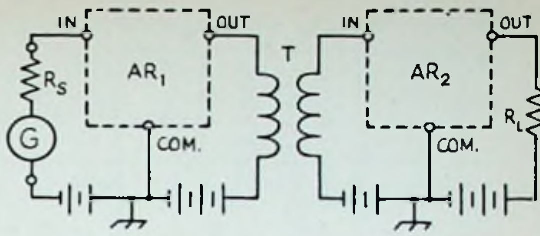


Fig. 25-5

level needed. Two or more stages will be used in such cases. When this is done the output of the first stage supplies the input for the next, and so on.

The output of a stage must be fed to the input of the following one without affecting the dc voltages in either; that is, the two stages must have dc isolation. (There are exceptions to this – direct-coupled stages – but they are too specialized for consideration here.) The dc isolation can be obtained either by using transformer coupling between stages, as in Fig. 25-5, or by separating the stages by means of capacitive coupling as in Fig. 25-6. The function of R_2 in Fig. 25-6 is to provide a dc return path for current flowing between the input and common electrodes of amplifier AR_2; without it, the circuit is "open" for direct current. It is also needed when there is no actual direct-current flow in the input circuit, as with negative-grid vacuum tubes and MOSFETs, because without it the average bias on the control electrode of AR_2 would depend on the average charge accumulated on coupling capacitor C over a period of time. As this charge has no way of being dissipated without R_2, the operating point of AR_2 would drift, an undesirable condition.

The amplifiers in these circuits may be any of the devices previously discussed. They are shown simply in block form in the drawings, with only the terminals indicated. No specific electrodes are assigned to these terminals, because any of the circuits described in this and the preceding two chapters may be used.

TRANSFORMER COUPLING: Transformer coupling offers a ready means for obtaining maximum power transfer when the output impedance of the first stage differs appreciably from the input impedance of the following one. It is particularly useful when a high output impedance must be

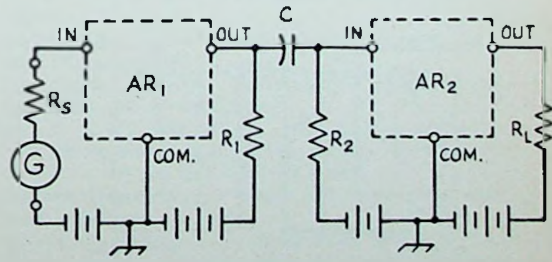


Fig. 25-6

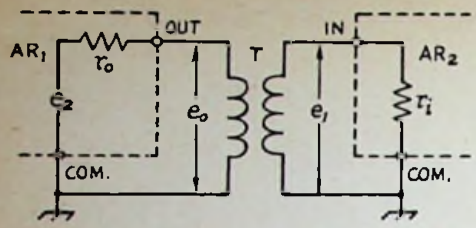


Fig. 25-7

coupled to a low input impedance. The properties of transformers in this and other respects have been discussed in Chapters 11, 13, and 14.

A simplified equivalent circuit, omitting the input details of amplifier AR_1 and the output details of AR_2, is given in Fig. 25-7. The input resistance, r_i, of AR_2 is the load for the secondary of transformer T. The transformer impedance ratio is chosen to match this resistance to the optimum load resistance, R_L, for AR_1. R_l is not shown physically in this drawing since it represents r_i as transformed by the impedance ratio of T. The values of both R_L and r_i can be determined by the methods outlined in this and the preceding chapters, for the particular circuit used.

At audio frequencies the transformer design can follow the principles set out in Chapter 11, since at such frequencies a coupling coefficient very close to 1 can be achieved and the impedance ratio therefore would be determined by the turns ratio. At radio frequencies this approach may be impracticable, in which case tuned transformers of the types described in Chapters 13 and 14 would be called for. Radio-frequency amplifiers are discussed further in Chapter 26.

CAPACITIVE COUPLING: The simplified equivalent circuit for capacitive coupling is shown in Fig. 25-8. The amplified ac signal from AR_1 is transferred to the input circuit of AR_2 through coupling capacitor C, which acts as an insulator for dc. Thus the supply potential on the output electrode of AR_1 does not appear at the input electrode of AR_2, and vice versa.

The reactance of C should be low enough at the operating frequency (or the lowest frequency of a band) so that the signal-voltage drop through it is small compared with the signal voltage that appears across R_2 and r_i in parallel. When this condition is met the actual output load on AR_1 is the resistance represented by R_1, R_2, and r_i in parallel.

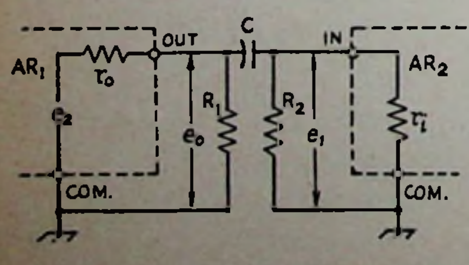


Fig. 25-8

That is, this combination of resistances is equivalent to R_L in the circuits discussed earlier. It is the value that must be used in calculating the voltage amplification.

If the reactance of C is not more than about 10 percent of the resistance of R_2 and r_i in parallel, the ac voltage across these two resistances is over 99 percent of the ac voltage across R_1. At high frequencies the capacitance required is relatively small, even when R_2 and r_i have a parallel resistance of less than 1000 ohms. The same input resistance, however, would need a capacitance of about 30 μF at 50 Hz (reactance = 100 ohms); in nearly all cases this would require an electrolytic capacitor. When such a capacitor is used its polarization must be observed; that is, the positive-marked side must be connected to the point of higher positive potential. This ordinarily will be at the output electrode of AR_1. The polarity must be reversed if the higher of the two potentials is negative, as when the amplifiers are of the pnp bipolar type.

The value of coupling capacitance also can be chosen on the basis of some permissible decrease in the voltage amplification. For example, if the gain is permitted to drop off 3 dB (voltage ratio 0.707) at the low end of the frequency range to be covered by a wide-band audio amplifier, the reactance of C at that frequency will be equal to the resistance of R_2 and r_i in parallel. If this resistance is, say, 1000 ohms the reactance of C also will be 1000 ohms. This calls for a capacitance of approximately 2 μF at 50 Hz. The required capacitance decreases, of course, as its "load" resistance (R_2 and r_i in parallel) is made larger.

In the actual circuit R_2 may not be a single resistance (as it is shown in Fig. 25-8) but may be the equivalent resistance of a bias network. Such networks are customarily used in bipolar-transistor common-emitter circuits. In the network shown in Fig. 22-9, Chapter 22, resistors R_1 and R_2 are effectively in parallel across the signal source because R_2 returns to ground through V_{CC}, which is assumed to have zero impedance. In such a case the parallel resistance of R_1 and R_2, Fig. 22-9, must be substituted for R_2, Fig. 25-8. Since the bias network places additional load on the preceding amplifier the values of resistance used in it should be as high as other design considerations, such as the regulation of the bias current, will permit.

UNDESIRED FEEDBACK: In practical amplifier circuits it is necessary to prevent feedback from one stage to an earlier one, unless the feedback is a deliberate part of the circuit design. In low-frequency amplifiers the principal cause of such feedback is ac coupling through power and bias supplies that are common to two or more amplifier stages. Such coupling has not been considered in Figs. 25-5 and 25-6 where separate supplies, all assumed to have zero internal impedance, are shown for each stage.

Actually, no source of dc power has zero internal impedance. Although the impedance may be small at the frequency or frequencies being amplified, it is nevertheless finite. And since the ac output currents of the amplifying stages flow through it, an ac voltage is developed in it. Fig. 25-9(A) illustrates the situation. The output ac currents of the three amplifiers all flow through the supply impedance, Z, since it is common to all three output circuits. If the supply also is the source of bias for bipolar circuits Z is also common to the *input* circuits of the amplifiers, a complication which need not be included here because the overall result is still unwanted feedback.

The output currents are larger in consecutive stages because of the amplification, and i_3 is ordinarily so large that it dominates the others. Thus a voltage may be generated in Z which is comparable to the output voltage of AR_1, for example. This voltage is transferred by the coupling circuit to the input electrode of AR_2. Its phase relationship to the output voltage of AR_1 depends on the exact relationship between Z and the various output impedances. If it has an in-phase component the output from both AR_2 and AR_3 is increased, and self-oscillation generally results. If it is out of phase the overall amplification is decreased below the value to be expected from ordinary design considerations.

Unwanted feedback of this type can be reduced to negligible proportions by using decoupling circuits, as shown in Fig. 25-9(B). If C_1 has low reactance, at some frequency well below the lowest operating frequency, compared with the resistance of R_1, very little of the ac voltage developed in Z can appear across C_1, and thus feedback is minimized. Similarly, in the second stage AR_2 is decoupled by C_2 and R_2. The resistance values usually are chosen so that the dc voltage drops in them will not be large enough to rob the amplifying device of a substantial amount of supply voltage. The use of decoupling resistance is frequently impracticable in the last stage because if this stage is required to deliver appreciable power it will have a relatively high direct current in its output circuit; in such case, the reactance of C_3 is simply made as low as practical components available will permit.

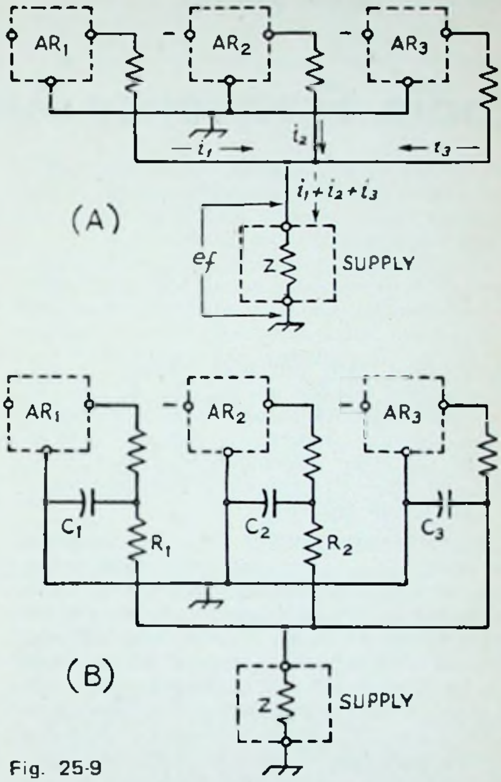


Fig. 25-9

Questions and Problems

25-1) How can amplifier circuits be classified?

25-2) What order of voltage gain is obtainable from a source follower? From an emitter follower?

25-3) Must all stages in a multistage amplifier use the same type of circuit?

25-4) If the voltage gain, A, is 7 in a common-gate circuit, what is the input resistance if R_L = 1500 ohms?

25-5) Add interelectrode capacitances to the equivalent circuit of Fig. 25-4. Can feedback occur through the grid-plate capacitance in this case?

25-6) If the voltage gain, A, in a common-grid circuit is 25, what part of the output power is supplied by the signal source? What part if A = 8?

25-7) If the amplifier of Q. 25-6 is changed to the common-cathode type, what is the effect on the voltage gain? What if it is changed into a cathode follower?

25-8) What are the advantages and disadvantages of transformer coupling? When must tuned transformers be used?

25-9) A common power supply having an internal resistance of 10 ohms is used for a multistage amplifier. If the ac output current from the last stage is 25 mA, what is the voltage fed back to the preceding stages if there is no decoupling? If a decoupling resistance of 1000 ohms and a decoupling capacitance having a reactance of 150 ohms are connected in the output circuit of the first stage, how much feedback voltage is applied to this stage?

25-10) An audio amplifier uses the interstage coupling circuit shown in Fig. 25-6. If R_2 = 0.5 megohm, C is 0.01 μF, and the frequency is 100 Hz, what proportion of the signal voltage across R_1 appears across R_2?

25-11) Why is decoupling usually necessary in a multistage amplifier?

25-12) Is a follower circuit using a resistive load free from positive feedback?

25-13) If the voltage gain, A, is 20 in a common-cathode circuit, what is the voltage gain if the circuit is changed to a follower type?

25-14) What order of power output is obtainable from a follower circuit?

25-15) How does a common-drain, -collector, or -plate circuit compare with a common-source, -emitter, or -cathode circuit with respect to overload from large input signals?

Equivalent circuit; interstage coupling; feedback; stabilization; reduction of C_f; neutralization; bridge circuits; bridge neutralization; other neutralizing circuits; phase relationships in practice; other amplifying circuits; frequency limitations; radio-frequency oscillators; frequency stability.

Radio-Frequency Amplification

ALTHOUGH THE principles outlined in the preceding chapters are valid for all frequencies, their application becomes more complicated as the operating frequency is increased. This is because at radio frequencies stray reactances, usually of little or no concern at low frequencies, no longer can be neglected.

EQUIVALENT CIRCUIT: A simplified equivalent circuit of an amplifying device operating at frequencies up to perhaps 30 MHz can be drawn as shown in Fig. 26-1. It is basically the same as the one described in Chapter 23 – i.e., the common electrode is the cathode, source, or emitter. However, an input capacitance, C_i, output capacitance, C_o, and feedback capacitance, C_f, have been added. C_i and C_o shunt the input and output resistances, respectively, and although the capacitances are small – of the order of a few pF in most cases – their reactances are not negligible at radio frequncies. A capacitance of 2 pF at 5 MHz, for example, has a reactance of 15,000 ohms, which is of the same order as the input and load resistances associated with a radio-frequency voltage amplifier using vacuum tubes of FETs.

Because input and output capacitances readily can be absorbed into the capacitances associated with tuned circuits, the input and output circuits of rf amplifiers used for communication always are tuned. (An exception can be made if the input and load impedances are so low that stray capacitances have little effect.) A second, and equally important, reason for using tuned circuits is that they provide selectivity that will permit amplifying a desired signal and suppressing undesired ones – an essential for good communication.

As shown in Fig. 26-2, C_1 is in parallel with C_i, and C_2 is in parallel with C_o. The input and output capacitances, then, become part of – that is, add

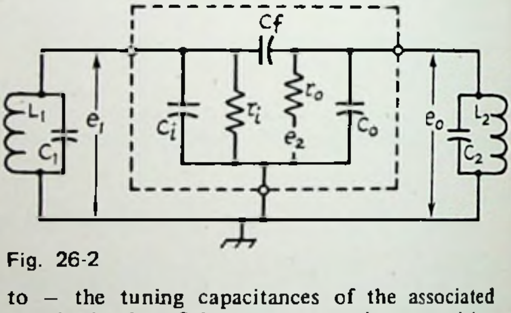


Fig. 26-2

to – the tuning capacitances of the associated tuned circuits. Other stray capacitances arising from, for example, the proximity of wires to a metallic chassis and shields, can be similarly absorbed in the tuning capacitances. Even so, it is important to minimize such stray capacitances as much as possible, because at the higher frequencies (particularly in the vhf range) a reasonable ratio of inductance to capacitance cannot be maintained if the irreducible capacitance in the circuit becomes large. A low L/C ratio makes it difficult to develop high-enough impedance in the tuned circuits to achieve the amplification the device is capable of giving.

INTERSTAGE COUPLING: A typical coupling circuit between two amplifiers is shown in Fig. 26-3. (Only the output equivalent of AR_1 and the input equivalent of AR_2 are shown.) This type of circuit has been discussed in detail in Chapter 14.

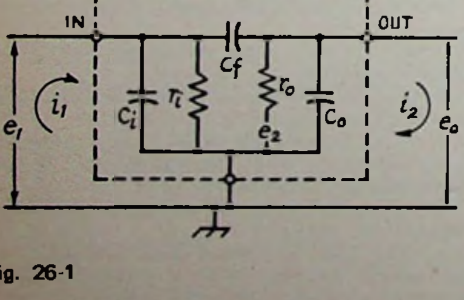


Fig. 26-1

The secondary, L_1C_1, is loaded by the input resistance, r_i, of amplifier AR_2, and the primary, L_2C_2, is loaded by the output resistance, r_o, of AR_1. The selectivity of the circuit depends on these resistances, the inherent Qs of the two circuits, and the coefficient of coupling, as described in Chapters 13 and 14. The ratio e_1/e_o will depend on these same factors; this ratio can be either step-up or step-down, depending on the circuit design.

The method shown in Fig. 26-3 would be most suitable when r_o and r_i are high resistances. Typical cases would be a vacuum-tube pentode for AR_1 followed by a negative-grid tube, or one FET supplying the input voltage for a following one. It would not ordinarily be suitable between two bipolar transistors because r_i is low in these devices and must be matched to the signal source for maximum power transfer. Fig. 26-4, using a low-impedance untuned secondary coil for transferring the signal to the input terminals of the amplifying device, would be better in that case. The lower circuit, in which the output terminal of AR_1 is tapped down on the tuned-circuit coil, can be used if r_o loads the tuned circuit too much to allow the desired selectivity, and also for stabilization as discussed later. These circuits also have been discussed in Chapter 14. Variations are possible; the point is that the coupling circuit must be adjusted to "fit" both the input and output load resistances if operation is to be optimum.

FEEDBACK: The effect of capacitance C_f in Figs. 26-1 and 26-2 is to cause positive feedback, as explained in Chapter 24, when the output circuit L_2C_2 has inductive reactance at the frequency being amplified. If the fed-back voltage (or current) is large enough, a condition which usually occurs with amplifying devices having appreciable feedback capacitance, the circuit will oscillate. Amplification of external signals then ceases, the amplifying device being saturated by its own self-generated signal.

STABILIZATION: Methods of stabilization or preventing self-oscillation are all based on making the fed-back signal at the input electrode (grid, gate, or base, in the circuit under consideration) too small to permit sustained oscillations to occur.

One method of doing this is to reduce the impedance "seen" by C_f looking back from the output electrode to the input electrode. If stray reactances are tuned out, this impedance is equal to r_i in parallel with the resonant resistance of the coupling circuit. This latter, in turn, is the circuit resistance looking back from the input terminals of the amplifier toward the preceding stage. In Fig. 26-4 this resistance can be made smaller by reducing the inductance of L_1 while maintaining tight coupling to L_2 (see Chapters 13 and 14). If the reactance of C_f is much higher than the total of these two resistances in parallel the fed-back current becomes essentially independent of the resistance and depends practically entirely on C_f. Hence, for a given value of C_f, the smaller the total input resistance the smaller the voltage developed across it by the fed-back current.

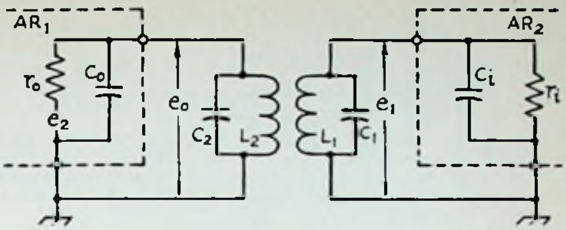


Fig. 26-3

The fed-back voltage also can be reduced by lowering the load impedance. This can be done by tapping the amplifying device's output terminal down on L_2; when this is done less output voltage is applied to C_f. A combination of both schemes can be used. The overall result, however, is that while the feedback voltage is reduced to the point where the amplifier becomes stable, the possible amplification (maximum available gain) cannot be realized. When the feedback has been reduced to the point of stabilization the **maximum usable gain** is obtained.

This method of stabilization is practical only if the transconductance of the amplifying device is great enough so that a reasonable amount of amplification is achieved even though some of the *possible* gain is sacrificed. It is a useful method with bipolar transistors because this type is inherently capable of large gains. Also, bipolar transistors have naturally low input impedance; this tends to discourage self-oscillation caused by C_f, although the impedance is seldom low enough to prevent self-oscillation entirely.

INPUT IMPEDANCE: In tuned rf amplifiers, even when stabilized, the changes in input impedance caused by feedback through C_f are especially disturbing because they affect both the tuning and Q of the input circuit (L_1C_1 in Fig.

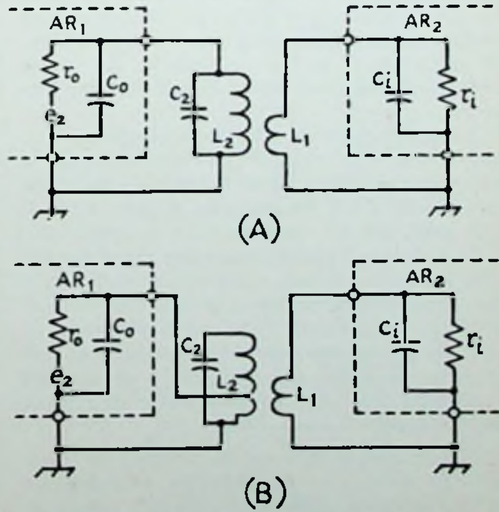


Fig. 26-4

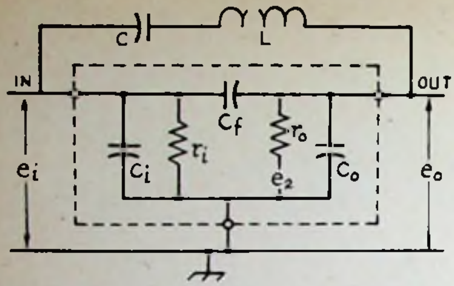


Fig. 26-5

26-2). If the effective input parallel resistance decreases, the Q becomes lower; if the resistance increases or actually becomes negative the Q becomes higher. Thus the selectivity and gain vary with the tuning of the output circuit.

Also, the variation in input capacitance with output tuning amounts to changing the tuning of the input circuit, so the input circuit goes "off tune" when the tuning of L_2C_2 is varied.

These factors make circuit adjustment difficult when the signal frequency is fixed. Also, if the tuning is fixed and the signal frequency is varied through the frequency to which the amplifier nominally is tuned, the response is unsymmetrical about the desired frequency.

REDUCTION OF C_f: An obvious method of reducing feedback of the type described above is to make C_f smaller. However, this capacitance is fixed by the construction of the amplifying device. In a general way it increases with the power-handling capability, since the electrodes in a power amplifier are necessarily larger than in one called upon to handle only small signals. C_f is larger in bipolar transistors than in FETs, and in turn is larger in junction-type FETs than in MOSFETs. It is lowest of all in small vacuum-tube pentodes designed especially for rf amplification. With such pentodes it is seldom necessary to use stabilizing methods except to prevent feedback *external* to the tube. This can be done by shielding the circuits and by maintaining adequate separation between input and output leads so that currents in the output circuit cannot induce appreciable rf voltages into the input circuit.

In amplifiers having tuned circuits connected as shown in Fig. 26-2 (in these cases r_i will be large, as a rule) it is generally necessary that C_f be of the order of 0.01 pF if the amplifier is to be stable. If C_f is large enough so that there is appreciable feedback, but not enough to require the stabilizing method described earlier, stabilization often can be achieved by loading the tuned circuits with external resistance. This lowers their impedances and reduces the gain to the point where oscillations will not occur. However, the selectivity is poorer, and the method may not be permissible in a given amplifier design.

NEUTRALIZATION: It is possible to overcome the effects of feedback by providing an *external* feedback path in an amplifier circuit to oppose the internal feedback through the device. This is called neutralization.

A simple method of neutralizing (coil neutralization) is shown in Fig. 26-5 as applied to the equivalent circuit of Fig. 26-1. It amounts to connecting an inductance, L, between the output and input terminals to balance out the capacitive reactance of C_f. Capacitance C is used simply for isolating the dc voltages on the two terminals from each other; it should have very low reactance compared with the reactance of L, so that the external reactance is essentailly that of L alone.

When the circuit formed by C_f and the external inductance L is parallel-resonant at the frequency being amplified, it acts as a very high resistive impedance at the operating frequency, so the amount of energy fed back through the circuit as a whole is negligible. However, since the circuit operates as described only over the range of frequencies for which the high parallel impedance is maintained, it is useful principally for a fixed-frequency amplifier or for one operating over a band which is only a small percentage of the center frequency.

BRIDGE CIRCUITS: A widely used method of neutralizing an amplifier is one in which the feedback capacitance is made part of a bridge or balancing circuit. In this type of circuit the voltage across one pair of terminals is unable to cause a voltage to appear across a second pair of terminals.

The basic principle of the bridge is readily understandable in terms of the simplest type, the resistance bridge (Wheatstone bridge) shown in Fig. 26-6(A). If an input voltage e_1 is applied between points *A* and *B* and the ratio of R_1 to R_2 is the same as the ratio of R_3 to R_4, Ohm's Law will show that the voltage at point *C*, measured with respect to, say, point *A*, is the same as the voltage at *D* measured with respect to *A*. In other words, there is no difference of potential between *C* and *D* no matter what the value of e_1, and the "output" voltage, e_2, always is zero. The bridge is said to be balanced when this is the case.

Bridge circuits are balanced in both directions. That is, if a voltage is applied between *C* and *D*, the voltage between *A* and *B* in turn will be zero.

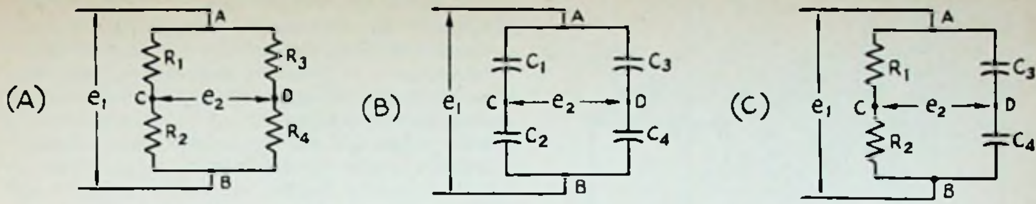
Fig. 26-6(B) shows a similar circuit for four capacitances with an ac voltage e_1 applied. As long as the ratio C_1/C_2 is the same as the ratio C_3/C_4 the bridge is balanced and e_2 will be zero for any voltage e_1. Additional variations are shown at (C) and (D). In all but (A), reactances form at least part of the bridge, but similar relationships hold. That is, in the four circuits

(A) $$\frac{R_1}{R_2} = \frac{R_3}{R_4}$$

(B) $$\frac{C_1}{C_2} = \frac{C_3}{C_4}$$

(C) $$\frac{R_1}{R_2} = \frac{X_{C3}}{X_{C4}} = \frac{C_4}{C_3}$$

(D) $$\frac{L_1}{L_2} = \frac{C_4}{C_3}$$



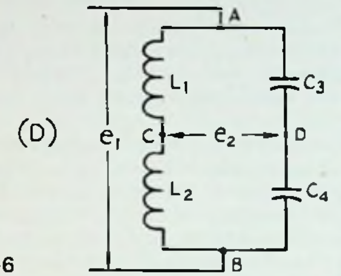


Fig. 26-6

Note that the capacitance ratio is the inversion of the capacitive-reactance ratio because of the way capacitive reactance varies with frequency.

In Fig. 26-6(D) L_1 and L_2 are shown as two separate inductances, but the two can actually be combined in a tapped coil, with the tap serving as point *C*. That is, mutual inductance can exist between the two inductances providing the ratio of the voltages across L_1 and L_2 is the same as the ratio of the voltages across C_4 and C_3.

The bridge circuits of Fig. 26-6 are independent of frequency since, as shown by the equations above, only the quantities R, L, or C appear in any one section or arm, and these quantities are themselves independent of frequency.

The ratios shown above may have any convenient values. The choice of a ratio depends on the capacitance of C_f, which is one of the arms of a neutralizing bridge, and the realizable values of circuit components needed for neutralizing it.

BRIDGE NEUTRALIZATION: Fig. 26-7 shows one form of bridge neutralization and its application to an amplifier, with dc supplies and other irrelevant details omitted. In (A) the bridge is formed by C_1C_2 and C_fC_n, so the circuit is that of Fig. 26-6(B) with C_f corresponding to C_3 and C_n corresponding to C_4. Inductance L between terminals *A-B* is not part of the neutralizing circuit but simply resonates with C_1C_2 at the operating frequency. L, C_1, and C_2 constitute the tuned output circuit for the amplifier in Fig. 26-7(B). When C_n, which is usually adjustable, is set to bring the bridge into balance, the output voltage of the amplifier, developed between *A-B*, is balanced out between *C-D*, the input circuit; thus there is no feedback voltage between the grid and cathode.

The circuit of Fig. 26-7(C) is the same except that one part of the bridge is in the input circuit rather than in the output circuit as in Fig. 26-7(B). Because the bridge works in both directions the overall result is the same.

Circuits of this type are used mainly with vacuum tubes, although in principle they can be used with any type of amplifying device. Fig. 26-7(B) is generally used with triodes, and the capacitance ratio is usually 1 to 1. Fig. 26-7(C) is commonly used with pentodes or tetrodes. Because C_f in these tubes is very small, C_n is usually larger than C_f, thus the capacitance ratio C_1/C_2 is smaller than 1. These two circuits are used almost exclusively in transmitting applications.

OTHER NEUTRALIZING CIRCUITS: The circuits in Fig. 26-7 are inconvenient in amplifiers that have to be continuously tuned over a wide frequency range, as in receivers. This is because they increase the number of variable capacitors that must be "ganged" – that is, operated from a common control shaft – in multistage amplifiers. It is necessary that C_1 and C_2 be varied together in such tuning, because the ratio of the two must conform to the fixed ratio of C_f to C_n.

Since the inductance in a tuned circuit is usually fixed, for a given tuning range, circuits of the general type shown in Fig. 26-6(D) are often used in such cases. Instead of using two coils, L_1

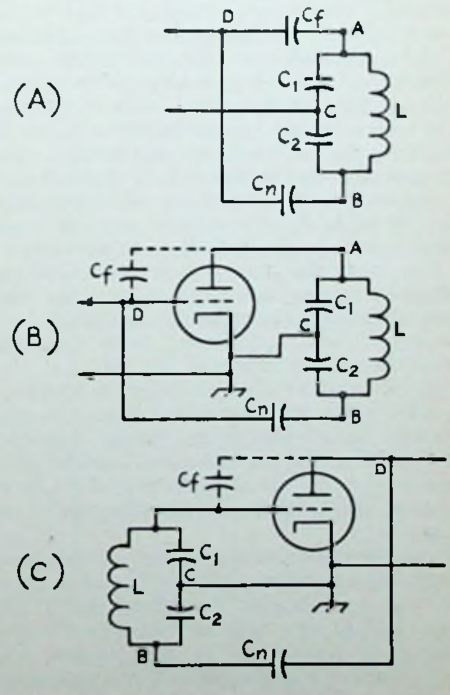


Fig. 26-7

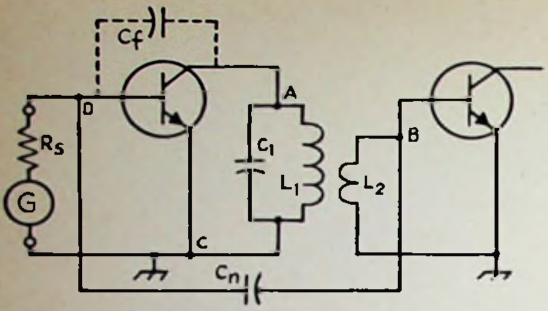


Fig. 26-8

and L_2, in series across the signal source (or the single-coil equivalent having a tap) the necessary division in voltage for balancing the bridge is often obtained by making L_2 a separate coil inductively coupled to L_1. This is shown in Fig. 26-8 as applied to a bipolar amplifier. The circuit makes use of the step-down transformer, $L_1 L_2$, normally used for matching the input impedance of AR_2 to the load impedance required by AR_1, so the voltage between *C* (ground) and *B* is considerably less than the voltage between *C* and *A*. Therefore C_n must be larger than C_f by the same ratio. The relative winding directions of L_1 and L_2 must be such that the fed-back voltage between *C* and *D* will be zero; if the phase is reversed, the feedback is positive and the circuit will oscillate.

PHASE RELATIONSHIPS IN PRACTICE: This discussion of neutralization has been based on the assumption that the unwanted feedback in an amplifying device is solely through a capacitance, C_f, between the input and output electrodes, and that phase relationships in the neutralizing circuit are determined entirely by reactances. In practice this is not the case. Resistance is always present, and it affects phase as described in Chapter 8. Thus, depending on the relationship between resistance and reactance, which in turn depends considerably on the components used, neutralization is always more or less incomplete. At any given frequency the circuits can be modified to make sure that both the phase and amplitude of the fed-back voltages are balanced accurately, but this leads to undesirable circuit complications.

In most cases, fortunately, really *complete* neutralization is not required. The object in neutralizing an amplifier is not necessarily to eliminate feedback entirely but to reduce it enough so that oscillations cannot start under normal operating conditions. Also, the neutralization should be good enough so that the effects on the amplifier's input impedance, described above and in Chapter 24, are not serious.

These two objectives can be met without great difficulty in amplifiers using FETs and vacuum tubes, but with bipolar transistors this is not always true. In this type of transistor C_f (the collector-base capacitance in the circuits considered) is shunted by the collector-to-base leakage resistance. Also, the characteristics of these devices, including both the resistances and capacitances, are temperature- and voltage-dependent, so a circuit can be neutralized only for a restricted set of conditions – constant temperature, constant bias and supply voltages, and a low-enough level of signal input and output voltages so that the device characteristics are not appreciably altered instantaneously by the signal-voltage excursions. When a bipolar transistor handles appreciable amounts of dc input and signal power, stabilization by overloading to reduce the gain, as described earlier, is the only practicable method. In addition, the usual precautions must be taken (see Chapter 22) to stabilize the bias voltage against temperature effects.

Neutralization of any type of amplifying device tends to become less effective as the operating frequency is increased. The actual circuits become much more complex, not only because of the resistances mentioned above (which generally become more troublesome as the frequency is raised) but also because the lengths of leads, both inside and outside the circuit components, introduce stray inductances in various parts of the neutralizing circuit. At vhf (and even somewhat lower) these inductances are not negligible. Since they are for the most part in series with the capacitances in the neutralizing circuit, they make the circuit frequency-sensitive. This restricts the range over which neutralization will be adequate. In addition, inductance in the lead (both internal and external) from the common electrode is a source of feedback which neutralizing circuits of the type described above do not overcome.

OTHER AMPLIFIER CIRCUITS: At radio frequencies tuned circuits can be substituted for the input and output resistances in Figs. 25-1 and 25-3, Chapter 25. Because the follower circuits do not give a voltage gain they are seldom used as small-signal rf amplifiers; their principal application is as step-down impedance transformers having a tuned input circuit and a resistive output circuit. The inherently large negative feedback makes these circuits quite stable, but it is necessary to eliminate unwanted feedback paths, such as through dc supplies or by external coupling between input and output, to ensure stability. (This, of course is also true of other circuits.)

The grounded-grid, -gate and -base circuits with tuned output, LC, and feedback capacitance, C_f, are shown in Fig. 26-9. It is possible for these circuits to go into oscillation, but a larger value of C_f is required than in the circuits discussed earlier in this chapter. The inherently low input resistance is responsible for this. The circuit is especially useful with bipolar transistors because these devices have very low base-emitter resistance. This in combination with the low input resistance of the circuit itself, offers very little opportunity for appreciable feedback voltage to be developed in the input circuit. The circuit has a moderate amount of voltage gain, and the power gain is considerable.

This type of circuit is frequently used in vhf small-signal amplifiers because of its power gain and stability, with all types of amplifying devices.

FREQUENCY LIMITATIONS: Various factors operate to set an upper frequency limit to useful amplification. Included in these are the internal capacitances and lead inductances of the amplifying devices, which cause phase shifts and unwanted feedback, and the finite transit time (see Chapter 19) of electrons in vacuum tubes. The useful upper frequency limit is often given in vacuum-tube and transistor data sheets.

Special considerations apply in bipolar transistors. When a bipolar transistor is switched on by the emitter-base signal voltage, a finite time is required for the carriers to cross the emitter-base junction, diffuse through the base, and begin crossing the base-collector junction. This is called the delay time, and represents an interval (typically of the order of micro- or nanoseconds) between turn-on in the emitter-base diode and the beginning of collector-current flow. When the transistor is switched off by the signal, many of these carriers are left (stored) in the base, and again a finite time (storage time) elapses before they all cross the base-collector junction. When the period of the ac signal becomes comparable with the delay and storage times very little alternating current can be developed in the collector circuit, and amplification ceases. Fig. 26-10 shows typical variation of α and β with frequency in a small transistor designed for use in receiving-type vhf amplifiers.

This behavior with frequency can be expressed in terms of a **cutoff frequency**, which is defined as the frequency at which the current gain of a bipolar transistor decreases to 0.707 times its value at 1000 Hz. The transistor is characterized by two such frequencies, α cutoff (common-base circuit) and β cutoff (common-emitter circuit). The frequency at which the common-emitter gain drops to 1 is called the gain-bandwidth product and is designated f_T. The operating frequency must be well below f_T if appreciable amplification is to take place. (Note that amplification can continue to take place above this frequency if the common-base circuit is substituted for the common-emitter circuit.) Data sheets on bipolar transistors usually specify f_T, although in some cases the highest *useful* frequency may be given, usually in terms of the ac current gain (h_{fe}) at a specified frequency.

RADIO-FREQUENCY OSCILLATORS: The use of an amplifying device to generate a signal frequency has been described in a general way in Chapter 24. The sine-wave type of oscillator has a variety of uses in radio communication. Several basic circuit forms, omitting power-supply and similar details, are shown in Fig. 26-11.

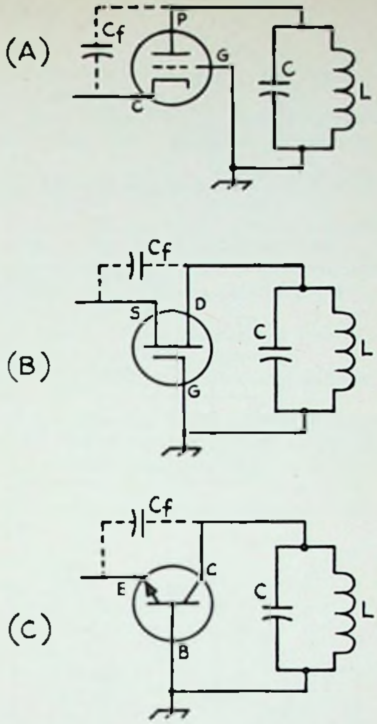


Fig. 26-9

The circuits at (A) and (B) are more or less equivalent. The difference is in the way the tuned circuit is divided so that one section, *X-Y*, is connected between the input and common electrodes and the other section, *Y-Z*, is between the output and common electrodes. The opposite ends, *X* and *Z*, of the tuned circuit are in opposite phase with respect to point *Y* when a voltage is developed across the circuit; therefore if *Y* is connected to the cathode, source, or emitter of the amplifying device used, the phases at the input and output electrodes will be the same as in normal amplification. That is, the circuit provides positive feedback.

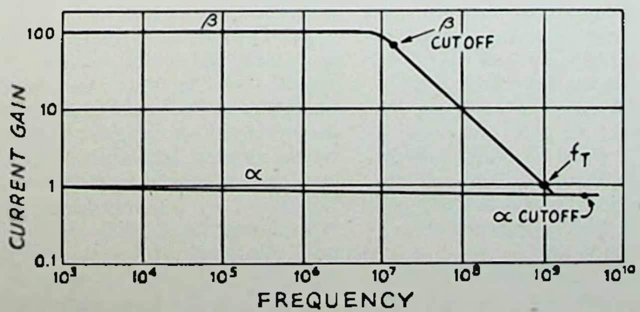


Fig. 26-10

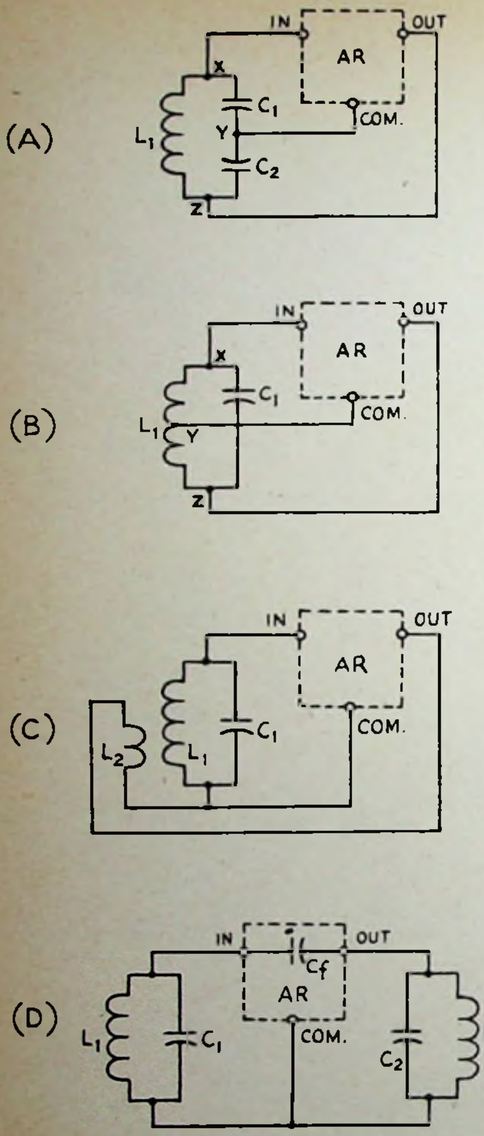


Fig. 26-11

If point Z is connected to the output electrode (plate, drain, or collector) the load on the amplifying device is the tuned circuit as seen through terminals Y-Z. The load can be varied by moving the common point, Y, along the tuned circuit. In (A) this is done by varying the ratio of C_2 to C_1, L_1 being adjusted as necessary to maintain the same resonant frequency. In (B) the load is adjusted by moving tap Y along L_1. Adjusting the position of Y in either circuit also adjusts the amount of feedback, because the proportion of the voltage between X and Z that appears between X and Y depends on the position of Y.

Fig. 26-11(C) is another form of (B). Instead of a tap, a second coil L_2 introduces the feedback voltage into the tuned circuit, L_1C_1. This simply substitutes mutual inductance for the tap. The feedback can be controlled by varying the mutual inductance, either by changing the number of turns on L_2 or by changing the coupling between L_2 and L_1. Since the relative phase is determined by the relative winding directions of the coils, the feedback can be either positive or negative with this circuit, and for oscillation to take place the voltage induced from L_2 into L_1 must be such as to reinforce the original signal at the input electrode. This circuit is often more convenient constructionally than Fig. 26-11(B).

Fig. 26-11(D) will be recognized as an unneutralized radio-frequency amplifier with parallel-tuned input and output circuits. For developing the positive feedback required for oscillation the net reactance of L_2C_2 must be inductive at the operating frequency; in turn, this means that L_2C_2 must be tuned to a slightly *higher* frequency than that to which L_1C_1 is tuned (see Chapter 24). There is an optimum degree of detuning for maintaining oscillation; this depends on the characteristics of the amplifying device and on the circuit constants and losses. The oscillator frequency also depends somewhat on these three factors.

FREQUENCY STABILITY: The oscillation frequency in all these circuits is determined primarily by the resonant frequency of the tuned circuit – or circuits, in Fig. 26-11(D). The stray capacitances shunting the circuits must be included in the total capacitance in each case, in calculating the approximate frequency at which oscillation will take place. The frequency is affected to some extent by the various resistances in the circuits themselves and in the amplifying devices, since these affect the phases of the input and output voltages. The generated frequency will automatically be the one at which the voltage at the input electrode is in the same phase as the voltage already present.

The ability of an oscillator to maintain a given frequency under changing conditions such as varying power-supply voltages is known as its stability. In general, the higher the Q of the tuned circuit the better the stability. The tuned circuit is loaded by the input and output resistances of the amplifying device, as well as by any other power-consuming resistances that might be present, so for high Q the loading must be as light as possible. There are various practical ways of adjusting the loading, all operating on the principles that have been described in Chapter 13.

The output of an oscillator using LC circuits is essentially sinusoidal, because a tuned circuit responds best to only one frequency. Although distortion in the amplifying device itself may result in generation of frequencies harmonically related to the desired fundamental frequency, these harmonics are poorly amplified, if amplified at all, because of the low impedance of the tuned circuit at such frequencies. The result is that there is very little feedback of harmonics, and therefore the harmonic amplitudes are not built up, provided the tuned-circuit Q is reasonably high.

Questions and Problems

26-1) Why is it usually necessary to use shielding on the input and output circuits of a radio-frequency amplifier?

26-2) Draw simple circuits showing interstage coupling circuits typically used between (a) two vacuum-tube pentodes; (b) two field-effect transistors; (c) two bipolar transistors.

26-3) What are the principal differences between amplifiers operating at radio frequencies and low frequencies? What accounts for these differences?

26-4) What is coil neutralization? What are its limitations?

26-5) Describe a method of stabilization commonly used with bipolar transistors.

26-6) Why is a vacuum-tube pentode usable without stabilization?

26-7) What is neutralization? What does it accomplish?

26-8) What is meant by the terms delay time and storage time?

26-9) What factors determine the oscillation frequency of an oscillator?

26-10) What is the effect, if any, of frequency on the input impedance of a radio-frequency amplifier?

26-11) Does a bridge stay balanced at all frequencies? If so, under what conditions?

26-12) What are the advantages of the grounded-grid, grounded-gate, or grounded-base circuit at radio frequencies? What are the disadvantages?

26-13) What is the effect of tuned-circuit Q on the stability of an oscillator?

26-14) How do delay and storage time affect the operation of a bipolar transistor?

26-15) Can amplifiers be stabilized without affecting gain?

26-16) Compare the common-base and common-emitter circuits in radio-frequency amplifiers.

26-17) On what principle does a bridge circuit operate?

26-18) What is meant by f_T? What is it called?

Experiments

The experiments which follow, designed to emphasize certain basic principles of circuit and amplifying-device operation, are confined to ones that can be performed with the simplest possible array of equipment. Only three principal items are needed: an electronic voltmeter, a multirange volt-ohm-milliammeter (20,000-ohms-per-volt or higher, if possible) and a dc power supply with adjustable output voltage and a rated current of about one-half ampere (500 mA). A voltage range from 0-30, or even 0-15, is adequate. The output voltage preferably should be electronically regulated in order to obviate frequent readjustment to a fixed value required in an experiment.

These three pieces of equipment are available at moderate prices in kit form. As they are practically indispensable for work at home in electricity, radio and electronics, their purchase should be regarded as a long-term investment. Some or all of them may, indeed, already be in the possession of the reader.

In additon, various small components will be required, chiefly resistors, capacitors, and the like. All components needed for a particular experiment are listed at the beginning of each. It is also helpful to have a dozen or so flexible leads, perhaps a foot long, with small alligator-jaw clips at both ends. Such leads make soldering unnecessary and permit quick assembly of an experimental circuit.

Performing an experiment, after having read and understood the text applying to its subject, is unquestionably the best way of fixing the principles in one's mind. Very often, too, the observed effects raise questions that illuminate a principle and bring out implications that perhaps were not fully appreciated before the actual attempt to demonstrate that principle. For these reasons the reader is urged to perform the experiments if it is at all possible. However, even if he cannot, a good deal can be gleaned from the data and explanations given as part of the discussion of each experiment, so it is worth studying them even if actual performance is not feasible.

One reason why actually working with circuits and instruments helps in gaining a good overall grasp of a subject is that while an electrical law may, theoretically, be quite simple and straightforward, practice and theory seldom agree *exactly*. This is not necessarily because something is wrong with the theory. More than likely it is the fault of measurements or components, neither of which ever comes up to the theoretical ideal. This discrepancy must always be held in mind in practical work. The following paragraphs discuss some of the more prominent causes, and should be read carefully before attempting the experimental work.

On Making Measurements

In most of the experiments in this book, measurement of one or more electrical quantities is a prominent feature, so it is necessary to appreciate what measurement really means. No measurement is "absolute." It can never give the "true" value of a quantity, in the sense that counting a number of discrete objects can lead to a true count. If the number of people in a gathering is 10, there can be no more and no less than 10; each person counts as one even though all may have (and will have) different sizes and shapes.

However, if you try to measure the *weight* of one individual, the best you can say is that within the limits of the inherent accuracy of the scale and your ability to read it correctly, the weight is probably not more than such-and-such and not less than so-and-so. The figure obtained, then, lies between limits.

These limits are usually expressed as the percentage departure from their mean. For example, a current may be given as 100 milliamperes plus or minus 5%, meaning that it will not be greater than 105 mA nor less than 95 mA. It is *not* 100 mA "absolutely", because with another instrument another observer might obtain an appreciably different value. In fact, even with the *same* instrument a different observer probably would come up with a slightly different figure.

Thus there are two sources of error in obtaining a reading from an instrument, the error inherent in the instrument itself, and the personal error of the observer in reading it.

Personal Error

Aside from simple mistakes in interpreting the meter scale, personal errors arise from parallax and from poor estimating of values between scale markings. The latter – interpolation error – can be reduced by training oneself in estimating decimal parts of a length. It is best not to attempt to split divisions into too-small intervals, especially when the space between the calibrated division lines is small. However, it is convenient to "guess" in tenths of a division, and satisfactory to do so if it is always kept in mind that such an estimate may be off by several tenths.

Parallax is the visual difference in readings when the meter is viewed from different angles. It arises because the pointer has to be a little above the scale in order to move, so there is a finite space between the pointer and scale. The best way to overcome it is to develop the habit of looking at the meter perpendicularly to the scale and the line of the pointer so that the visual angle between the two is always the same. Some meters have mirrors under the scale to help reduce parallax; by lining up the pointer and its reflection in the mirror so that the two coincide, parallax is minimized.

The personal error also can be reduced by averaging several readings to get a "most probable" value. For example, the voltage across a circuit element may be read five different times, with the following results.

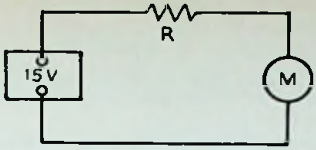
No. 1 – 24.5 volts
No. 2 – 24.3 volts
No. 3 – 25.1 volts
No. 4 – 24.4 volts
No. 5 – 24.8 volts

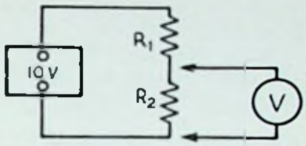
Unless some extenuating conditions make it possible to say without doubt that one or more of these readings is definitely wrong, the *average* of the five – in this example, 24.6 volts – should be used as the true reading.

Instrumental Error

Instrumental error is a matter of the quality of the instrument, which naturally is related to its cost. The accuracy is usually specified by the manufacturer for instruments of good quality.

A consequence of instrumental error is that no two instruments, even two of the same make and model, can be expected to give exactly the same reading on a given quantity. A common rating for accuracy of dc milliammeters and voltmeters, for example, is "plus-or-minus 2% of full scale". This means that a voltmeter having a 0-10 range can be expected to be "off" by as much as 2% of 10 volts – that is, 0.2 volt – *at any part of the scale*. Let us say that the voltmeter reads 5.4 volts in a given circuit. Since the error can be as much as 0.2 volt either way, the actual voltage will lie somewhere between 5.2 and 5.6 volts. Suppose that a precision voltmeter having 0.1% tolerance is available and shows that the voltage actually lies between plus or minus 0.1% of 5.25 volts – that is, within about 0.005 volt either way. Since this is far closer to the truth than the value given by the instrument with 2% tolerance, the actual voltage can be taken to be 5.25 volts. The 2% instrument is within its tolerances in calling the voltage 5.4, an error of 0.15 volt. But a second instrument of the same type might give any reading between 5.05 and 5.45 volts and be within *its* rated tolerances. If its readings happen to be low and it reads 5.05 volts, the difference between the two instruments is 0.35 volt. However, the difference *could* be as much as 0.4 volt at the limits of the tolerances. Don't expect too much in the way of exact agreement between two instruments even though they are supposedly identical!





In-Circuit Errors

Unfortunately, the inherent accuracy of the instrument is only one factor. The way in which it is used may be even more important.

No instrument can be inserted into a circuit for a measurement without becoming a *part* of the circuit. As such, it has current flowing through it, and this current may be large enough to have an appreciable effect on the measurement.

An example or two will make this clear. In the upper circuit of the accompanying figure an emf of 1.5 volts is applied to a resistor, R, and a meter, M, in series. If R is 15,000 ohms one would expect the current to be 100 microamperes, from Ohm's Law, and a 0-100 microammeter should read full scale. However, if such a circuit is tried the microammeter is likely to read less than 90 microamperes, for the reason that a practical microammeter of such range might have an internal resistance of 2000 ohms. The *actual* resistance in the circuit, therefore, is 17,000 ohms, not 15,000. On the other hand, if R were 150,000 ohms the same meter would read 10 microamperes, very closely, because its 2000-ohm internal resistance adds only a negligible amount in comparison to 150,000, and the error caused by the difference between the actual and ostensible resistances is far less than the measurement errors already discussed.

In the lower drawing a voltmeter, V, is being used to measure the voltage across resistor R_2, which is in series with R_1, with an emf of 10 volts applied to the two. If R_1 and R_2 are both 10,000 ohms, exactly 5 volts should appear across R_2. But suppose that V is a 0-10 voltmeter having a resistance of 1000 ohms per volt, which might be typical of a panel-type instrument having a built-in multiplier. Then the meter resistance is 10,000 ohms, and when the meter is connected into the circuit this resistance is in parallel with R_2. The total resistance of the two is 5000 ohms, so the total resistance in the circuit is 15,000 ohms, not 20,000. As a result, only 1/3 of the applied voltage appears across R_2, so the meter reads 3-1/3 volts. A different instrument having the same range but a resistance of 20,000 ohms per volt would place 200,000 ohms across R_2, making the combination of the two in parallel approximately 9500 ohms. In this case the reading would be about 4.87 volts – much nearer the value without the meter connected, although still a larger error than the meter accuracy warrants. A meter of still larger resistance would have to be used to approach the "true" reading in such a circuit – at least 100 times as large as the resistance across which the meter is connected.

The lesson in this is that meter resistance must never be neglected in interpreting measurements and estimating their accuracy. No reading can be regarded as accurate (within the meter's rating) unless it is known that inserting the meter will result in essentially no observable effect on the circuit conditions. If these conditions and the available meter make this impracticable, due allowance must be made for the effect of the meter on the circuit.

Component Tolerances

Apparent discrepancies between measurement and calculation can arise from another source. Electrical values of components – resistance, capacitance, and so on – are always given by the manufacturer in terms of a tolerance, sometimes implied, sometimes expressed, but never absent. Rated tolerances of components compare with rated accuracies of instruments and, like the instrument accuracies, are expressed in terms of percentage departure from a mean value. A resistor might be rated at 1000 ohms ± 10%, for example, meaning that its *actual* value can lie anywhere between 1000 + 10% and 1000 - 10%; that is, between 900 and 1100 ohms.

It is obvious that the voltage drop across such a resistor when carrying a constant current can also vary as much as ± 10%, even with ideally-accurate voltage readings. If these tolerances are not kept in mind, discrepancies may be ascribed to instrument errors which, in fact, are not to blame.

In summary, there are many unavoidable reasons why a simple measurement – such as, for example, the current and voltage associated with a resistance – may lead to seeming departures from well-established electrical laws. While these uncertainties pervade all measurements, the process of allowing for them should not be carried to the extreme of blaming them for what actually is careless work in making measurements. When a discrepancy occurs between measured and theoretical values, every effort should be made to account for the difference. This, in turn, may suggest a technique by which it can be reduced to a minimum.

Experimental Procedure

Observations taken in the course of the experimental work should be recorded in a notebook kept especially for the purpose. Observations should be as complete and detailed as possible. It will be helpful, too, to include with the data a written explanation of the phenomena observed and the results secured, including reasons why (when necessary) there is a difference between experiment and theory. Such written notes are invaluable as a means of fixing the principles and practice firmly in mind. Should the subject matter under study suggest further experimental possibilities within the scope of the equipment, the exploration of such possibilities cannot help but be beneficial, if the work is carried out in a methodical way and an attempt is made to provide a logical interpretation of the results.

Experiment 1

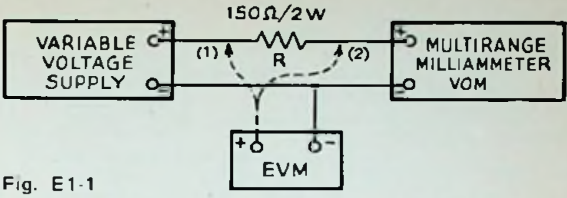


Fig. E1-1

PURPOSE: To demonstrate some of the relationships between voltage, current and resistance as expressed by Ohm's Law. This experiment also serves to bring out the importance of tolerances, and illustrates a measurement-circuit error that frequently is overlooked.

APPARATUS: Electronic voltmeter, (EVM) multirange milliammeter (in VOM), variable-output dc power supply, and an assortment of 10%-tolerance carbon resistors.

PROCEDURE (A): Connect the apparatus as shown in Fig. E1-1. Note that the positive prod of the EVM can be touched to either (1) or (2) for measuring the voltage at these two points. The negative or common lead can be clipped to the negative output terminal of the power supply throughout this experiment. Only one resistor (R) is used: it should be 150 ohms, rated at 2 watts.

Starting with 2 volts at the power-supply terminals, take current readings at 2-volt intervals up to 14 volts. Record the voltage or current range of the instruments in each case. Following is a typical set of data:

Power-supply Voltage (1)	*Voltage Range*	*Voltage at (2)*	*Voltage Range*	*Current mA*	*Current Range*
2	0-10	0.13	0-1	13.2	0-50
4	"	0.26	"	26.0	"
6	"	0.4	"	39.0	"
8	"	0.08	"	57	0-500
10	"	0.1	"	69	"
12	"	0.115	"	81	"
14	"	0.135	"	97	"

During the measurements check the EVM's zero reading frequently, with no voltage applied. An electronic voltmeter will "drift", particularly when first turned on, and to prevent the readings from being in error, the meter must be "re-zeroed" when the pointer drifts off the zero mark. Also, it is a good idea to go through the procedure three or four times and average the readings, as explained in the section "Personal Error."

Comment: If there are no other variables, the current through a fixed resistance is exactly proportional to the voltage applied to it. A plot of this relationship, Ohm's Law, on rectangular cross-section paper will be a straight line, hence the relationship is called linear. Such a plot for a 150-ohm resistor over the voltage range used in this experiment is shown in Fig. E1-2. However, no series of measurements can be expected to lead to such ideal results; the best that can be done is to approach them as closely as possible by holding the measurement errors to a minimum.

The points from the first and fifth columns of the data tabulated above are shown by small circular dots in Fig. E1-2. None of them fall exactly on the straight line. Also note that while the points tend to fall *below* the line up to 6 volts applied, there is a shift at 8 volts which puts all the remaining points *above* the line. This coincides with the change in milliammeter scale, and emphasizes the desirability of recording *every* pertinent fact about a set of measurements. Without the knowledge of the scale change it would be hard to account for this abrupt shift.

Actually, the shift is caused by the circuit, not necessarily by meter inaccuracy. The concealed source of error here is the meter resistance (see section on "Instrumental Error"). It is much lower on the 0-500 scale of the instrument than on the 0-50 range. Since the resistance is in series with resistor R the *total* resistance of the circuit is less when the 0-500 scale is in use, hence the current is relatively larger. This is also shown by the voltage across the milliammeter alone, listed in the third column of the tabulation. There is a sudden drop in this voltage when the current scale is changed.

The *difference* between the voltages in the first and third columns represents the actual voltage applied to resistor R. If these differences also are plotted on the graph, as shown by the small crosses in Fig. E1-2, it can be seen that all points actually fall on the same side of the ideal straight line (the differences are so small when the 0-500 range is used that they cannot be seen on a graph of this size and so these additional points have been omitted). Furthermore, the corrected points form almost a straight line themselves. The inference to be drawn from this is that the measurement accuracy is not too bad, and that the resistance of R actually is a little less than 150 ohms. If R is calculated for each of the actual applied voltages and the results averaged, it is found that R is approximately 143 ohms, instead of its nominal value of 150. This is

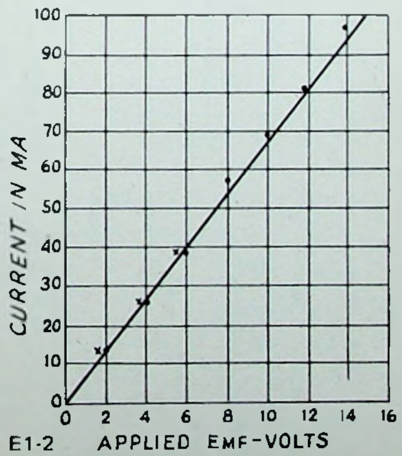


Fig. E1-2

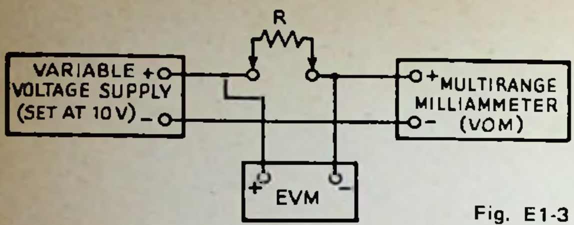


Fig. E1-3

well within the 10% tolerance rating of the resistor, and so is not surprising.

PROCEDURE (B): Connect the apparatus as shown in Fig. E1-3. In this experiment, change R for each measurement so that the dependence of current on the resistance in the circuit can be shown. Use clip leads to make connections to R. Adjust the applied voltage to exactly 10 volts across R each time. Note that the EVM is connected only across R, so that the resistance of the milliammeter is not a factor.

Measure an assortment of resistors covering a fairly wide range of resistance. In the lower values, make sure that the power ratings of the resistors are adequate; above 200 ohms all resistors can be 1/2 watt.

A typical series of measurements is shown below:

Nominal resistance, ohms	*Current, mA*	*Current Range*	*Calculated Current, mA*
47 (2 watts)	220	0-500	230
100 (1 watt)	98	"	100
150 (1 watt)	62	"	67
330	28.7	0-50	30.3
680	15.2	"	14.7
1000	9.6	"	10
2200	4.58	0-5	4.55
6800	1.43	"	1.47
10000	1.05	"	1

The "Calculated Current" column lists the theoretical value from Ohm's Law, for the exact resistance in the first column. The two values of current would match exactly only under ideal conditions.

In addition to making the measurements, check the resistances used in the experiment with an ohmmeter. Because ohmmeter scales are more cramped than the voltmeter and milliammeter scales and are not linear, such a check will not, in general be as accurate as a voltage/current measurement from which the value of the resistance can be calculated. However, the ohmmeter check serves to give an idea of the relative measurement dependability of the ohmmeter and the voltmeter/milliammeter combination. It also will uncover gross errors such as might be caused by reading the wrong current scale in the measurements.

Comment: Even if measurement errors somehow could be overcome, it is not to be expected that the measured values of current would exactly equal the calculated values, because the tolerance on the resistances is plus or minus 10%. To some extent the presence of measurement errors can be detected, because in a single series of measurements such as this the measured values tend to be consistently either high or low, while the resistor variations can be expected to be random in a sample of 8 or 10 resistors. There is a trace of a measurement-error pattern in the above data, as most of the measured currents are lower than calculated. However, in no case is the discrepancy between measured and calculated as much as 10%, so it can be assumed that all the resistor values are within tolerances, and likewise that the measurement errors are within the meter tolerances.

In this experiment a principal objective has been to build up an awareness of the ever-present errors in making measurements (the experiment itself is quite simple in principle) and, equally, an awareness of tolerances in the marked values of components. Tolerances are an integral part of every measurement, whether mentioned explicitly or not.

Experiment 2

PURPOSE: To demonstrate the rules governing resistances in series, parallel, and series-parallel; internal resistance and voltage regulation; voltage-divider operation.

APPARATUS: Same as in Experiment 1.

PROCEDURE (A): Three resistors are used; values of 100, 220, and 680 ohms, 10%, are suitable. The first two should be rated at 1 watt; the last can be 1/2 watt. Four different circuit arrangements are used, as shown in Fig. E2-1(A). Adjust the power-supply voltage to be exactly 10 volts with positive prod of the EVM at point *X*; readjustment will be required for each circuit change and each change of milliammeter scale.

In the series arrangement measure the current and the voltages (to common) at points *Y* and *Z*. Form the parallel circuit progressively by first using only the 680-ohm resistor, then adding the 220-ohm, and finally the 100-ohm. Measure the total current each time. In the first series-parallel circuit, first leave the 220-ohm resistor unconnected and measure the current, also the voltage at point *Z*. Then connect the 220-ohm resistor as shown and repeat the measurements. In the second series-parallel circuit (move the lower connection of the 220-ohm resistor to form this) measure the total current.

Typical data are as follows:

	Measured Values	*Theoretical Values*
Series connection:		
Current	10.5 mA	10 mA
Volts at Z	1.1 V	1 V
Volts at Y	3.15 V	3.2 V
Parallel connection:		
Current with 680 ohms	15.5 mA	14.7 mA
Current with 680 and 220 ohms	65 mA	60.2 mA
Current with 680, 220, and 100 ohms	160 mA	160 mA
First series-parallel:		
Current without 220 ohms	13.3 mA	12.8 mA
Volts at Z	1.4 V	
Current with 220 ohms	38.4 mA	37.5 mA
Volts at Z	4.0 V	3.76 V
Second series-parallel:		
Total current	63 mA	58.3 mA

Comment: The theoretical values listed in the data are based on exact values of resistance (100, 220, and 680 ohms). The measured and theoretical values of current are seldom identical, because of measurement errors and resistor tolerances. However, if a measured value departs more than 10% or so from the theoretical, it would be well to recheck the measurements, as well as the resistors used.

The measured data contain enough values to give three independent voltage and current measurements on each individual resistor. The actual resistance in each case can be closely approximated by using Ohm's Law and taking the average of the three resistance values so obtained. This procedure

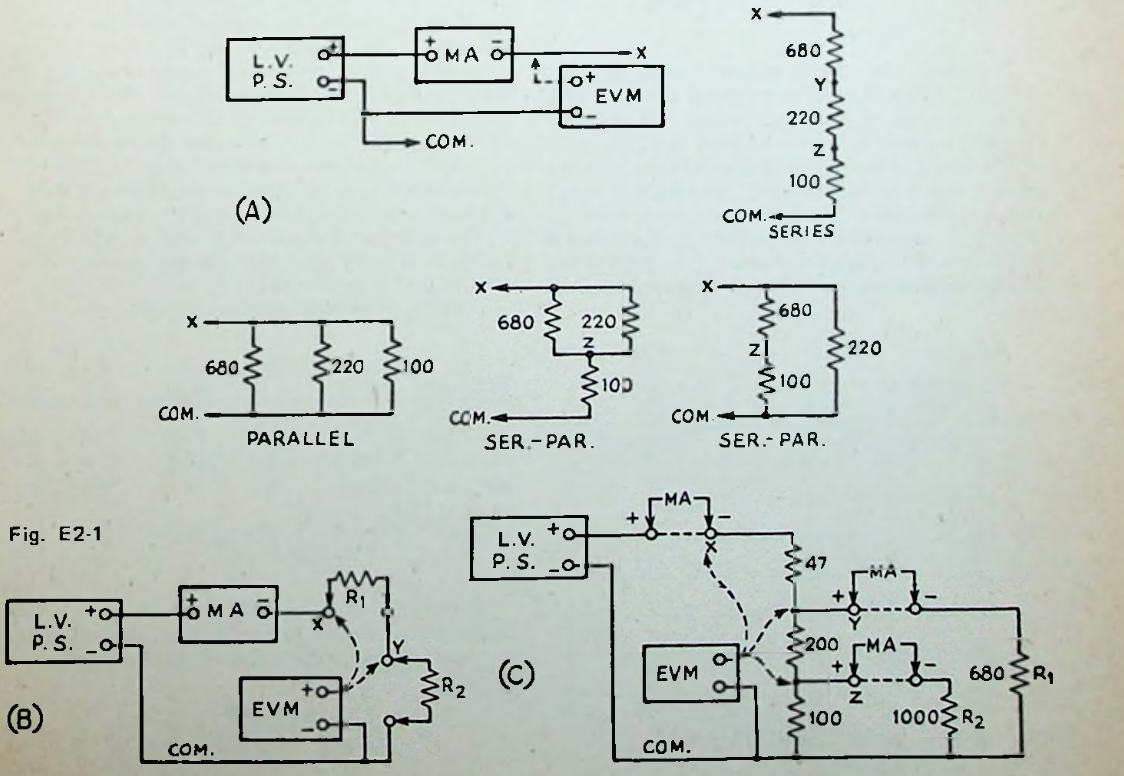


Fig. E2-1

applied to the data above leads to values of 105, 199, and 650 ohms for the three resistors. Working out the theoretical currents and voltages using these figures leads to much closer agreement between measured and theoretical.

PROCEDURE (B): In this experiment R_1 simulates the internal resistance of a power supply and R_2 is the load. Several values of both resistors are used; suitable ones for R_1 are 5, 10, and 25 ohms. A slider-type 25-ohm wire-wound resistor may be used if fixed resistors having these low values are not obtainable. Set the slider to the proper value using an ohmmeter. Power ratings of 1/2 watt will suffice for R_1. For R_2, use about five separated values – 47 ohms, 2 watts; 100 ohms, 1 watt; and 330, 680, and 1000 ohms, 1/2 watt.

Maintain the voltage between point *X* and the common negative at exactly 10 volts. Using 5 ohms at R_1, insert various resistances at R_2 and measure the voltage at point *Y*, along with the current. Repeat with 10 ohms at R_1, and finally with 25 ohms at R_1.

The following currents and voltages are typical, with 10 volts at *X* for every measurement:

R_2 ohms	R_1 = 5 ohms Volts at Y	mA	R_1 = 10 ohms Volts at Y	mA	R_1 = 25 ohms Volts at Y	mA
47	9.25	205	8.25	185	6.65	145
100	9.7	95	9.15	88	8.2	80
330	9.85	28.5	9.7	28.4	9.35	27.2
680	9.92	15.4	9.85	15.3	9.65	15.0
1000	9.96	9.5	9.90	9.6	9.80	9.5

Comment: By maintaining the voltage constant at *X* and measuring the voltage at *Y* across various values of load resistance, (output voltage) R_1 becomes equivalent to the internal resistance of a power supply. The equivalent circuit of every source of power is a constant-voltage generator in series with an impedance. The impedance, being internal, cannot be separated from the generator for separate measurement. Voltage-vs.-current data such as are tabulated above can be used to plot regulation curves of the type shown in Fig. E2-2. The curves are linear, since the generator impedance in this case is a simple resistance. The larger the internal resistance the more rapidly the output voltage drops off as the current is increased. The points from which any such curve is constructed should fall very close to a straight line; if one does not, the measurement should be checked.

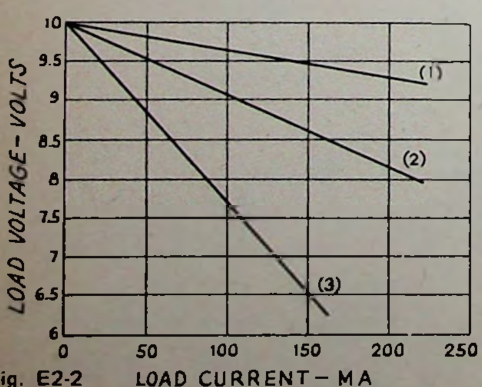


Fig. E2-2

The internal resistance of the supply can be found from regulation curves such as these. The slope of the curve – i.e., the number of volts per ampere – gives the internal resistance in ohms. In the case of linear curves it can be found quite easily by taking the *difference* between the open-circuit voltage and the voltage at any selected value of current, and dividing this difference by the selected current in amperes. For example, from curve (3) the output voltage at 150 mA is 6.5 V, so the internal resistance is

$$\frac{10 - 6.5}{0.015} = 22.5 \text{ ohms.}$$

The nominal value of the resistance (25 ohms) used for getting the data for this curve had been obtained by ohmmeter measurement on a slider-type adjustable resistor. The value obtained by the above method is more accurate.

PROCEDURE (C): In this procedure measurements are made on a voltage divider of the type shown in Fig. 3-5, Chapter 3, to check its voltage regulation. The test circuit is given in Fig. E2-1(C). It is necessary to shift the milliammeter for measuring the current at three points in the circuit. One (*X*) is the total current to the system. At *Y* the current through R_1 is measured, and at *Z* the current through R_2. The voltage at *X* should be kept at 10 volts throughout the measurements. Voltage measurements at *Y* and *Z* include the voltage drop in the meter as well as the resistor drops. This is negligible at *Y* (0-50 mA scale) but is sufficient to affec the voltage reading at the 1000-ohm load (0-5 mA scale), although it does not affect measurements at *X* and *Y*. The dashed lines in the mA positions indicate that a jumper must be used when the milliammeter is in another part of the circuit.

Typical measurements are as follows:

Resistors	mA at X	V at Y	mA at Y	V at Z	mA at Z
R_1 and R_2 in use	39.3	8.2	12.6	2.45	2.18
R_1 only	39.0	8.2	12.6	2.6	0
R_2 only	38.5	8.7	0	2.6	2.3
None in use	27.2	8.7	0	2.8	0

Comment: If R_1 and R_2 together represent the working load on the voltage divider, the measured voltages at *Y* and *Z* show that the percent regulation at *Y* is

$$\frac{8.7 - 8.2}{8.2} \times 100 = 6.1\%$$

and at Z is

$$\frac{2.8 - 2.45}{2.8} \times 100 = 14.3\%.$$

These are rather good regulation figures for a divider, and can be explained on the basis that the larger load current flows through a comparatively low divider resistance (47 ohms) and that the load current through the larger divider resistance is small. If the same load voltages and currents were wanted from a 12-volt source, for instance, the 47-ohm resistor would have to be increased to about 100 ohms and the resulting regulation at Y would be 34%, over five times as great. The regulation at Z also would increase.

Note also that holding the voltage constant at X is equivalent to having a power source with zero internal resistance – that is, perfect regulation. If the source is not voltage-regulated its output voltage, too, will vary with load, as in Procedure (B), and this would have to be taken into account in the design of a complete system.

Experiment 3

PURPOSE: To demonstrate electromagnetic induction.

APPARATUS: Low-voltage variable dc power supply, multirange milliammeter, electronic voltmeter, telegraph key, and small iron-core transformer. The transformer used in the description below is a Chicago-Stancor PS-8415, a small 115-volt power transformer having 125-volt and 6.3-volt secondaries.

PROCEDURE: Connect the apparatus as shown in Fig. E3-1. (A single-pole switch or a push-button can be used in place of the telegraph key, but the latter is a little more convenient for the purpose.) With the dc output voltage at minimum, close the key and gradually raise the voltage until the current through the primary is 50 mA. Set the electronic voltmeter for + readings. Open the key and note whether the EVM pointer deflection is up or down. If it is up, reverse the EVM connections to the 125-volt secondary.

Now close the key and the pointer will kick up to about half scale on the 1-volt range. With the key still closed, switch the EVM to − and then open the key. The meter will again kick upward, although the deflection polarity actually is opposite to that on closing. Changing the meter ± switch avoids an off-scale deflection below zero.

Now shift the EVM to the 6.3-volt secondary and repeat the procedure. The deflection should be about 0.05 volt with this winding, with the same current in the primary.

Reconnect the EVM to the 125-volt winding. Still on the EVM 1-volt scale, unbalance the meter zero (with the control provided on the EVM) so that the pointer is at 0.2 volt with no voltage applied. Close the key and reduce the primary current to 25 mA. Rock the power-supply voltage control above and below the 25-mA setting so that the EVM reading varies about 0.1 volt above and below the 0.2-volt rest reading. If the control is rocked about four times a second the EVM reading will follow, showing increasing voltage with increasing power-supply voltage (which also increases the primary current), and decreasing EVM reading with decreasing power-supply voltage.

Comment: The experiment gives a qualitative indication of induction and mutual inductance. The voltage induced in the secondary, although read on a dc scale on the EVM, actually swings *both* positive and negative when the key is opened or closed, because a purely unidirectional current cannot be transferred through the magnetic field (this is shown by the fact that the pointer returns to zero when the primary current is not changing). However, one swing has higher amplitude than the other, and the mechanical inertia of the pointer gives the impression that a momentary dc voltage has been induced. An oscilloscope would show deflections in both directions, with one much larger than the other. Nevertheless, the *average* of both together over a sufficient time interval is zero. The predominant induced polarity on closing the key is opposite to that on opening it.

The lower induced-voltage reading from the 6.3-volt winding, which has considerably fewer turns than the 125-volt winding and thus much less inductance, is qualitative confirmation of the formula given earlier for mutual inductance. The primary is the same in both cases and the coefficient of coupling between windings is essentially constant and practically equal to 1 in an iron-core transformer. Thus the mutual inductance between the primary and the 125-volt winding is greater than the mutual between the primary and the 6.3-volt winding.

Rocking the primary voltage control gives a visual demonstration of how the induced polarity depends on whether the primary current is increasing or decreasing. In this experiment the EVM polarity is chosen so that the secondary voltage increases in the positive direction with an increase in primary current, but this is simply a question of choice. Reversing the EVM connections to the coil would cause the EVM reading to decrease with increasing primary current. However, the latter is a little confusing, for observation purposes, because one meter pointer swings up while the other is swinging down. By Lenz's Law the induced polarity (both windings are in the same field) will be such that the change in primary current will be opposed. The polarity *observed* at the secondary terminals depends on the relative winding directions of the primary and secondary coils.

There are induced voltages in the primary coil as well as in the secondaries, but it is not practicable to measure them because of the difficulty of separating them from the dc voltage applied to the primary. However, the *existence* of an induced voltage in a single coil can be demonstrated. Adjust the current in the primary to about 10 mA. As there will be less than 1 volt applied to the primary, add a 50-ohm resistor in series at *X* if necessary. Lay a finger lightly across the bare primary terminals. Closing the key will produce no effect, but on opening the key a distinct, although mild, shock will be felt, indicating

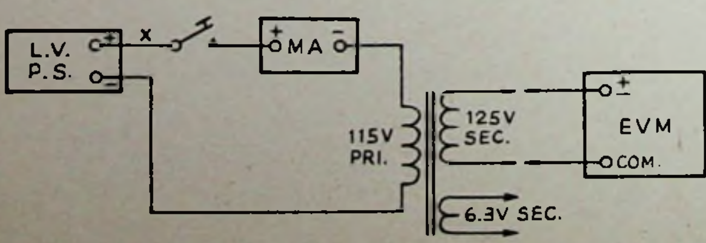


Fig. E3-1

that there is instantaneously a considerable voltage across the coil terminals.

If the primary current is increased a little the shock becomes more evident. At 50 mA it is distinctly unpleasant – probably enough so to discourage repetition of the experiment, and to inspire caution in future handling of iron-core inductances having direct current flowing in them! The intensity of the shock varies with different persons, depending principally on skin moisture. Accordingly, the current should be started at a low value and increased until the demonstration is satisfactory. The induced voltage on *closing* the circuit cannot exceed the supply voltage (if it did, the coil would be supplying energy to the power supply, which is impossible), but the only limitation on *opening* the circuit is the rapidity with which the current changes from its steady value to zero.

Electrical and Radio Units

Quantity	*Fundamental Unit*	*Auxiliary Unit*	*To Convert to Fundamental Unit-Multiply by*
Voltage	volt	kilovolt millivolt microvolt	10^{3} 10^{-3} 10^{-6}
Current	ampere	milliampere microampere	10^{-3} 10^{-6}
Power	watt	kilowatt milliwatt microwatt	10^{3} 10^{-3} 10^{-6}
Frequency	cycle per second	kilocycle megacycle Gigacycle	10^{3} 10^{6} 10^{9}
Resistance	ohm	kilohm* megohm**	10^{3} 10^{6}
Reactance Impedance	ohm ohm	Same as for resistance	
Inductance	henry	millihenry microhenry	10^{-3} 10^{-6}
Capacitance	farad	microfarad micromicrofarad or picofarad	10^{-6} 10^{-12}

*In circuit diagrams, the letter K written after the number is used to represent 1000; e.g., 15K indicates 15 kilohms or 15,000 ohms.

**Generally written "meg." in values given on circuit diagrams; e.g., 2.2 meg. = 2.2 megohms or 2,200,000 ohms.

Metric Multiplier Prefixes

Multiples and submultiples of fundamental units (e.g., ampere, farad, gram, meter, watt)) may be indicated by the following prefixes.

Prefix	*Abbreviation*	*Multiplier*
tera	T	10^{12}
giga	G	10^{9}
mega	M	10^{6}
kilo	k	10^{3}
hecto	h	10^{2}
deci	d	10^{-1}
centi	c	10^{-2}
milli	m	10^{-3}
micro	μ	10^{-6}
nano	n	10^{-9}
pico	p	10^{-12}

Resistor-Capacitor Color Code

Color	*Significant Figure*	*Decimal Multiplier*	*Tolerance (%)*	*Voltage Rating**
Black	0	1	–	–
Brown	1	10	1*	100
Red	2	100	2*	200
Orange	3	1,000	3*	300
Yellow	4	10,000	4*	400
Green	5	100,000	5*	500
Blue	6	1,000,000	6*	600
Violet	7	10,000,000	7*	700
Gray	8	100,000,000	8*	800
White	9	1,000,000,000	9*	900
Gold	–	0.1	5	1000
Silver	–	0.01	10	2000
No color	–		20	500

* Applies to capacitors only.

Experiment 4

PURPOSE: To demonstrate the time constants of capacitance-resistance circuits.

APPARATUS: Low-voltage adjustable power supply, electronic voltmeter, 50-μF electrolytic capacitor (50 V or higher rating), 0.1-μF paper capacitor, 1-megohm resistor.

PROCEDURE (A): Connect the apparatus as in Fig. E4-1(A) using a clip lead from the positive terminal of the power supply to C. Use the 50-μF capacitor at C. If the capacitor is new or has not been used for some time, first apply the highest voltage available from the power supply for a half hour or so to re-form the electrodes. Then reduce the voltage to 10 volts. Voltmeter readings are to be taken at 1-minute intervals immediately after disconnecting the positive lead from the power supply. Time these intervals as accurately as possible; use the second hand of a clock or watch to determine them. Also note the elapsed time when the EVM shows 3.7 volts (37% of the initial 10 volts).

Following is a typical set of data:

Elapsed Time, minutes	*EVM Reading, volts*
0	10.0
1	9.1
2	8.3
3	7.65
4	7.0
5	6.45
6	5.95
7	5.45
8	5.0
9	4.6
10	4.25
11	3.9
12	3.65
13	3.3
14	3.1
15	2.85
20	1.9
25	1.35

The last two readings were taken at 5-minute intervals because the change in voltage in this region is such that the pointer on a linear meter moves very slowly. The 3.7-volt reading occurred at 11 minutes, 45 seconds after "zero" time.

Plot your data on cross-section paper and draw a smooth curve through the points. A curve based on the above data is shown in Fig. E4-2. The time at which the 3.7-volt reading occurred also is shown.

Now substitute the 0.1-μF paper capacitor at C. The time constant for this capacitor is so much shorter that the same method of taking data cannot be used. Instead, read the voltage as accurately as possible (considering that the pointer will be moving fairly rapidly) at intervals of 1 and 2 seconds after disconnecting the charging source. The 1-second time ticks from WWV are useful for this. Approximate readings based on the average of several trials are

1 sec – 5.5 V
2 sec – 2.7 V

PROCEDURE (B): Use the circuit of Fig. E4-1(B) with a clip lead from R (1 megohm) so that the left-hand end of R can be connected to the power supply or to the common lead.

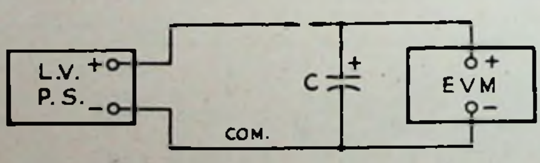
First, connect R to the common lead and then charge C to 10 volts directly from the power supply. Then disconnect the voltage source, and at the instant of breaking the contact start counting seconds until the EVM reads 3.7 volts. Repeat several times and average the readings.

Next, discharge the capacitor by short-circuiting its terminals for a few minutes. Capacitors tend to "recover" from a short-circuit and often will retain a small charge, so make sure that the EVM shows no reading when the short-circuit is removed.

Now set the source at 10 volts and charge C through R, counting seconds from the instant contact is made until the time when the EVM reads 5.75 volts (this is 63% of 9.15 volts, the final voltage across C with 1 megohm in series with the EVM). The time should be about the same as that obtained on measuring discharge.

Because capacitance tolerances on electrolytic capacitors are so wide (usually minus zero and plus 100% – really a guarantee that the capacitance will not be *less* than the rated value) the actual times may

(A)



(B)

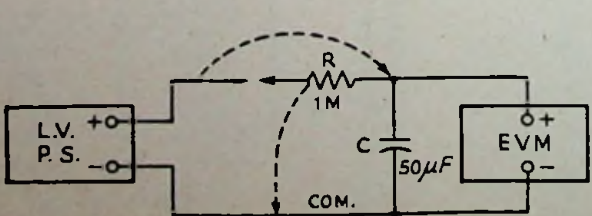


Fig. E4-1

vary from 50 seconds upward, depending on the particular capacitor used.

Comment: The time constants of inductive circuits are too short, with practicable components, to be measured with simple equipment, so this experiment is confined to CR circuits.

In Procedure (A) the entire resistance on discharge is ostensibly the 11-megohm resistance of the EVM; in reality, this resistance is shunted by a leakage path in the capacitor which allows a small additional current to drain off. The leakage can be expressed as an equivalent resistance, and in the case of electrolytic capacitors the leakage resistance decreases with an increase in applied voltage. It was well over 100 megohms at 10 volts in the capacitor (150-volt rating) used for getting the data shown, a fact that could be inferred by comparison of the capacitances calculated from the time constants obtained in this procedure and in Procedure (B).

The second part of Procedure (A) shows how much more rapidly the voltage changes when the time constant is short. Unfortunately, reliable measurements cannot be made with ordinary instruments in a case like this; the errors arise because the meter pointer is moving too rapidly for accurate reading, and because the inertia time of the pointer is of the same order of magnitude (with these CR values) as the time constant. The measurement accuracy therefore is poor, and the experiment is included mainly to bring out the fact that meter characteristics cannot always be ignored. Using the measured data and Fig. 5-4, Chapter 5, the capacitance is found to be 0.15 μF. This is too high by about 50%, since the rated capacitance tolerance is $\pm$10% on paper capacitors of this type. The principal reason for the discrepancy is meter inertia.

Procedure (B) shows voltage rise as well as voltage decay, and illustrates another possible pitfall

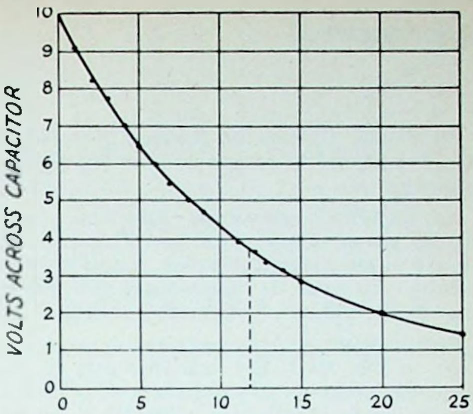


Fig. E4-2

in making measurements if meter characteristics are forgotten. On charging, the meter resistance becomes part of a voltage divider across the dc source, so the capacitor does not charge to 10 V when the source is set to 10 V. Also, on discharge the meter resistance is in parallel with the 1-megohm resistor, reducing the total to about 0.92 megohm. Thus the time constant on discharge can be expected to be slightly shorter than on charge. In the actual experiment the charge time was 67 seconds and the discharge time 65 seconds. The capacitance values calculated from these times (67 and 70 μF, respectively) are in fairly good agreement. They also agree quite well with the value obtained by Procedure (A), which is 64 μF if 11 megohms is used for R, confirming that the leakage resistance of the capacitor is high enough to be neglected.

Experiment 5

PURPOSE: To show the behavior of ac and dc when simultaneously present in linear and nonlinear circuits.

APPARATUS: Low-voltage adjustable regulated dc power supply; small power transformer having 6.3-volt winding; 330-ohm, 1-watt resistor; electronic voltmeter; dc milliammeter; spst switch; silicon diode, 100-volt, 50-mA rating or higher.

PROCEDURE (A): Connect the apparatus as shown in Fig. E5-1. For this first part of the experiment, short-circuit the diode, CR, or omit it from the circuit. With S open, adjust the dc voltage to 15 as read by the electronic voltmeter. Record the current as given by the milliammeter. Then close S, thereby adding a nominal 6.3 volts ac to the dc already present, and record the voltage and current again. Shift the EVM to ac (change probes) and record the ac voltage across the 330-ohm resistor. Repeat the procedure with 12, 9, 6, 3, and 0 volts dc (turn off the low-voltage power supply for the last, if necessary).

Typical data are as follows:

Without ac		*With ac*		
Dc volts	*Dc mA*	*Dc volts*	*Dc mA*	*Ac volts*
15	44.5	15	44.5	7.7
12	34.5	12	34.5	7.7
9	26.5	9	26.5	7.7
6	17.5	6	17.5	7.7
3	9.0	3	9.0	7.7
0	0	0	0	7.7

The observed value of ac voltage, 7.7, is fairly typical of the voltage output of a "6.3-volt" winding at low currents, with a small transformer. Also, the line voltage was higher than the nominal rating of the transformer primary (115 volts) when these data were taken.

PROCEDURE (B): Connect the diode into the circuit. Be sure to observe proper polarity; there will be no output voltage if the polarity is reversed. Follow Procedure (A), except that readings should be taken at 2-volt rather than at 3-volt intervals. Do not readjust the power supply output when ac is added.

A representative set of data is given below:

Without ac		*With ac*		
Dc volts	*Dc mA*	*Dc volts*	*Dc mA*	*Ac volts*
15	44.6	15	44.6	7.7
13	39.0	13	39.0	7.7
11	32.5	11	32.5	7.7
9	26.8	9.1	27.0	7.6
7	21.0	7.5	22.4	7.4
5	15.2	6.1	18.0	6.9
3	9.0	4.7	14.2	6.35
1	3.0	3.6	10.5	5.55
0	0	2.15	6.5	4.2

As a check on your data, apply Ohm's Law to corresponding dc voltage and current columns to determine the apparent resistance. The answer should be close to the rated resistance, 330 ohms, in every case. The data above are well within resistor and measurement tolerances.

Comment: Without the diode the circuit is linear, since the output circuit of the regulated power supply has only a negligible effect on the total circuit resistance. In this case the voltage-vs.-current characteristic of the whole circuit is essentially that of the 330-ohm resistor alone, because the resistance of the 6.3-volt winding of the transformer is likewise negligible. (The milliammeter adds a little resistance to the circuit, but the voltage is measured across the *resistor* rather than the resistor and meter in series. In any case, the meter is a simple resistance and therefore linear.)

The current-vs.-voltage characteristic is shown graphically in Fig. E5-2 by the straight line passing through zero. To the left of zero the current and voltage both reverse polarity.

In a linear circuit, ac and dc simultaneously present are independent of one another. This is shown by the fact that the ac voltage remains constant while the dc voltage is varied. If the ac voltage could be varied readily the converse also would be found to be true.

With the diode in the circuit, current is prevented from flowing whenever the total applied voltage is negative, as shown by the horizontal portion of the voltage-vs.-current characteristic to the left of zero. In this case the dc-only measurements are the same as in (A), but conditions change when the ac voltage is present. In the data, the rms voltage is 7.7, which means that the peak ac voltage is 7.7 × 1.4, or 10.8 volts. Therefore *at least* this much dc voltage must be present in order to preserve the linearity of the circuit. This is shown in the data by the fact that dc voltages from 15 to 11 duplicate the measurements made without the diode. At lower dc voltages the current is intermittently cut off by part of the negative swing of the ac voltage, with the result, as demonstrated by Fig. 6-5(B), Chapter 6, that the average current on the positive swing is greater than the average current on the negative swing. This causes a net

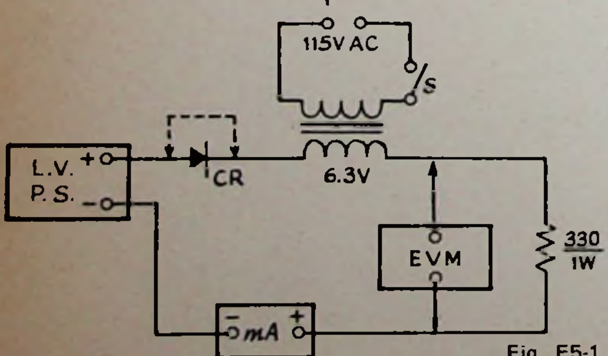


Fig. E5-1

increase in current and a consequent increase in the dc voltage across the resistor.

Simultaneously, the ac voltage begins to decrease, because the ac axis has to shift in such a way that the *average* voltage is the same in each direction. (The EVM, being capacitively coupled for ac, eliminates the dc component.) This, too, can be visualized from Fig. 6-5(B). Because part of the negative half cycle is clipped off, the peak-to-peak voltage is reduced and the axis shifts upward. As the EVM actually reads the positive peak (rather than true rms) its reading decreases, and the greater the ac overswing (which accompanies the reduction of dc voltage in the experiment) the more pronounced this effect becomes.

The measured values of ac voltage are affected to some extent by the voltage drop in the diode, but this is not important for the purposes of this experiment.

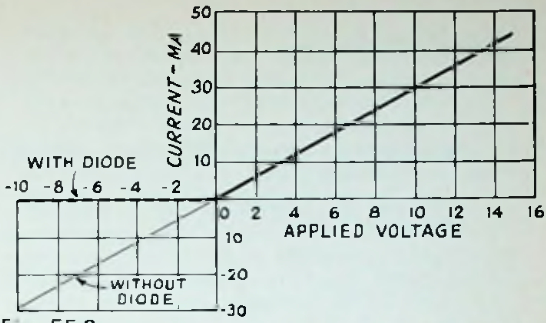


Fig. E5-2

Experiment 6

PURPOSE: To show the effect of combining resistance with reactance in an ac circuit.

APPARATUS: 6.3-volt (nominal) ac source, negligible current (6.3-volt filament transformer or the filament secondary of the small power transformer used in earlier experiments); electronic voltmeter; VOM; iron-core filter choke, approximately 15 H (Stancor C-1515 or equivalent); 0.5-μF paper capacitor, any voltage rating; 4700- and 6800-ohm resistors, 1/2 watt, 10% tolerance.

PROCEDURE (A): The circuit is given in Fig. E6-1(A), R_1 being 4700 ohms and R_2 6800 ohms. Measure the voltages indicated (E, E_{R1}, and E_{R2}) using the EVM. The VOM is used to monitor the input voltage to make sure that the voltage remains constant while E_{R1} and E_{R2} are measured; its use is not strictly necessary in this experiment but it is convenient.

Typical measured voltages are as follows:

E – 7.4 volts
E_{R1} – 3.0 volts
E_{R2} – 4.4 volts

From the measured voltages, calculate the current, $E/(R_1 + R_2)$, and the estimated voltage drops in R_1 and R_2 based on their nominal values. Compare with the measured voltages.

PROCEDURE (B): Substitute Fig. E6-1(B) for the section of E6-1(A) between the vertical dashed lines (inductor L replaces R_2), and measure the voltages as indicated. Representative measurements:

E – 7.4 volts
E_L – 5.4 volts
E_R – 4.1 volts

Calculate the current ($I = E_R/R$), the impedance Z_L of L ($Z_L = E_L/I$) and the total impedance ($Z = E/I$).

Calculated values:

I – 0.000873 A
Z_L – 6190 ohms
Z – 8470 ohms

PROCEDURE (C): Substitute Fig. E6-1(C) and measure the voltages indicated. Calculate I, Z_C, and Z. Representative measurements:

E – 7.4 volts
E_C – 5.5 volts
E_R – 4.75 volts

Calculated values:

I – 0.00101 A
Z_C – 5450 ohms
Z – 7330 ohms

PROCEDURE (D): Substitute Fig. E6-1(D) and measure the voltages indicated. Calculate I, Z_C, Z_L, Z_X ($Z_X = E_X/I$) and Z. Representative measurements:

E – 7.4 volts
E_C – 6.7 volts
E_L – 8.1 volts
E_R – 5.95 volts
E_X – 1.85 volts

Calculated values:

I – 0.001265 A
Z_C – 5300 ohms
Z_L – 6400 ohms
Z_X – 1460 ohms
Z – 5850 ohms

PROCEDURE (E): Substitute Fig. E6-1(E), measure the indicated voltages, and calculate I, Z_X, and Z. Representative measurements:

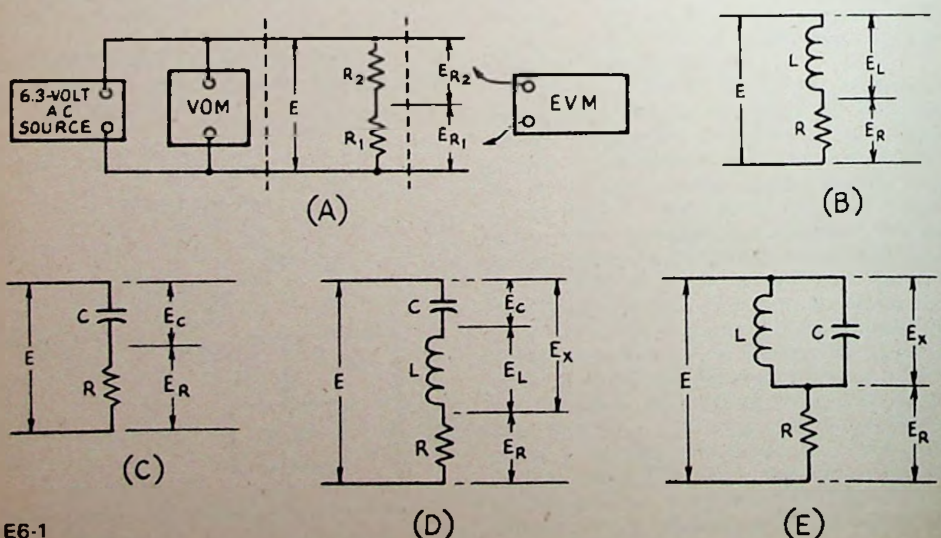


Fig. E6-1

E – 7.4 volts
E_X – 6.5 volts
E_R – 1.1 volts

Calculated values:

I – 0.000234 A
Z_X – 27800 ohms
Z – 31600 ohms

Comment: This experiment brings out the point that the presence of both resistance and reactance in an ac circuit leads to considerable departures from the simple Ohm's Law relationships of dc circuits (and the equally simple relationships of ac circuits which consist solely of pure resistance or pure reactance). In such mixtures of resistance and reactance the measured voltages in a series circuit do not add up numerically to the total voltage, nor do the currents add up numerically in a parallel circuit. The reason, of course, is that the voltages and currents in different components have different phases.

The equipment provided for these experiments does not include means for direct measurement of alternating currents. Consequently, the values of current must be determined indirectly – by measuring the voltage across a pure resistance and using Ohm's Law. This method cannot be used with inductors and capacitors, because these components usually have internal resistances which cannot be isolated for measurement. The resistances can be approximated, however, by using phasor diagrams, as described in Chapter 9.

The inductance and capacitance used in the present experiment approach, but do not equal, resonance at 60 Hz, the applied frequency. This choice emphasizes the peculiarities of behavior of circuits containing both kinds of reactance along with resistance.

Procedure (A) simply demonstrates that Ohm's Law does apply without modification when the circuit is purely resistive. (If the two voltages do not add to the proper total, check the EVM response against another ac voltmeter such as the one in the VOM.)

In the remaining procedures the measurements do not lead to values for the reactances of the inductor and capacitor because of their internal resistance. Rather, their impedances are measured; thus the symbol Z rather than X. From Procedures (B) and (C), Z_L is larger than Z_C, so it may be inferred that in the circuit containing both in series, Fig. E6-1(C), the current will lag the voltage in the total impedance, Z. In the parallel circuit, Fig. E6-1(C), this same relationship means that the current *leads* the voltage, as explained in Chapter 8. Note that Z_X, the impedance of the LC combination, is considerably smaller than either Z_L or Z_C in the series circuit (D) but is considerably larger than either in the parallel circuit (E).

Experiment 7

PURPOSE: To demonstrate the effect of phase on voltages and currents in circuits containing both reactance and resistance.

APPARATUS: Transformer furnishing 6.3 volts, nominal, 60 Hz (current is negligible); electronic voltmeter; VOM; iron-core filter choke, approximately 15 H inductance, 400 ohms dc resistance (Stancor C-1515 or equivalent); 0.5-μF paper capacitor; resistors: 6800, 4700, 1000, and 560 (2) ohms, 1/2 watt, 10% tolerance or better.

PROCEDURE (A): Connect the equipment as shown in Fig. E7-1, using (successively) 6800, 4700, and 100 ohms at R. Make all measurements with the EVM, but leave the VOM connected across the 6.3-volt source to monitor the applied voltage. Line-voltage changes of a few percent are common on commercial power lines, and since small changes in voltage will have a considerable effect on the data, make each group of measurements at constant input voltage even if this requires waiting for the opportune moment.

Using 6800 ohms at R, (1) measure E, E_L, and E_R; (2) Calculate the current from E_R and R; (3) Calculate the circuit impedance ($Z = E/I$); (4) Calculate the impedance of the inductor alone ($Z_L = E_L/I$); (5) Repeat with 4700 ohms and 1000 ohms.

Typical measurements are as follows:

R	*E*	E_R	E_L
6800 ohms	7.45 volts	5.2 volts	4.48 volts
4700 "	7.45 "	4.2 "	5.5 "
1000 "	7.45 "	0.95 "	7.2 "

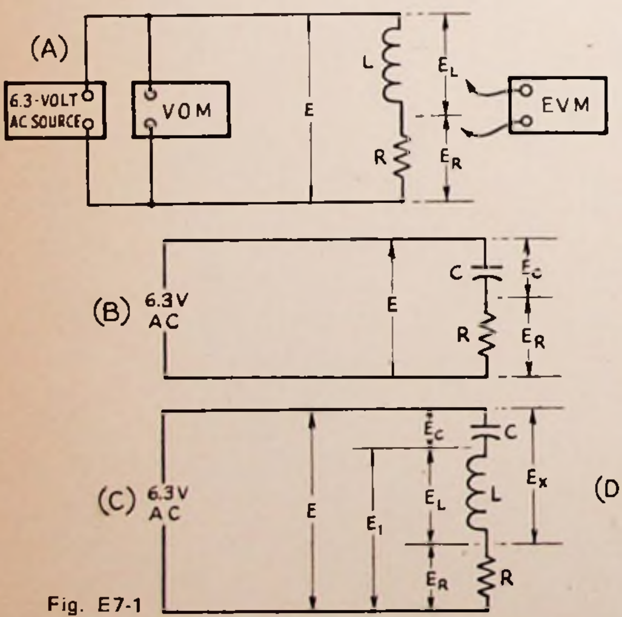


Fig. E7-1

From these figures, the calculated results are

R	*I*	*Z*	Z_L
6800 ohms	0.000765 A	9750 ohms	5850 ohms
4700 "	0.000895 "	8330 "	6210 "
1000 "	0.00095 "	7850 "	7580 "

PROCEDURE (B): The circuit is given in Fig. E7-1(B). Equipment is the same; although the meters in Fig. E7-1(A) are not repeated in the subsequent parts of the figure, they are used in exactly the same way as in (A).

Follow the same procedure as in (A), using 6800, 4700, and 1000 ohms at R, and measuring the voltage E_C across the capacitor rather than the inductor. Typical measurements follow:

R	*E*	E_R	E_C
6800 ohms	7.4 volts	5.83 volts	4.44 volts
4700 "	7.4 "	4.73 "	5.5 "
1000 "	7.4 "	1.13 "	7.25 "

Results based on these data are:

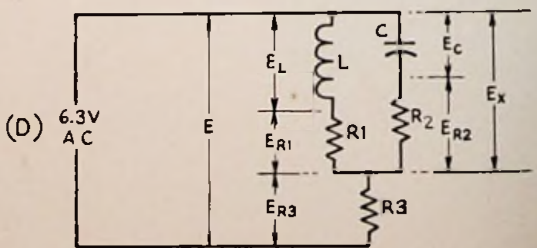
R	*I*	*Z*	Z_C
6800 ohms	0.000858 A	8630 ohms	5170 ohms
4700 "	0.001005 "	7360 "	5480 "
1000 "	0.00113 "	6550 "	6410 "

PROCEDURE (C): Fig. E7-1(C) gives the circuit, which has L, C, and R in series. The L and C values are the same as in (A) and (B), respectively. Procedure is similar to that already described, extended to include determination of the combined impedance of L and C, as well as separate determinations of the individual impedances. To this end the voltages across L and C in series (E_X) and L and R in series (E_1) are also measured. One set of measurements, using 4700 ohms at R, will suffice.

Typical data:

E – 7.55 volts
E_R – 6.1 "
E_L – 7.9 "
E_C – 6.62 "
E_X – 2.05 "
E_1 – 11.0 "

from which the following calculated values are obtained:



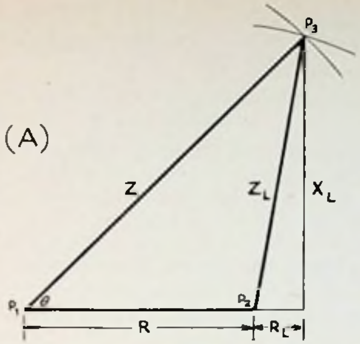
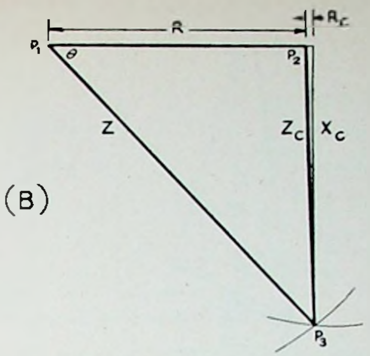


Fig. E7-2



I	– 0.00129 A
Z_L	– 6350 ohms
Z_C	– 5280 "
Z_X	– 1470 "
Z	– 5850 "
Z_1	– 8550 "

PROCEDURE (D): Fig. E7-1(D) shows the voltages to be measured in a series-parallel circuit arrangement. Resistors R_1 and R_2 (560 ohms) should be matched as closely as possible; pick out two that read the same on an ohmmeter. Although the actual values are not especially critical, they should be *known* as accurately as possible. The value of R_3, nominally 1000 ohms, also should be carefully measured. The accuracy of the results of the measurements depends fairly critically on the accuracy with which these resistances are known, so they should be checked on the ohmmeter scales of both the EVM and VOM; if there is a discrepancy of more than a few percent, determine the resistance by measuring the current and voltage using a dc source.

When the measurements are complete, calculate the current I_L flowing through L and R_1, the current I_C flowing through C and R_2, and the total current I flowing into and out of the entire circuit.

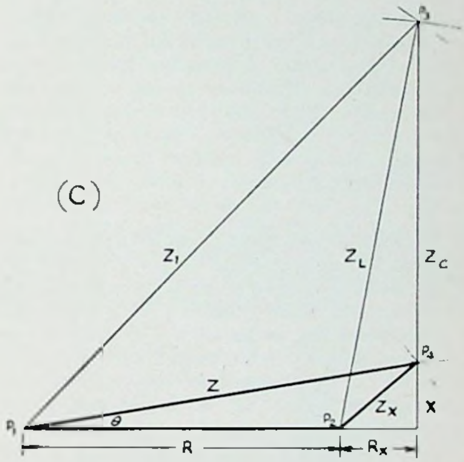
Representative measurements are given below:

E	– 7.4 volts
E_L	– 6.86 "
E_C	– 6.95 "
E_X	– 7.0 "
E_{R1}	– 0.58 "
E_{R2}	– 0.68 "
E_{R3}	– 0.44 "

These data lead to the following figures for the impedances of the inductive branch ($Z_L = E_X/I_L$), the capacitive branch ($Z_C = E_X/I_C$), both in parallel ($Z_X = E_X/I$), and the total impedance ($Z = E/I$):

Z_L	– 6750 ohms
Z_C	– 5760 "
Z_X	– 15,900 "
Z	– 16,800 "

Comment: The measurements in this experiment are used to demonstrate how phasor diagrams

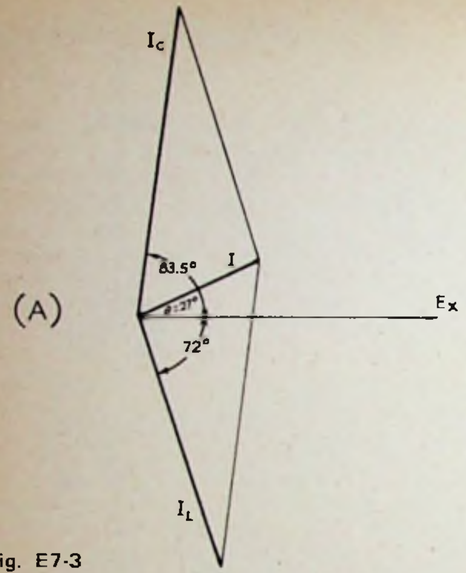


can be employed for determining quantities that are not directly measurable.

Fig. E7-2(A) is the diagram corresponding to Fig. E7-1(A). It is drawn as an impedance diagram, and is based on the impedances calculated from the measured data. To construct it, the resistance R (4700 ohms in this case) is drawn to a suitable scale, then an arc of radius Z is drawn from one end, P_1, the origin of the phasor. Next, another arc of radius Z_L is drawn from the other end, P_2, to intersect the Z arc. P_3, the point where the arcs intersect, determines the third point of the impedance triangle with sides Z, Z_L, and R. The angle θ is the phase angle between R and Z; that is, θ is the phase angle of the overall circuit.

To determine the resistance and reactance of the inductor, the base line is extended far enough so that a perpendicular can be drawn to it from P_3. The scaled lengths of X_L and R_L then give the reactance and resistance, respectively, of the inductor. In Fig. E7-2(A) X_L measures 6050 ohms while R_L is 1050 ohms. Of R_L, 400 ohms is the dc resistance of the inductor and the remainder is the resistance equivalent of the ac loss, which is principally in the iron core.

Similar diagrams can be drawn for the other two sets of data obtained from Fig. E7-1(A), and it is recommended that they be constructed. How-



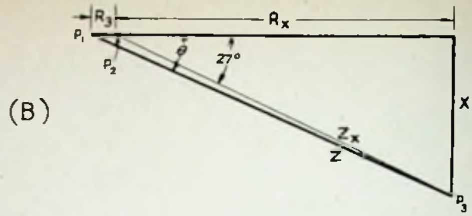


Fig. E7-3

ever, it is not to be expected that all such drawings will agree exactly, because of measurement and drafting inaccuracies. In general, the accuracy will be highest when no very small or very large angles are used; when the angles are about equal the arc intersection points can be located more definitely. The angles can be governed to some extent by the choice of the value for R in relation to the characteristics of the inductor.

Fig. E7-2(B) is a diagram, also for R = 4700 ohms, corresponding to the data for Fig. E7-1(B). The losses in a paper capacitor usually will be so small that Z_C and X_C coincide, but the capacitor used in securing the data was old and exhibited enough resistance to be fairly significant. The scaled resistance is 210 ohms and the reactance 5450 ohms.

The impedance diagram for the circuit of Fig. E7-1(C) is formed by Z, R, and Z_X in Fig. E7-2(C). The composition of Z_X is found in the same way as before: the scaled values are 1160 ohms for R_X and 1000 ohms (inductive) for X. R_X is the sum of the resistances of the inductor and capacitor, and X is the *algebraic* sum of the inductive and capacitive reactances.

By following the same procedure using Z_1 (based on E_1) and Z_L (based on E_L) the resistance and reactance of the inductor L can be found; similarly with capacitor C. Point P_4 is determined by swinging an arc having a radius Z_C from P_3 to intersect an arc of radius Z_X centered on P_2. As the reactance of C is negative, it is drawn to *subtract* from the reactance of L.

The method of constructing the diagram (Fig. E7-3) for the series-parallel circuit of Fig. E7-1(D) is basically similar. However, E and E_X are very nearly equal, and E_{R3} is small compared to either; this makes it difficult to use these voltages or the corresponding impedances for determining the phase angle with any accuracy. Therefore, as a first step it is better to find the angle between current and voltage in the LR_1 circuit and do the same thing for the CR_2 circuit. The methods are those used in Figs. E7-2(A) and E7-2(B). The graphical construction leads to a phase angle of 72 deg between E_X and I_L, and an angle of 83.5 deg between E_X and I_C. Then by drawing the currents to scale, using E_X as the reference, they may be added vectorially to find the total current, I, as shown in Fig. E7-3(A). The value of I scaled from this diagram is 0.00045 A, which is in good agreement with the value calculated from E_R/R_3. The final step is to draw Z_X to the proper scale and angle from P_2, Fig. E7-3(B), and then draw Z from P_1 to P_3, the end of the Z_X line. The values of effective resistance and reactance may be found by dropping a perpendicular from P_3 to the reference line, as shown.

Radio-frequency circuits used in communication frequently work at or close to resonance at some particular frequency, and at this frequency the overall phase angle between voltage and current is small. In such cases there are approximate methods for solving the circuits without the necessity for determining phase angles and resorting to phasor diagrams. However, there is no better way of gaining insight into the actual operation of ac circuits than to become acquainted with relationships such as are discussed in this experiment and Chapter 9. In particular, Fig. E7-3 demonstrates how, in a parallel circuit having opposing reactances, it comes about that the total impedance Z and its resistive component $R_3 + R_X$ increase while the reactive component X decreases, as compared with the values making up the individual parts of the circuit. Although it is not possible to vary the frequency in Exp. 7 it is not hard to visualize that as the currents I_L and I_C become more nearly equal (circuit approaches resonance) X in Fig. E7-3(B) becomes smaller, the phase angle approaches zero, and Z becomes not only purely resistive but much larger than the individual reactances or impedances in the circuit.

Note: The series circuits in this experiment are identical with those of Exp. 6, and the measurements and calculations in the two experiments may be compared directly as a check on the work. The data presented in the two experiments agree within the measurement tolerances. The parallel circuits in the two experiments have different resistances, however, and the voltages and impedances cannot be expected to check. Note that the presence of resistance in each leg of the parallel circuit in Exp. 7 leads to lower values of both Z and Z_X than in Exp. 6.

Experiment 8

PURPOSE: To measure forward and reverse characteristics of typical diodes.

APPARATUS: Low-voltage adjustable power supply; electronic voltmeter; VOM (used for current measurement); 100-ohm 1/2-watt resistor; other 1/2-watt resistors as required for preventing off-scale current readings, based on 10 volts input to the circuit of Fig. E8-1 (e.g., approximately 220K for 50-μA scale, 2200 ohms for 5-mA scale, 220 ohms for 50-mA scale, etc.); diodes as described below.

PROCEDURE (A): Connect the apparatus as shown in Fig. E8-1, using the proper value of resistance at R for confining current to the meter scale in use. For CR, use a silicon diode of the type used in small power supplies; one rated at 500 mA dc and a peak-inverse voltage of 200 or more will be suitable. Meter ranges of 0-50 and 0-500 mA (or equivalent scales covering about the same total range) will be satisfactory.

Choose values of current at intervals about as shown in the data below, and measure the voltage across the diode when the input voltage is set to give the selected current values.

Following is a typical set of data:

Current, mA	*Volts*
0.1	0.53
0.2	0.56
0.5	0.59
1.0	0.62
2	0.65
3	0.67
5	0.69
10	0.74
20	0.75
30	0.77
40	0.78
50	0.8
100	0.86
200	0.95

Next, reverse the connections to the diode and measure the current and voltage at intervals of a few volts up to the maximum voltage available. Typical data are:

Volts	*Current, μA*
1	0
2	0
5	0.3
10	0.6
15	1.0
20	1.4
25	2.0

Note that only the low microampere scale is used for this set of measurements.

Plot the data on cross-section paper and draw a smooth curve as closely as possible to the points. Such a curve for the data in the above tabulation is given in Fig. E8-2. Only the forward voltage and current are shown. The reverse data would not show significant information because of the limited voltage range. The silicon diode used for this experiment has a PIV rating of 400 volts, so a voltage in excess of 400 would be required for reaching the avalanche point.

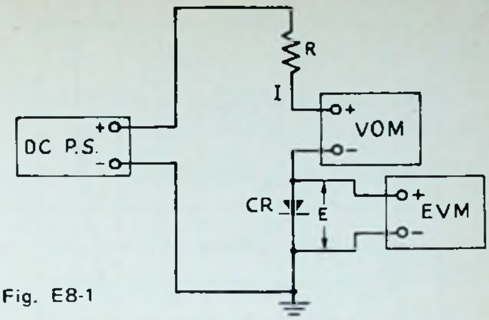


Fig. E8-1

PROCEDURE (B): The circuit is the same as for Procedure (A), but with a germanium point-contact diode substituted for the silicon rectifier used in (A). Current and voltage readings should be started at the lowest practicable values. At very low voltages it may be difficult to get consistent voltage readings because of stray currents induced on the wiring from ac power-frequency fields. The earth connection shown in Fig. E8-1 will help reduce such effects. In taking the data listed below, voltage readings for currents below 3 μA would not "repeat" in different measurements and so are not included. Typical data are as follows:

Current	*Volts*
3 μA	0.05
5	0.06
10	0.08
20	0.1
30	0.115
40	0.125
50	0.135
0.1 mA	0.16
0.2	0.185
0.5	0.22
1.0	0.24
2	0.27
3	0.29
5	0.32
10	0.36
25	0.42

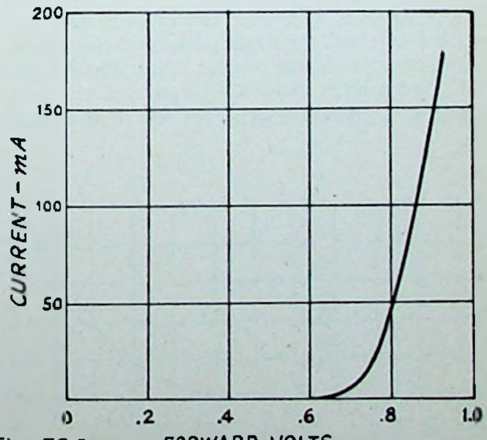


Fig. E8-2

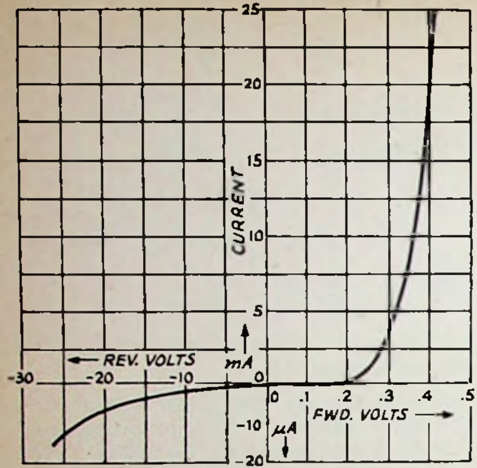


Fig. E8-3

Next, reverse the diode connections and measure the reverse current over the available voltage range. Typical data:

Volts	*Current, μA*
1	0.75
2	0.9
5	1.3
10	2.5
15	4.5
20	7.2
25	12.7

Plot the data on cross-section paper as in Procedure (A), but include the reverse voltage/current plot. Fig. E8-3 is a plot of the above data. Note the change of scale of both current and voltage when reverse current and voltage are plotted. The current scale (mA) does not permit showing the shape of the curve for very small currents, and these may be plotted on an expanded current scale as in Fig. E8-4. In Fig. E8-4 the scales for forward and reverse are the same.

PROCEDURE (C): The circuit is as given in Fig. E8-1 but with the diode polarities reversed for measurement of reverse current. Use a Zener diode at CR. One rated at about 6 volts and 1/2 to 1 watt is suitable. A resistance of about 100 ohms should be used at R to limit the current through the diode to the safe value based on the voltage and power rating. Measure the reverse voltage and current, taking equal intervals of voltage up to the beginning of avalanche, and thereafter reducing the voltage intervals so that the behavior of the reverse current can be observed.

Typical data are as follows:

Volts	*Current*
1	0 μA
2	0.5
3	3.3
4	22
5	0.2 mA
5.2	0.32
5.4	0.6
5.6	1.1
5.8	2.7
5.85	4.1
5.9	8.2
6.0	50

Plot the results as shown in Fig. E8-5. Since the current scale in the graph is in milliamperes, currents at applied voltages below 5 V are too small to show, and the voltage values below this may be omitted.

Comment: Comparison of the data obtained in Procedures (A) and (B) clearly brings out the differences between silicon and germanium semiconductor diodes. The curves of Figs. E8-2 and E8-3 make the forward voltages at which appreciable current begins to flow a little more obvious than the tabulated data does, because of the overall perspective of current flow that the graphs offer. The forward current rises rapidly, although not linearly, once the cutoff point is passed.

The two sets of data also emphasize the difference between the two types with respect to reverse currents. The reverse current for the silicon diode was not shown in Fig. E8-2 because it was very small (see data) for this diode, which had a PIV rating of 400 V, at the voltages available (up to about 25 V) for this test. The PIV rating for the germanium diode would be of the order of 50 V, however, and the reverse current becomes appreciable with about 20 volts applied in the reverse direction. It also increases from zero at a relatively high rate, which is typical of germanium compared with silicon.

Germanium diodes of the type tested are frequently used for rectification of small signals, and Fig. E8-4 shows how the diode would respond to signal voltages that are smaller than the nominal cutoff voltage. For equal signal-voltage swings either side of the zero voltage axis, but less than about 0.15 volt, the reverse current is negligible, indicating good rectifier performance.

Procedure (C) measures the avalanche characteristics of a Zener diode, as representative of what happens at the avalanche point in ordinary diodes. The Zener diode was chosen for this because its avalanche voltage was within the range of the low-voltage power supply. Zener diodes can be manufactured to have a low-voltage avalanche;

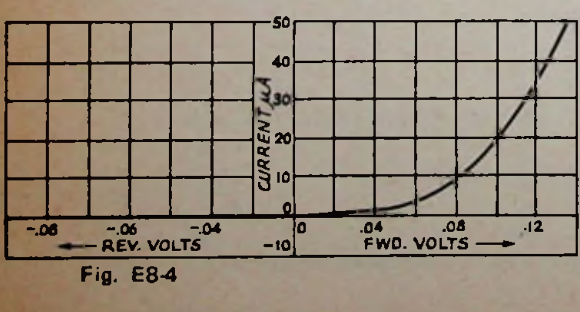


Fig. E8-4

ordinary diodes have considerably higher avalanche voltages. The diode used was rated nominally at about 6 V, and as shown by Fig. E8-5 the "knee" of the curve is at slightly over 5.8 volts. From that point on the current increases very rapidly for a small change in reverse voltage. If the voltage scale is multiplied by 100, such a curve might be representative of a rectifier having a PIV rating of 500 V, although it no doubt would differ in details such as the shape of the knee and the exact voltage at which avalanche would occur. The shape will, in fact, be different for different types of diodes, although the general characteristics would remain the same.

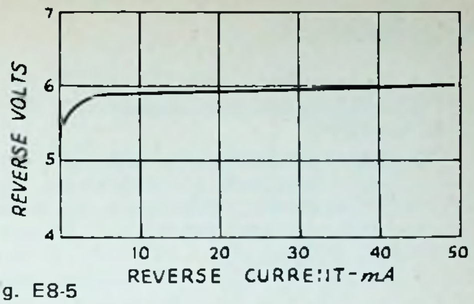


Fig. E8-5

In all three sets of measurements the current, and thus the power dissipated in the diodes, was kept well below ratings so that temperature effects would be minimized. To the same end, the test voltage should be applied only long enough to obtain satisfactory readings. If the device is operated continuously at ratings, the measured values will differ noticeably from those obtained at lower temperatures. This will be particularly so in the case of the germanium diode, but the difference is also observable with the silicon devices.

Experiment 9

PURPOSE: To measure characteristics of a bipolar transistor.

APPARATUS: Low-voltage adjustable dc power supply; VOM, preferably with 50-mA current scale (or equivalent milliammeter); electronic voltmeter; 150- and 1000-ohm 1/2-watt resistors; 100,000-ohm potentiometer; 1.5-volt dry cell (size D) and holder; audio-type npn transistor rated nominally at 500 mW at room temperature (25 deg C); transistor socket.

PROCEDURE (A): Connect the components as shown in Fig. E9-1(A). Base current is measured by measuring the voltage drop across the 1000-ohm resistor, using the lowest available EVM voltage scale; with 1000 ohms, 100 μA = 0.1 V, etc. (A 0-500 μA or 0-1 mA current-reading instrument can be substituted for the 1000-ohm resistor, if such an instrument is available, for greater convenience in making and interpreting base-current readings.) Use the 0-50 mA scale of the VOM, if such scale is provided in the instrument; otherwise, use the scale or scales that will give the equivalent range with as much readability as possible.

(A)

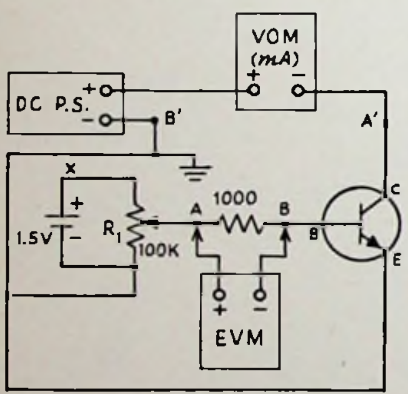
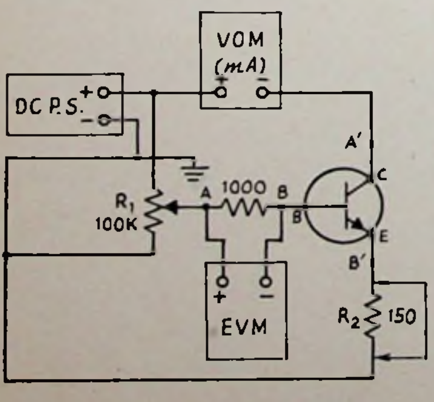


Fig. E9-1

(B)



It is advisable to connect an earth ground to the common point, as indicated in the circuit. The ground helps eliminate the effects of stray ac fields which tend to cause false EVM readings at very low voltages (below about 0.3 volt).

Temporarily connect the EVM to measure the power-supply output voltage (A' and B') and set the power supply for 15 volts with zero bias on the base. As the base current must be cut off frequently, it is convenient to use an easily removable clip connection at X.

With the power supply voltage set at 15, move the EVM connections to A-B, set the movable arm of R_1 to the ground end, and connect the bias battery. Now move the arm to increase the positive bias on the base until the collector current, I_C, is about 30 mA. Disconnect the bias and allow the transistor to cool for a few minutes. Then reconnect the battery and *immediately* note I_C and the base current, I_B. If I_C is not approximately 30 mA at the instant of connecting the bias, try another setting of R_1 and then allow the transistor to cool once more. The object is to have the initial current close to 30 mA when the transistor is at room temperature. The initial adjustment is somewhat critical, since the transistor will heat rapidly at this collector dissipation (about 500 mW).

Once the proper setting of R_1 has been found, and after thorough cooling of the transistor, connect the bias and note both I_C and I_B at frequent timed intervals until the collector current approaches 50 mA, when the bias should be shut off in the interests of preventing damage to the transistor. The current changes are most rapid at the beginning, and it is advisable to make measurements at 15-second intervals. Typical readings for a transistor of the type specified are given in the table below, and are shown graphically in Fig. E9-2 by the curves marked "Without Sink."

Time		I_C, *mA*	I_B, μA
0		29	160
	15 sec	32	170
	30 "	35.5	175
	45 "	39	180
1 min		40.5	185
	15 sec	43.5	188
	30 "	45.5	190
	45 "	47	190
2 min		48	192

As the meter pointers were in motion throughout the test that gave the above figures, third-place figures cannot be taken literally; they should be interpreted to mean that the reading was definitely higher than the associated second-place figure. After the two-minute test period the transistor case temperature was well above the boiling point of water, as attested by touching a moistened finger to the case.

As the figures above indicate, the temperature rise is far too great and too rapid to permit taking data for characteristic curves. It is therefore necessary to use a heat sink. If a suitable manufactured sink cannot be obtained readily, one can easily be made using a strip of metal obtained from an ordinary tin can. The dimensions given in Fig. E9-3 are satisfactory for the purpose. The end can be formed around a round bar of about the same diameter as the transistor case (an ordinary drill bit can be used), final adjustment for a snug fit being made by slight bending with the fingers. Such a heat sink was used in obtaining the following data:

Time	I_C, *mA*	I_B, μA
0	29	160
1 min	33	168
2 "	35	172
3 "	35.8	174
5 "	36.3	175
10 "	36.5	175

The test was discontinued at this point as the temperature had obviously stabilized. The corresponding curves are shown in Fig. E9-2 marked "With Sink."

Comment: These measurements show the importance of temperature rise in the operation of bipolar transistors, and indicate the necessity for taking steps to counteract it. As shown by Fig. E9-2, a simple heat sink will do a great deal toward holding the temperature to a safe level; with this sink, the heat was just sufficient to be noticeable when a finger was touched to it where it fitted over the transistor case. A sink of the same dimensions, but made of copper or aluminum, should dissipate the heat even more rapidly because these metals have considerably greater heat conductivity than the mild steel used in "tin" cans.

Although the sink used would be adequate for operating the transistor continuously with a collector input of 36 mA at 15 V or somewhat over 500 mW, it should be observed that it does not entirely solve the problem of making measurements over a wide range of currents and voltages, since for this purpose the transistor internal temperature must be constant regardless of variations in dissipation during the measurements. However, it helps to make the current drift smaller than it would be if the sink were not present. If the

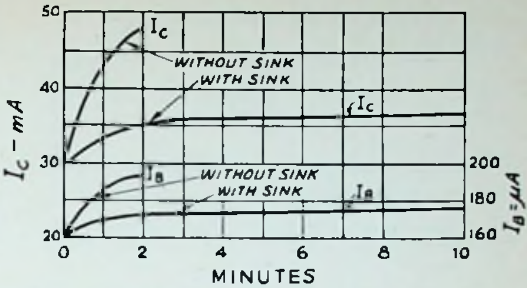


Fig. E9-2

transistor has a metal case, the case is probably connected to the collector internally, so make sure that the sink does not touch any part of the circuit when in use.

PROCEDURE (B): The circuit for this set of measurements is the same as in Procedure (A). The object is to obtain data for constructing the collector family; i.e., to measure the collector current, I_C, for various values of collector-to-emitter voltage, V_{CE}, at constant base current. Repeat the measurements for different selected values of base current. To stay within safe dissipation limits, the collector current should not be allowed to exceed 50 mA. The collector-voltage intervals should be small at low values of V_{CE} because I_C changes rapidly in this region. Also, at the higher values of collector voltage and current, readings must be made as rapidly as possible because the current drifts, even with the heat sink, as the transistor temperature rises. The most satisfactory procedure is to set the base current to zero by means of R_1, then adjust the power-supply voltage to the desired value, after which the base current should be increased to the value used in a run, I_C read as quickly as possible, and the base current then immediately returned to zero. After a short period to allow the transistor to return to room temperature, V_{CE} can be set to the next interval and the procedure repeated.

The following set of data is representative for the type of transistor specified:

Collector current in mA for base currents taken at 50-μA intervals

V_{CE}	*50*	*100*	*150*	*200*	*250*	*300*	*350*	μA (I_B)
15	7	17	27	37	48			*mA* (I_C)
10	6.5	16	25.3	35	44.5			
7						49		
5	6.2	15	24	31.5	40.3	46		
3	6	15	23	30.5	39	44.5		
2	5.95	14.8	23	30.1	38	43.5	50	
1	5.9	14.7	22.7	29.8	37	42.5	48.5	
0.8	5.85	14.6	22.2	29.4	36.5	42.0	48	
0.6	5.8	14.5	22.1	29.0	35.5	38.3	40.2	
0.5	5.75	14.4	21.9	28.0	31.5	32.2	33.5	
0.4	4.95	14.1	20.1	24.0	25.3	25	25.5	
0.3	3.8	12.1	15.2	17.0	17.5	18	19.8	
0.2	2.2	7.3	10.2	10.9	11.5	11.2	12.0	
0.1	1.0	2.1	3.8	3.0	4.0	4.0	4.8	

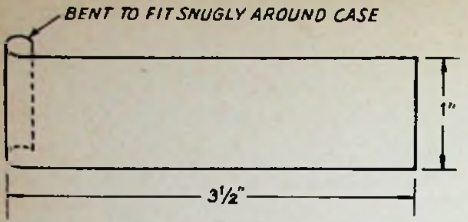


Fig. E9-3

Comment: Fig. E9-4 is a plot of the data in the above table. Note that in this transistor the curves are nearly horizontal at collector voltages above 1 V, indicating high internal collector resistance. Also, a change in base current of 50 μA causes approximately-equal changes in collector current. This indicates that the common-emitter current gain, h_{FE} or β, does not vary a great deal with either collector voltage or collector current. If operation is confined to the currents and voltages shown by these curves, a suitable operating point would be in the region of V_{CE} = 8 volts and I_C = 25 mA, for which a base current, I_B, of 150 μA is required.

At this operating point a base-current change of 50 μA causes an average change in I_C of 8 mA. β is therefore equal to 8×10^{-3} divided by 50×10^{-6}, or 160. The particular transistor used in obtaining the above data was rated to have a value of $h_{FE}(\beta)$ between 100 and 200, so the value obtained by measurement is well within the transistor specifications. Other transistors having similar dissipation ratings and designed for the same applications would of course have different characteristics; the exact type used for this experiment does not matter particularly since measurement will show what the characteristics are. A transistor having about the same dissipation rating is desirable, however, because the currents fall in a range which can be measured with reasonable accuracy.

The same data can be used for plotting curves of collector current vs. base current for various fixed values of collector voltage. However, a transfer characteristic – collector current vs. base voltage – is not readily obtainable because it is difficult to read small changes in base voltage with enough accuracy.

PROCEDURE (C): The circuit for this set of measurements, shown in Fig. E9-1(B), is the same as for Procedures (A) and (B) except for the addition of a 150-ohm resistor, R_2, in the emitter lead, with provision for eliminating it from the circuit by short-circuiting (indicated) or otherwise. Also, point *B'* is at the emitter terminal of the transistor instead of being at common or ground. This is so the actual voltage between the collector and emitter, V_{CE}, can be measured, rather than the power-supply output voltage.

The initial adjustment, to set the operating conditions at V_{CE} = 10 V and I_C = 25 mA, is critical. It is necessary that these two values be established with the transistor cool – i.e., before any internal warming occurs because of collector-current flow. This means that after each adjustment the transistor must be turned off, allowed to cool, and then turned on again when back to room temperature. If V_{CE} and I_C do not immediately show the correct values, a new set of adjustments must be tried and the process repeated, until cold turn-on shows V_{CE} = 10 V and I_C = 25 mA. The heat sink should be on the transistor when this adjustment is made.

Two separate sets of adjustments are required, one for the case where R_2 is shorted out and one with R_2 in the circuit. Aside from the heating problem just mentioned, the former presents no difficulties since the power-supply output voltage can be set directly to 10 V and left alone. R_1 then sets the bias current. This adjustment is less critical if the 1.5-volt cell is used for bias, as in Fig. E9-1(A). However, when R_2 is in the circuit, the bias must come from the power supply, and the power-supply voltage and R_1 must *both* be adjusted for the required conditions. These two adjustments interlock, so setting up the initial voltage and current is likely to become somewhat tedious, particularly in view of the necessity for keeping the transistor at room temperature. Whether or not R_2 is in the circuit, the initial adjustments should be left alone during a run.

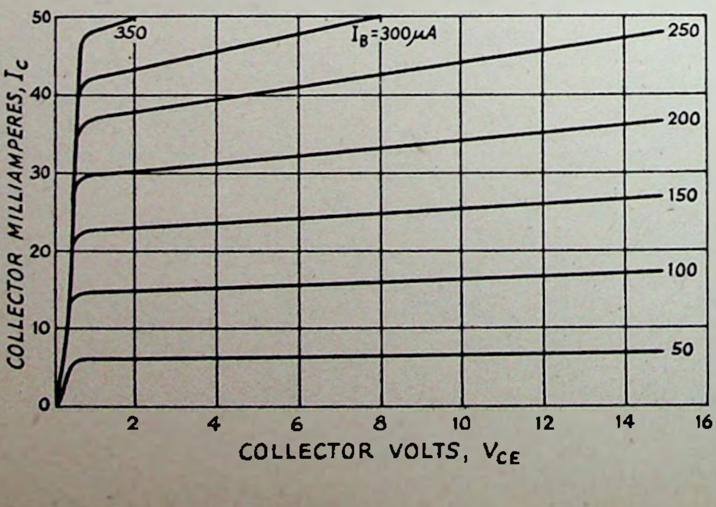


Fig. E9-4

Readings should be taken of collector current as it varies with time, for four different conditions: (A) without the heat sink and with R_2 shorted out; (B) with the heat sink, but with R_2 shorted out; (C) without the heat sink but with R_2 in the circuit; (D) with both the heat sink and R_2 in use. Also measure the base current at the beginning and end of each run. Except for these two measurements, the EVM may be left connected to *A'* and *B'* throughout a run, serving as a monitor for V_{CE}.

A representative set of readings is as follows:

Time	*A*	*B*	*C*	*D*	
0	25	25	25	25	mA
15 sec	27.3	26	25.6	25.4	
30 "	29.6	26.8	26	25.7	
45 "	31.1	27	26.2	25.9	
1 min	32.8	27.3	26.5	26	
30 sec	34.7	28	27	26	
2 min	36	28.1	27.2	26.1	
30 sec	36.5	28.2	27.3	26.2	
3 min	37.2	28.3	27.3	26.2	
4 "	38	28.3	28	26.2	
5 "	38.2	28.3	28.2	26.2	

Base currents as measured at the beginning and end:

	A	*B*	*C*	*D*	
Start	160	160	145	155	μA
Finish	180	165	145	155	

Comment: The variation of collector current with time for the four cases given in the above tabulation is shown in Fig. E9-5. This experiment brings out the advantage of the customarily used biasing circuit in reducing the effects of temperature rise in the transistor. Note that in this particular case there is very little difference between curves *B* and *C*, indicating that the bias circuit is equally as effective as the heat sink in stabilizing the circuit when either is used alone.

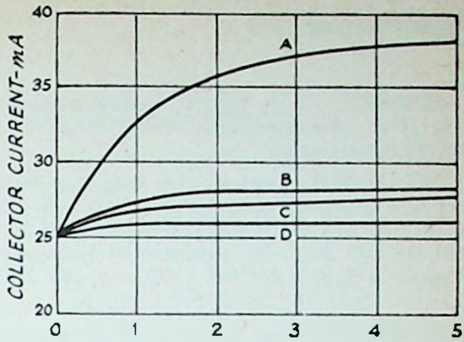


Fig. E9-5

(This would not necessarily be true of other transistors, heat sinks, and bias-circuit values.) Best of all, of course, is a combination of both. The bias circuit does not directly reduce the operating temperature, as the heat sink does, but simply *compensates* for the rise in temperature by reducing the collector current through a reduction in base current. In this case the transistor case was uncomfortably hot to the touch at the end of the 5-min run under condition *C*, while with the heat sink, condition *B*, the rise in temperature was just perceptible, on touching the sink over the transistor case.

A further improvement in stabilization can be effected by reducing the resistance of the bias divider R_1. With the transistor tested, substituting a 10,000-ohm potentiometer for the 100,000-ohm one used in obtaining the above data reduced the total rise in collector current to less than 1/2 milliampere, starting from 25 mA. With the same value at R_1 and the collector input increased to approximately 650 mW (14.7 V at 45 mA) by increasing the bias current, the collector current rose 1.5 mA in 2 min and then remained constant for 1-1/2 hours, at which time the test was terminated.

Experiment 10

PURPOSE: To determine characteristics of a field-effect transistor and check behavior of grounded-drain circuit.

APPARATUS: Same as for Exp. 9 with the following changes: Substitute a junction-type FET (about 300 mW rated dissipation) for the bipolar type used in Exp. 9; substitute a 9-V battery ("transistor type") for the 1.5-V cell; add 1500- and 470-ohm, 1/2-watt resistors.

PROCEDURE (A): Connect the components as indicated in Fig. E10-1(A). It is assumed in this diagram that an n-channel FET will be used, so the polarity of the gate-bias battery should be as shown. (An n-channel depletion-type MOSFET can be substituted for the junction type without changing the procedure.) The object of the measurements is to obtain data on drain current vs. drain voltage for various fixed values of gate bias. The bias voltages, which are always negative with respect to the source and the common point of the circuit, should be taken at 0.5-volt intervals, beginning with zero bias, until the value is reached at which the drain current is cut off. The VOM scale should be 0-10 mA, or the next-larger one available if the instrument used does not have a 0-10 scale.

A typical set of data is as follows:

Readings should be taken at small intervals of V_{DS} when the drain voltage is low, because (just as in the case of the bipolar transistor in Exp. 9) the drain current changes most rapidly in this region. Also, as was remarked in Exp. 9, third-place figures in the tabulated data should not be taken too literally; they simply indicate that there is a *difference* detectable on the meter as compared with the preceding digit.

Comment: In this procedure the dissipation in the transistor was kept below 100 mW, about one-third the rating, in order to prevent undue heating. In making measurements such as these it will be observed that there is very little drift in drain current, as compared with the drift in collector current of a bipolar transistor, even though no heat sink is used on the FET. This is because the thermal generation of carriers, while the same in both types if they are made of the same semiconductor material and have comparable dimensions, is the *only* factor operating in the FET. In the bipolar type the major part of the drift is caused by the amplified leakage current (see Chapter 22); in addition, regular thermal generation is increased by the rising power input caused by the amplified leakage.

A series of curves can be plotted from the tabulated data as shown in Fig. E10-2. In this particular transistor the curves become almost horizontal in the pinch-off region – i.e., to the right of the points at which the curves begin to bend over as the drain voltage is increased. (See

V_{DS}	Drain current, mA, for									
	0	*−0.5*	*−1*	*−1.5*	*−2*	*−2.5*	*−3*	*−3.5*	*−4*	*V* *(V_{GS})*
15		8.8	7.0	5.4	3.8	2.4	1.2	0.45	0	*mA* *(I_D)*
12		8.8	7.0	5.4	3.8	2.4	1.2	0.45		
10	11.0	8.8	7.0	5.4	3.75	2.4	1.2	0.45		
7	10.5	8.8	6.95	5.4	3.7	2.4	1.15	0.45		
5	10.3	8.8	6.8	5.3	3.6	2.3	1.15	0.45		
4	10.1	8.4	6.5	5.2	3.5	2.2	1.15	0.45		
3	9.0	7.6	6.2	5.0	3.5	2.2	1.11	0.43		
2	6.9	6.0	5.0	4.2	3.1	2.0	1.08	0.4		
1	3.8	3.4	3.0	2.6	3.05	1.6	0.9	0.37		
0.5	2.2	1.8	1.7	1.4	1.2	0.9	0.55	0.27		

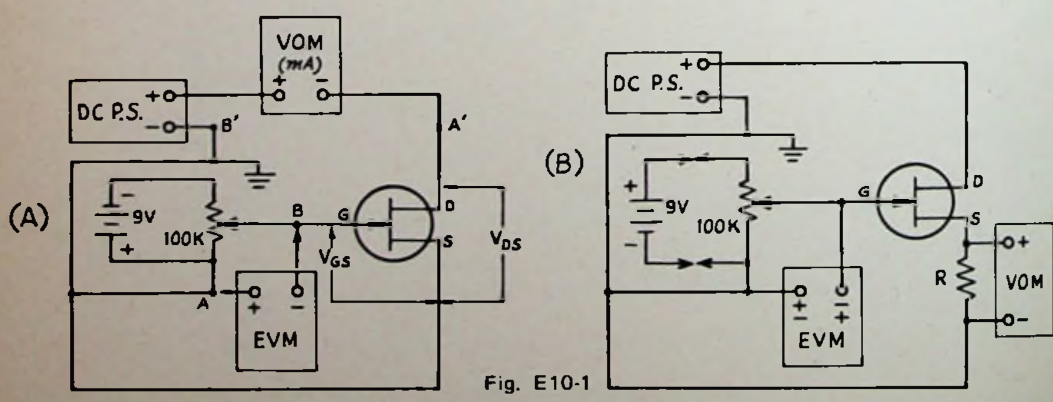


Fig. E10-1

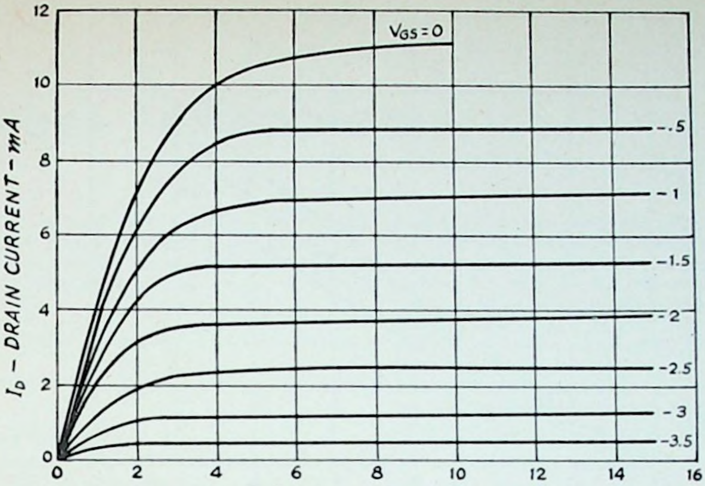


Fig. E10-2

Chapter 21.) The dynamic drain resistance is very high in this region. Load lines representing various values of load resistance can be drawn on these curves, as described in Chapter 23, and the voltage amplification can then be calculated from the intersections of the load line with the curves for various values of V_{GS}.

The same data supplies the information necessary for plotting a curve of drain current vs. gate voltage for any value of drain voltage that may be selected. Such a curve is given in Fig. E10-3 for V_{DS} = 10 volts. The curve is quite straight for gate voltages between −2 and 0, but at gate voltages more negative than −2 it becomes increasingly curved. Since the slope of this curve represents the rate of change of drain current with respect to gate voltage – i.e., the forward transconductance, g_{FS} – a curve of g_{FS} can be constructed by measuring the slope at intervals of V_{GS} over the selected range. In this case the slope is about 3.5 mA per volt at $V_{GS} = -1$, and is approximately 1.1 mA per volt at $V_{GS} = -3.5$. That is, the transconductance can be varied from 3500 μmho to 1100 μmho by varying the gate bias between these two limits. The transconductance will decrease to zero at the drain-current cutoff point, which is at slightly less than −4 volts on the gate with this particular transistor.

As explained in Chapter 23, the linearity of the drain current vs. gate voltage curve depends on the operating point and the load resistance. The curves of Fig. E10-2 can be used for obtaining the information necessary for constructing an *operating* curve of the type given in Fig. E10-3 (see Fig. 23-4 and accompanying discussion in Chapter 23).

PROCEDURE (B): The circuit for this procedure is shown in Fig. E10-1(B): The EVM and the VOM are used as dc voltmeters. A 0-10-volt scale is suitable for both. Resistor R is the load resistor for the common-drain or source-follower circuit.

Measurements should be made of the voltage across R (V_R) for several values of gate-to-common voltage, V_G. The actual gate-source bias voltage is equal to $V_G - V_R$, since the voltage drop across R opposes the battery voltage. In order to cover an adequate range of bias voltages it is necessary to reverse the polarity of the battery for some of the measurements. In the data tabulated below, the measurements were taken at intervals of 1 volt between gate and common, over the range from −4 to +9 volts.

In making these measurements it is necessary to make sure that V_R is always greater than V_G, both being positive with respect to ground. If V_R is less than V_G by more than about 0.2 volt, which would put that amount of *positive* bias between the gate and source, a measurable current will flow between the gate and channel, and as explained in

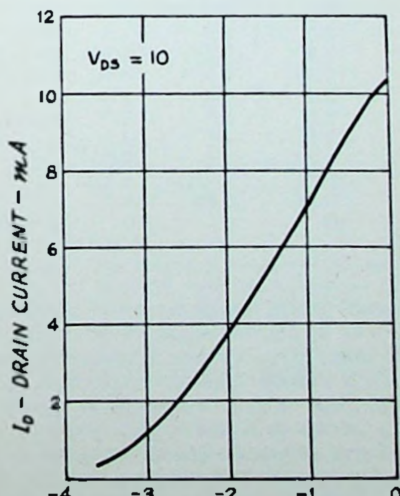


Fig. E10-3 GATE-TO-SOURCE VOLTS

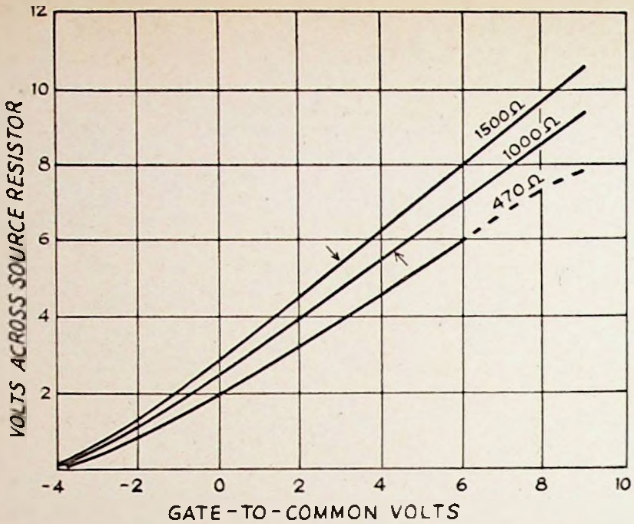


Fig. E10-4

Chapter 21, operation with positive gate-source bias is not desirable.

In this case, with a 470-ohm load resistor at R the gate-to-channel current was just detectable (less than 1 μA) when the gate-common voltage, V_G, was 6.2 volts. At V_G = 8 V the current was 500 μA (0.5 mA). The actual gate-source bias was 0.2 V in the first instance and 0.6 V in the second. For this reason no measurements are given in the table below for V_G greater than 6 volts. With the 1000- and 1500-ohm load resistors V_R was larger than V_G for any value of V_G up to the maximum V_G available, 9 volts.

A typical set of data is as follows:

	Voltage across R			
V_G	*R=470*	*R=1000*	*R=1500*	*ohms*
+9		9.35	10.6	*volts*
8		8.65	9.7	
7		7.85	8.85	
6	5.9	7.05	8.0	
5	5.2	6.3	7.1	
4	4.5	5.45	6.28	
3	3.85	4.7	5.42	
2	3.2	3.9	4.5	
1	2.6	3.2	3.7	
0	2.0	2.5	2.9	
−1	1.38	1.8	2.05	
−2	0.8	1.1	1.3	
−3	0.35	0.45	0.6	
−4	0	0	0	

Comment: These readings, when plotted in curve form as in Fig. E10-4, illustrate some facts about "follower"-type circuits. The improvement in linearity is evident from comparison of these curves with Fig. E10-3, especially at the low-current (or, in this case, low output voltage) ends. Perhaps a fairer comparison would be between the output voltage vs. gate-source bias curves that can be obtained from Fig. E10-2 by plotting load lines for the three values of load resistance on those curves; this can be done by the method outlined in Chapter 23.

The fact that the voltage gain of the follower circuit is less than 1 is also demonstrated by the curves. The small arrows on the 1000-ohm and 1500-ohm curves indicate possible operating points for the two load resistances, being at about the midpoints of the straight portions. With 1000 ohms, the operating point is at V_G = −4.3 V, and a swing of ±1 V, from 3.3 to 5.3 volts, causes the output voltage, V_R, to change from 4.9 to 6.4 volts. Thus a peak-to-peak voltage of 2 on the gate causes a peak-to-peak voltage of 6.4 − 4.9 = 1.5 at the source, so the voltage amplification is 1.5/2, or 0:75. The corresponding ratio is 1.7/2 for the 1500-ohm load, a voltage amplification ratio of 0.85. Still higher values of load resistance would increase the voltage amplification although, as explained in Chapter 25, it cannot ever quite reach unity. The amplification for the 470-ohm load could be worked out similarly, although it is evident from comparison of the curves that this value is lower than optimum as a source load for this transistor.

The increased signal-handling capability (gate swing) of this circuit as compared with the common-source circuit is also obvious when these curves are compared with those of Fig. E10-2 if the same values of load resistance are drawn on the latter curves as load lines. If this is done, it will be observed that for equally straight portions of the output-voltage curves (which can be constructed from Fig. E10-2) that the signal input and output voltages are much more limited when the common-source circuit is used.

Finally, it may be noted that Fig. E10-4 also shows that the input (gate) voltage and output (source) voltage swings are in phase in the common-drain or source-follower circuit. This contrasts with the phase relationship between input and output voltage swings in the common-source circuit when curves are constructed as outlined in the preceding paragraph.

Answers to Problems

The answers to problems requiring numerical solution are given below. (General questions are all based on the text, which should be consulted to check the correctness of answers.)

Chapter 1:

Q. 1-13) 0.0008 coulomb (8×10^{-4} coulomb).

Q. 1-14) 800 V.

Chapter 2:

Q. 2-3) 6.25 watt-hours.

Q. 2-4) 2.5 ohms.

Q. 2-10) 2500 circular mils.

Q. 2-14) The resistance of the wire of Q. 2-10 is 1/4 as large.

Q. 2-15) 50-mA range, R = 10 ohms. 500-mA range, R = 1.42 ohms. Both figures are averages of all measurements on each scale.

Q. 2-16)

Resistor (ohms):	Watts
47	2.2
100	0.98
150	0.62
330	0.287
680	0.152
1000	0.096 = 96 mW
2200	0.0458 = 45.8 mW
6800	0.0143 = 14.3 mW
10K	0.0105 = 10.5 mW

Q. 2-17)

13.2 mA:	0.025 W
26 "	0.097 W
39 "	0.217 W
57 "	0.465 W
69 "	0.681 W
81 "	0.938 W
97 "	1.346 W

Chapter 3:

Q. 3-1) Resultant = 3.16 ohms.
Total I = 1.90 A.
5-ohm resistor: I = 1.2 A; P = 7.2 W
14-ohm resistor: I = 0.429 A; P = 2.57 W
22-ohm resistor: I = 0.273 A; P = 1.64 W.

Q. 3-2) 114.3K (114,300 ohms).

Q. 3-3) 16,666 ohms.

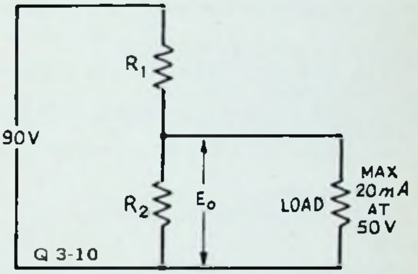
Q. 3-4) 1000-ohm resistor: 35.7 V; 0.0357 A (35.7 mA)
500-ohm resistor: 35.7 V; 0.0714 A (71.4 mA)
250-ohm resistor: 26.8 V; 0.1071 A (107.1 mA)
300-ohm resistor: 18.36 V; 0.0612 A (61.2 mA)
150-ohm resistor: 9.18 V; 0.0612 A (61.2 mA)
600-ohm resistor: 27.55 V; 0.0450 A (45.9 mA)

Q. 3-5) Connect 10K and 40K in parallel, giving 8K; add 12K in series for a total of 20K.

Q. 3-6) 316 V; 6.33 mA.

Q. 3-7) First tap, 20 V; second tap, 120 V; third tap, 220 V; fourth tap (top), 300 V.

Q. 3-10) The accompanying diagram is the circuit of the divider and load resistance. As explained in the text, the selection of a value for R_2 determines the divider design and the voltage regulation of the divider. In a simple divider such as this, R_2 may be chosen at will, including being left out of the circuit entirely; in that case, R_1 is a simple dropping resistor. However, the voltage regulation is poor unless R_2 is comparable with the load resistance. Arbitrarily, R_2 may be chosen equal to the load resistance, 2500 ohms; in that case R_1 must be 1000 ohms to meet the conditions of the problem. Other values may be tried.



Q 3-10

Q. 3-11) For the divider consisting of R_1 = 1000 ohms, R_2 = 2500 ohms, the output voltage E_o rises from 50 V at full load to 64.3 V at no load, a change of 14.3 V.

Q. 3-12) For the divider above, regulation = 28.6%; current at no load (25.7 mA) = 64.3% of current at full load (40 mA). Thus this divider meets the specified conditions.

Q. 3-13) For the divider above, power loss at full load: R_1, 1.6 W; R_2, 1.0 W. At no load: R_1, 0.66 W; R_2, 1.65 W.

Q. 3-14) Resistor powers in the four cases are as follows:

Resistance	*1st Case*	*2nd Case*	*3rd Case*	*4th Case*
680 ohms	0.068 W	0.147 W	0.057 W	0.112 W
220 ohms	0.022	0.455	0.177	0.455
100 ohms	0.01	1.0	0.141	0.016

Chapter 4:

Q. 4-14) k = 0.4.

Q. 4-18) The relative magnetizing force is twice as great in the second case as in the first.

Chapter 5:

Q. 5-1) 25 H.

Q. 5-2) No power is required for maintaining the fields, but the resistance loss in the 10-H inductance is 1 watt and in the 15-H inductance, 1.5 W.

Q. 5-3) 6 H.

Q. 5-4) 0.015 sec; from Fig. 5-4, 0.045 sec.

Q. 5-5) 4 μF.

Q. 5-6) Zero.

Q. 5-7) 0.6 sec.

Q. 5-8) From Fig. 5-4, for this change in voltage the ratio of actual time, T, to time constant, t, is 2.57. From this t = 0.39 sec and R = 39,000 ohms.

Chapter 6:

Q. 6-2) 4250 kHz.
Q. 6-3) 3.5 MHz.
Q. 6-4) 163 V.

Chapter 7:

Q. 7-2) 11,300 ohms.
Q. 7-3) 63.7 ohms; 3185 ohms.
Q. 7-4) 79.6 ohms; 318 ohsm.
Q. 7-5) At 2 MHz: 607 ohms, capacitive; at 6 MHz: 300 ohms, inductive; at 11.5 MHz: 946 ohms, inductive.
Q. 7-6) 4.107 MHz.
Q. 7-7) At 2 MHz: 247 ohms, inductive: at 6 MHz: 500 ohms, capacitive; at 11.5 MHz: 159 ohms, capacitive.
Q. 7-8) 2920 ohms, capacitive.
Q. 7-9) 85.5 ohms, capacitive.
Q. 7-10) 7.042 MHz.

Chapter 8:

Q. 8-2) Z = 722 ohms; power factor = 55.5%; no answer possible unless type of reactance is known.
Q. 8-3) Current leads; 707 ohms.
Q. 8-4) At 250 Hz: 2240 ohms; at 1 kHz: 629 ohms; at 40 Hz: 4500 ohms.
Q. 8-7) 12.5 W at any frequency; 0.282 A at 60 Hz; 0.2505 A at 500 Hz.
Q. 8-8) At 60 Hz: 0.142 W; at 1200 Hz: 34.5 W.
Q. 8-9) 552 ohms.
Q. 8-10) 0.00181 mho (at 100 Hz).

Chapter 9:

Q. 9-5) Impedance = 327 ohms, inductive; admittance = 0.00306 mho, inductive; phase angle = 40.2 deg.
Q. 9-6) Equivalent consists of a resistance of 429 ohms (G = 0.00233 mho) and an inductance of 101 mH (B = 0.00198 mho).
Q. 9-7) Equivalent consists of a resistance of 72.3 ohms in series with a capacitance of 116.8 pF (X = 195 ohms, capacitive).

Chapter 10:

Q. 10-3) 66.7%; 89% (from Fig. 10-6).
Q. 10-4) 8.9 watts; 4.5 W.
Q. 10-6) 85 dB.
Q. 10-7) Voltage gain = 250; power gain = 312.5 million.
Q. 10-8) Eff. = 75%; power in source = 33.3 W.
Q. 10-9) Eff. = 66.6%; safe output = 120 W.
Q. 10-10) 800 ohms, 60 W; 500 ohms, 37.5 W; 3000 ohms, 226 W.

Chapter 11:

Q. 11-1) 10 to 1.
Q. 11-4) Approximately 3 H minimum, assuming that the primary reactance should be at least 10 times the reflected primary impedance with load.
Q. 11-5) No; the primary L should be at least 30 H.

Chapter 12:

Q. 12-5) It is 35 times the line current (no *specific* current can be determined unless the circuit impedance, reactance, or resistance are known).
Q. 12-7) 6250 ohms; 80 kHz.
Q. 12-8) 2.6 A, neglecting the effect of the 10-ohm resistance.
Q. 12-9) 1692 ohms; 0.938 MHz.
Q. 12-16) At 2 MHz, Q = 50.2; at 3 MHz, 47.1.
Q. 12-17) Approximate solution: A parallel circuit consisting of 1200 ohms, 265 pF and 6.79 μH; or a series circuit consisting of 21.3 ohms and the same L and C. These figures are based on the simplifying assumptions that L and C have negligible internal loss and that the resonance curve is symmetrical about 3.75 MHz. Neither is strictly true, but the approximation would be a useful one in practice.

Chapter 13:

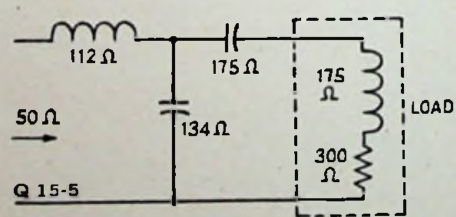
Q. 13-1) 667 ohms.
Q. 13-2) Parallel, 5 ohms; series, 500 ohms.
Q. 13-4) Reactance in primary, 125 ohms; in secondary, 333 ohms; mutual reactance, 17 ohms; coefficient of coupling, 0.83. (Note: It is unnecessary to consider source resistance because the optimum load resistance for the source is given.)
Q. 13-5) 11.1 μH; 44.2 pF.
Q. 13-6) 0.791.
Q. 13-9) None, provided the internal resistance of the tuned circuit can be neglected in both cases.

Chapter 14:

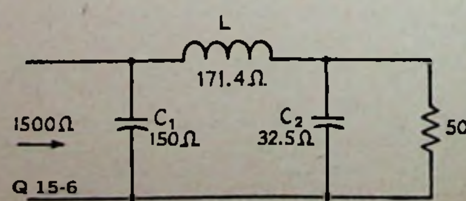
Q. 14-4) Approximately 70, by interpolation.
Q. 14-5) 10.3 ohms (series).
Q. 14-8) From Fig. 14-8, the bandwidth at 2X critical coupling is 0.03 times the center frequency. The required center frequency is therefore 100 kHz.

Chapter 15:

Q. 15-3) Series X, 172.5 ohms; parallel X, 174 ohms; C = 44 pF; L = 1.32 μH.
Q. 15-4) Parallel input reactance, 133 ohms; series reactance, 157 ohms; parallel output reactance, 27.5 ohms. No, because the minimum possible Q for a single L network would be 5.08.
Q. 15-5) Use an L network to step up 50 ohms to 300 ohms. The series reactance in the load must first be balanced out by adding 175 ohms of reactance of the *opposite* type in series, leaving a resistive load of 300 ohms to be matched. The matching network for this requires a parallel reactance of 134 ohms and a series reactance of 112 ohms. One such combination is shown in the accompanying figure.

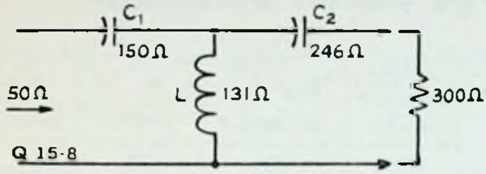


Q. 15-6) Reactance values are shown in the accompanying drawing (low-pass type circuit illustrated). Corresponding L and C values for 4 MHz are C_1, 265 pF; C_2, 1225 pF; L, 6.82 μH.



Q. 15-7) At 3.5 MHz (see circuit accompanying Q. 15-6) C_1 = 303 pF; C_2 = 1400 pF; L = 7.79 μH.

Q. 15-8) The network is shown in the accompanying diagram. At 14 MHz, C_1 = 75.8 pF; C_2 = 46.3 pF; L = 1.49 μH.

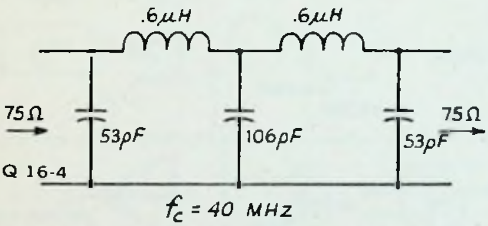


Chapter 16:

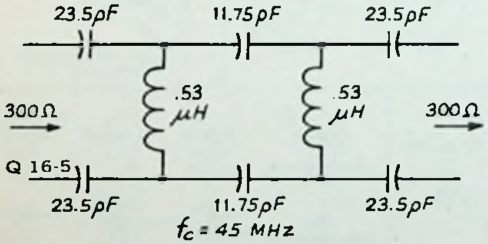
Q. 16-1) None; there is no dissipation in an ideal filter.

Q. 16-3) L_k = 0.299 μH; C_k = 53.1 pF.

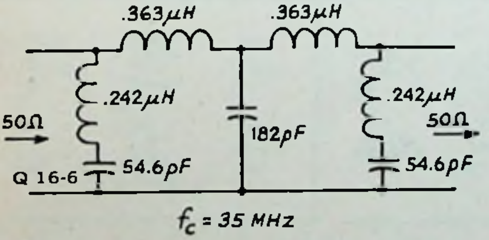
Q. 16-4) See accompanying circuit.



Q. 16-5) See accompanying circuit.



Q. 16-6) See accompanying circuit.



Chapter 17:

Q. 17-3) 80 ohms, inductive; 300 ohms, inductive.

Q. 17-5) 14.75 inches multiplied by the propagation velocity relative to that of light.

Q. 17-7) 66 ohms, if the dielectric is air.

Q. 17-8) 11.5 inches (50.7 deg).

Chapter 18:

Q. 18-3) 0.375 dB.

Q. 18-8) 31.6 ohms.

Q. 18-14) By calculation, using the standard transmission-line formula based on complex numbers, the input impedance is 77.9 ohms, consisting of 35.3 ohms resistance in series with 66.2 ohms inductive reactance. A good estimate from the curves of Fig. 18-4 would be R = $0.5Z_0$ and X = $0.9Z_0$ (line length = 135 deg). Accurate interpolation is difficult because of the divergence of the VSWR curves.

Q. 18-18) From Fig. 18-10, the additional loss is about 0.65 dB, making the total loss app. 1 dB.

Chapter 19:

Q. 19-3) 3 watts.

Chapter 20:

Q. 20-3) 212 volts.

Chapter 22:

Q. 22-10) β = 65; α = 0.985. (α is found by transposing the equation in the text for β, leading to

$$\alpha = \frac{\beta}{1+\beta}$$

Chapter 23:

Q. 23-7) By trying various load lines from V_{DS} = 30 V, a peak-to-peak drain swing of 20 V, from V_{DS} = 25.5 V to 5.5 V, can be obtained with a gate swing of −2 to −4 V, thus giving the required amplification of 10. The slope of this load line is 4000 ohms. The operating point is set by V_{GS} = −3, midway between −2 and −4 V. At this point the drain current is 3.5 mA and V_{DS} = 16 V.

Q. 23-8) 20 V pk-pk is about maximum, because with V_{GS} more negative than −4 V the drain current is approaching cutoff.

Q. 23-11) For 5000-ohm load, V.G. = 20; for 25K, 55.5; for 100K, 83.3.

Q. 23-15) Input resistance 530 ohms; V.G. signal source to collector, 7.2.

Chapter 24:

Q. 24-7) 286 ohms (the standard value of 270 ohms would be close enough in practice).

Q. 24-8) 32 V.

Q. 24-11) Gain with feedback = 7.5.

Q. 24-13) V.G. with cathode resistor = 9.7; gain with bypass would be essentially the same as without feedback – i.e., 12.

Q. 24-15) Input capacitance = 33.5 pF.

Chapter 25:

Q. 25-4) 188 ohms.

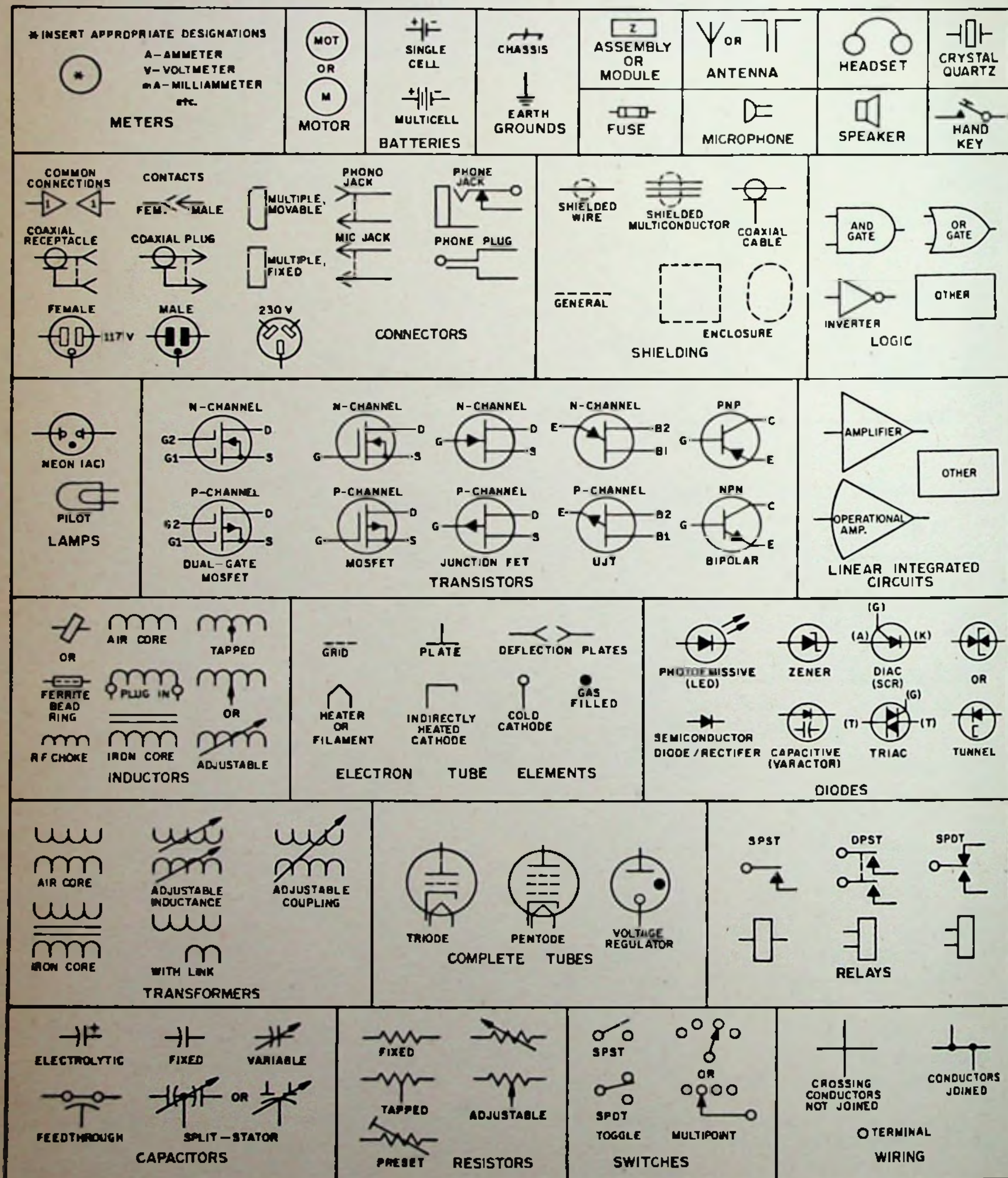
Q. 25-6) For A = 25, 3.9%; for A = 8, 11.1%.

Q. 25-9) 0.037 volt.

Q. 25-10) 95%.

Q. 25-13) Voltage gain with 100% feedback (cathode follower) = 0.95.

SCHEMATIC SYMBOLS USED IN CIRCUIT DIAGRAMS



Also see page 157.

The whole picture of amateur radio from basic fundamentals through the most complex phases of this appealing hobby is covered in ARRL publications. Whether novice or old-time amateur, student or engineer, League publications will help you to keep abreast of the times in the ever-expanding field of electronics.

THE RADIO AMATEUR'S HANDBOOK

Internationally recognized, universally consulted. The all purpose volume of radio. Packed with information useful to the amateur and professional alike. Written in a clear, concise manner. Contains hundreds of photos, diagrams, charts and tables. $7.50

SINGLE SIDEBAND FOR THE RADIO AMATEUR

A digest of the best articles from *QST*. Includes discussions of theory and practical how-to-build-it descriptions of equipment. Covers reception and transmission. $4.00

THE ARRL ANTENNA BOOK

Theoretical explanation and complete instructions for building different types of antennas for amateur work. Simple doublets, multi-element arrays, mobile types, rotaries, long wires, rhombics and others. Transmission lines are exhaustively discussed. Profusely illustrated. $4.00

HINTS AND KINKS

No builder should be without this collection of practical ideas for the workshop and station. Gathered together from the "Hints and Kinks" column of *QST*. $2.00

THE RADIO AMATEUR'S VHF MANUAL

A thorough treatment of vhf. Covers receiving and transmitting principles, techniques and construction, antenna and feed system design, uhf and microwaves, test equipment, interference causes and cures. Emphasis throughout is on tried and tested equipment and practice. $4.00

SPECIALIZED COMMUNICATIONS TECHNIQUES

Seven chapters cover the more esoteric forms of amateur radio: ATV, SSTV, FAX, RTTY, Satellite Communication, and advanced communication techniques. $4.00

ARRL HAM RADIO OPERATING GUIDE

Written for the amateur who prides himself on good operating procedures. A ready reference source and guide to good operating practices. Ideal for the amateur who wishes to brush up on operating procedures and who wishes information on all facets of amateur operating. $4.00

THE RADIO AMATEUR'S LICENSE MANUAL

Study guide and reference book. Points the way toward the coveted amateur license. Complete with typical questions and answers to the FCC amateur exams for Technician, General, Advanced and Extra Class. Continuously kept up to date. $3.00

ARRL ELECTRONICS DATA BOOK

Includes data on radio frequency circuits, filter design, *L*, *C*, and *R* networks, antennas and feed systems. A catalog of solid-state circuits and much more. This 128-page reference is a useful addition to the amateur's technical library. $4.00

TUNE IN THE WORLD WITH HAM RADIO

The complete beginner's package. Everything needed to obtain a Novice license: Theory, rules, how to assemble a station, and operating practices. With one-hour code cassette and U.S. call-district wall map. $7.00

UNDERSTANDING AMATEUR RADIO

For the beginner. Explains in simple language the elementary principles of electronic and radio circuits. Includes how-to-build-it information on low-cost receivers, transmitters and antennas. A "must" guide for the newcomer. $4.00

SOLID STATE DESIGN FOR THE RADIO AMATEUR

Takes the mystery out of solid state theory and design. Provides a wealth of tried and proven circuitry. Keeps mathematical treatment at a relatively simple level while providing a wealth of practical application data. $7.00

FM AND REPEATERS FOR THE RADIO AMATEUR

A complete treatment of fm and repeaters. Includes chapters on receivers, transmitters, antennas as well as repeater operation. Of great value to the fm buff as well as to those just becoming interested in the subject. $4.00

ARRL CODE KIT

Two top-quality cassettes provide practice at 5, 7-1/2, 10 and 13 wpm. Step-by-step progression and use of random code characters plus proven suggestions and hints in the accompanying instruction book make increasing your code proficiency easy! $8.00

Available from THE AMERICAN RADIO RELAY LEAGUE, NEWINGTON, CT 06111
PRICES ARE SUBJECT TO CHANGE WITHOUT NOTICE