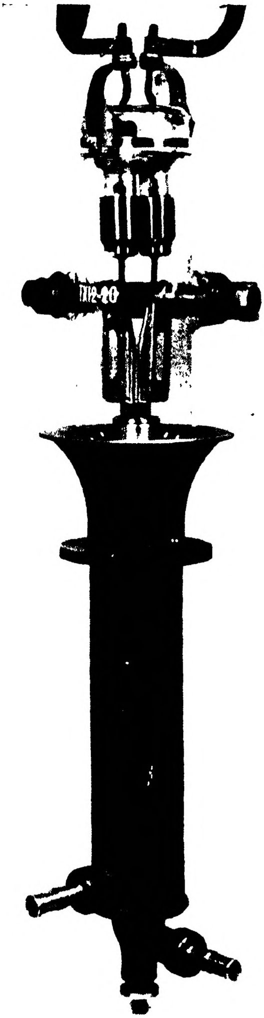
A TEXT BOOK OF ELECTRONICS



(*Mullard W.S. Co. Ltd.*)

(*R.C.A. Man. Co.*)

Left: A Large Water Cooled transmitting valve (Mullard TX 12-20)

Right: Electron Microscope Photographs
(*Upper*) Particles in zinc oxide smoke.
(*Centre*) Micro-organism E. coli, about to be attacked by the bacteriophage anti-coli.
(*Lower*) The same after 30 minutes, showing the destruction of the E. coli. (Magnification about 15,000).

A TEXT BOOK OF ELECTRONICS

by

J. M. A. LENIHAN, M.Sc.
Lecturer in Natural Philosophy in the University of Glasgow

ENGLISH UNIVERSITIES PRESS LTD.
SAINT PAUL'S HOUSE
WARWICK SQUARE
LONDON

First Printed March, 1948.

Made and Printed in Great Britain for the English Universities Press, Ltd.,
London, in complete conformity with the authorized economy standards,
by C. Tinling & Co. Ltd., Liverpool, London and Prescot.

PREFACE

THE purpose of this book is to explain, in the simplest language consistent with accuracy, the physical principles of electronics, and their application to some of the more interesting devices which they govern.

The first eight chapters are devoted to a consideration of fundamentals; after an introduction defining the terms to be used, a description is given of the phenomena which take place when a current passes through a gas or vapour in a discharge tube. Chapter III is concerned with the liberation of electrons from metals by the thermionic and photo-electric effects, and with the generation and properties of X-rays. The next chapter, dealing with the properties of alternating currents, is the only one in which any mathematics occurs, and is followed, in Chapter V, by a discussion of the construction and operation of valve rectifiers and their use in power supply circuits. The three succeeding chapters, describing the characteristics of various types of thermionic valves and their employment in amplifiers and oscillators, conclude the first part of the book.

The remainder of the book illustrates the application to practical problems of the principles enunciated in the first eight chapters. Chapter IX covers the essentials of radio communication and Chapter X takes up again the photo-electric effect, with an account of some of its uses in industrial processes. The succeeding chapter deals with discharge lamps and high-power arcs and a discussion of electronic heating follows in Chapter XII. Chapter XIII outlines the use of electronic methods in electrotherapy and Chapter XIV is devoted to the cathode ray oscillograph, with an account of its construction and operation, and some common applications. The final chapters are concerned with television and with the electron microscope.

Emphasis has been placed throughout on the physical laws and methods which are the foundation of electronics, rather than on the multitudinous applications and special techniques developed from them. Thus, although a conscientious reader may obtain from these pages a reasonably clear picture of what goes on inside his radio receiver, he will not (unless he is of uncommon perception) discover how to mend it when it stops working.

PREFACE

It is assumed that the reader has an elementary knowledge of mathematics and physics; with this equipment, he should find little difficulty in the pages which follow. The book is not intended to cover the syllabus of any particular examination, but it is hoped that it may be useful to students of physics and electrical engineering, as well as to more advanced workers in other sciences who may wish to know something of the principles underlying the applications of electronics to the subjects in which they are interested. The author is grateful to the manufacturers and publishers who have helped in the provision of illustrations, and to a number of friends who have offered comments on parts of the manuscript—but are in no way responsible for the imperfections which remain!

Contents

CHAPTER I

INTRODUCTION

The Science of Electricity. The first electrical experiments ever performed were probably those of Thales (about 600 B.C.), who observed that a piece of amber when rubbed would attract small objects such as pieces of paper and thread. The next developments were initiated by Gilbert (1540—1603) who found that glass, sealing wax, and a variety of other substances behaved in the same way. The position of the science after Gilbert's work may be summarised as follows.

1. The effect produced on amber, glass and sealing wax by rubbing them is called electrification and a body which is subjected to this process is said to be electrified or electrically charged.
2. There are two kinds of electricity :—vitreous electricity, produced, for example, by rubbing a glass rod with silk, and resinous electricity, produced, for example, on sealing wax rubbed with flannel.
3. A body charged with vitreous electricity repels another body similarly charged and attracts a body charged with resinous electricity. In other words, like charges repel and unlike charges attract each other.
4. Equal charges of resinous and vitreous electricity, if compounded, will neutralise each other.

An important advance was made by Franklin (1706—1790) who suggested that there is only one kind of electricity, which is possessed by uncharged bodies in certain definite amounts. An excess of this electricity constitutes a vitreous charge and a deficiency constitutes a resinous charge. Franklin further proposed that vitreous and resinous electricity should be termed positive and negative respectively. This fits in with his other proposals, since a deficiency of positive charge is the same thing as an excess of negative charge. In fact, Franklin's ideas about electricity are not greatly different from those now generally held. It is known that one form of electricity is fundamental ; unfortunately this is resinous (negative) electricity. As Franklin's sign convention was firmly established before this became known, it has not been altered. Most of our work in this book will deal with negative charges and negative electricity in general.

It is assumed that the reader is familiar with the common phenomena of electrostatics, but it will be convenient to introduce at this stage the law governing the force between two electric charges. We define the unit of electrostatic charge as that charge which repels an equal charge placed one centimetre away in air with a force of one dyne. For charges not of unit magnitude and not at unit distance the force is directly proportional to the product of the charges and inversely proportional to the square of the distance between them, so that we may write

$$F = \frac{Q_1Q_2}{d^2}$$

F being the force, Q_1, and Q_2 the two charges and d the distance between them. If the medium containing the two charges is not air, the law becomes

$$F = \frac{Q_1Q_2}{kd^2}$$

k is the dielectric constant of the medium.

The Electron. Present knowledge of the foundations of electricity may be summarised as follows. The fundamental unit of electric charge is the *electron*. The precise value of the electronic charge, denoted by the symbol e, is a matter for experiment and discussion, but it is approximately

$$4{\cdot}8 \times 10^{-10} \text{ electrostatic units (e.s.u.)}$$

This is the smallest charge which can exist and every electric charge must therefore be a multiple of e. Though the electron is primarily an electric charge and nothing more, it behaves as though it possessed a small mass—approximately

$$9 \times 10^{-28} \text{ gm.}$$

The exact nature of the electron is not clear, but we shall commit no serious error by regarding it as simply a small charged body.

The matter which surrounds us is composed of atoms and molecules. An atom may be defined as the smallest possible particle of a chemical element, and a molecule as an aggregation of atoms. Each atom is characteristic of the element from which it is derived, and as 92 elements are known there are 92 varieties of the atom.*

Before discussing the structure of the atom it is necessary to mention two more elementary particles which, with the electron, are the building stones of matter.

1. The *proton* has mass approximately 1840 times that of the electron and carries a charge of $+e$ (taking the electronic charge as $-e$.)
2. The *neutron* has the same mass as the proton but is uncharged.

The Atom. The atom consists of a positively charged central nucleus, which contains most of the mass, and a number of electrons which occupy a region surrounding the nucleus. The number of these electrons is the atomic number of the element and ranges from 1 to 92 for the naturally occurring elements. The negative charge borne by the electrons is balanced by the positive charge on the nucleus. The mass of the atom, expressed in suitable units (mass of oxygen atom = 16) is the atomic weight of the element concerned; thus the atomic weight of hydrogen is 1 and that of oxygen is 16. When a body is electrified, the distribution of charge on some of its atoms is altered; in a negatively charged body some of the atoms have a surplus of electrons that is to say, more than enough to balance the positive charge of the nucleus, and in the atoms of a positively charged body there is a corresponding deficiency of electrons. A charged atom or molecule is called an ion, and usually carries one unit of charge, though it is possible to produce multiply ionised atoms, carrying two or more units of charge.

The Size of an Atom. It is difficult to give definite figures as to the size of an atom, since the term connotes, not a single particle, but a region of space in which a number of sub-atomic particles are congregated. In certain circumstances, however, the number of atoms in a given quantity of matter may be calculated from known chemical and physical data, and an estimate may be made of the volume occupied by each. In this way it is found that the diameter of an atom is of the order of 10^{-8} cms. By means of other indirect calculations it is estimated

---

* Or 96, including Neptunium, Plutonium and two others which have been made by artificial means in the course of work on the atomic bomb (1940-1945).

that the mass and charge of a nucleus are concentrated into a region about 10^{-13} cms. across, and approximately the same figure is obtained for the electron. None of these estimates should be regarded as in any way exact; they serve merely to indicate the orders of magnitude which are involved in the study of atomic processes.

So far we have mentioned only stationary charges. In electronics, however, we shall be concerned mainly with moving charges. A stream of electrons (or for that matter one moving electron) constitutes an electric current. The magnitude of the current is proportional to the number of electrons passing a given point per second—not to the velocity with which they move. Here we may observe another source of confusion arising from Franklin's arbitrary allocation of positive and negative labels. A current is conventionally regarded as a transfer of positive electricity and, for example, a battery is considered to send current from the positive pole through the external circuit and back to the negative pole. From what has been written it is apparent that what actually takes place is a transfer of electrons in the opposite sense. The statement that a current flows from A to B really means that a stream of electrons flows from B to A. For many purposes either of these viewpoints may be adopted, but in electronics the latter is usually necessary.

Ions. We have seen that the atom in its normal state is electrically neutral. It is possible to increase or decrease temporarily the number of electrons revolving round the nucleus and thus to produce negatively or positively charged atoms. An atom or molecule which has been treated in this way is said to be ionised and the charged atom or group of atoms is called an *ion*. A gas may be ionised by exposing it to electromagnetic radiation of short wavelength, such as ultra-violet light or X-rays. Then equal numbers of molecules acquire positive and negative charges; that is to say, some of the molecules lose electrons and an equal number gain them.* Each molecule usually gains or loses one electron, but it is possible to obtain double ionisation by removing or adding two electrons and still higher stages are possible. As the ions are charged, they move towards the electrodes between which

---

* In some gases, such as nitrogen, hydrogen and helium, the electrons released by the formation of positive ions remain free and do not form negative ions by attaching themselves to neutral molecules.

the electric field exists. The positive ions move to the negative plate or cathode and each removes an electron from it, thus becoming neutral again. The negative ions move to the positive plate or anode and give up their surplus electrons. While these processes are going on, numerous collisions occur between ions and neutral molecules, as well as between negative and positive ions. Collision of an ion with a neutral molecule may produce further ionisation, especially if the ion is moving rapidly, and collisions between ions of opposite charges generally produce two uncharged molecules. The collisions therefore lead to further ionisation and to recombination. Eventually (actually after a very short time) an equilibrium is established and the number of ions present remains steady thereafter. Then the net effect of the electric field is a transfer of electrons from the cathode to the anode, the ions acting as carriers. This means that the gas conducts electricity, since a flow of electrons constitutes a current.

Behaviour of the electron in electric and magnetic fields. This is a subject of some importance, and it will be useful to outline here the principal conclusions. An electric field is set up between any two conductors which are at different potentials, and behaves towards electric charges in much the same way as does a magnetic field to magnetic poles. We shall consider the case of an electron in a uniform field formed between two parallel plates separated by air ; one plate is at a higher potential than the other (fig. 1/1). By definition, the intensity of this field, that is, the force exerted by it on a unit electric charge, is equal to the potential gradient. If the plates are at a distance d apart and the potential difference between them is V, then the intensity is $\frac{V}{d}$; V is in electrostatic units (1 e.s.u. = 300 volts). The force exerted on the electron in this field is therefore $\frac{Ve}{d}$ dynes.

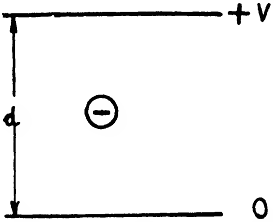


Fig. 1/1. In an electric field, an electron moves towards the positive electrode.

This is the force with which the electron is attracted towards the anode. If the electron moves through a distance s cm. under this force, the work done on it by the field is $\frac{Ves}{d}$ dynes. Now $\frac{V}{d}$ is the potential gradient, and the potential difference through which the electron falls is therefore $\frac{Vs}{d}$. We may therefore conclude that the work done on an electron when it falls through a potential difference X is Xe. This work appears as kinetic energy, the electron of mass m acquiring a velocity v given by

$$\tfrac{1}{2}mv^2 = Xe, \qquad \text{1·1}$$

a result which will be useful. The electron-volt is a convenient unit of energy in atomic physics ; it is the energy acquired by a particle of charge e in falling through a potential difference of one volt and is approximately equal to $1{\cdot}6 \times 10^{-12}$ ergs.

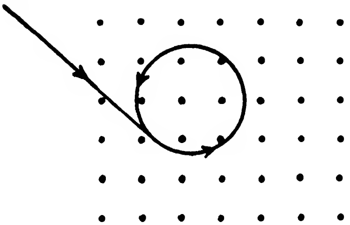


Fig. 1/2. A charged particle, projected into a magnetic field at right angles to the lines of force, executes a circular path.

Now consider an electron shot into a magnetic field of strength H with speed v, in a direction at right angles to the lines of force (fig. 1/2;) the lines of force are perpendicular to the paper, with which their intersections are shown by dots). This electron is acted on by a force of Hev in a direction perpendicular both to the path of the particle and to the lines of magnetic force. It therefore traces out a circular path, with the force acting towards the centre ; the circle is in a plane perpendicular to the lines of force, and has a radius r given by the equation :

$$Hev = \frac{mv^2}{r} \qquad \text{1.2}$$

CHAPTER II

DISCHARGE PHENOMENA IN GASES AND VAPOURS

THE possibility of electrical conduction through a gas, resulting from ionisation of its molecules, has already been mentioned (p. **10**). In the present chapter the various processes involved in this form of conduction will be discussed more fully.

Ionisation in gases. Under normal conditions a gas is always ionised to a small extent by the action of such agencies as cosmic radiation and emanations from radioactive substances. If a potential difference is applied between two metal electrodes in a gas, the positive ions move towards the negative electrode and the electrons towards the anode. These movements of charge constitute a current I, which depends on the applied voltage E, in the manner illustrated by fig. 2/1. (This type of conduction is known as the dark discharge.) As the voltage across the discharge tube is increased, the current increases—along AB in fig. 2/1—up to a value which may be about 10^{-17} amps.* at which it remains almost constant over a wide range of voltage (BC). At C a sudden increase of current takes place and the discharge assumes characteristics which will be discussed later.

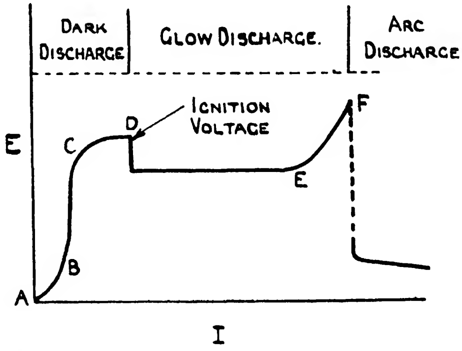


Fig. 2/1. Illustrating the relation between current and applied potential difference in a gas or vapour discharge.

The processes occurring in the dark discharge are briefly as follows. At first the current increases because the increasing electric field between the electrodes draws more ions across. At B the field is strong enough to remove ions from the space between the electrodes as rapidly as they are formed. As long as the formation of ions depends on an external influence, no further increase in current is possible. However, as the electric field

---

* Depending on the pressure of the gas, and on the dimensions of the tube.

continues to increase, the ions and electrons moving about in the interelectrode space acquire greater and greater energies, eventually moving quickly enough to ionise the gas by collision with its molecules. This fresh supply of ions permits an increased current, but the discharge becomes more complicated since the tube now contains ions and excited atoms as well as the original gas. The energy changes involved in ionisation, excitation and the various associated processes may give rise to radiation of energy in the visible or non-visible regions of the spectrum. In the region beyond *D* in fig. 2/1 the rate at which ions are produced is sufficient to keep up the discharge without the assistance of any external agency, and the discharge is said to be self-maintaining. The potential difference which must be established across the tube to initiate a discharge of this kind is referred to as the ***breakdown or ignition voltage.*** For a given gas it depends on the pressure and on the nature and separation of the electrodes (fig. 2/2). With plane parallel electrodes at a given distance, the ignition voltage is least for one particular value of the pressure and there is similarly an optimum separation for any given pressure. Thus if the separation of the electrodes is kept constant, discharge does not readily occur at a low pressure, since the distance between the gas molecules is then comparatively great, and an ion or electron travelling across the inter-electrode space makes few collisions. Ionisation sufficient in extent to maintain the discharge can only be produced if the energy of the ionising particles is increased by raising the voltage across the tube. At high pressures, the ionising particles travel only a short distance between collisions

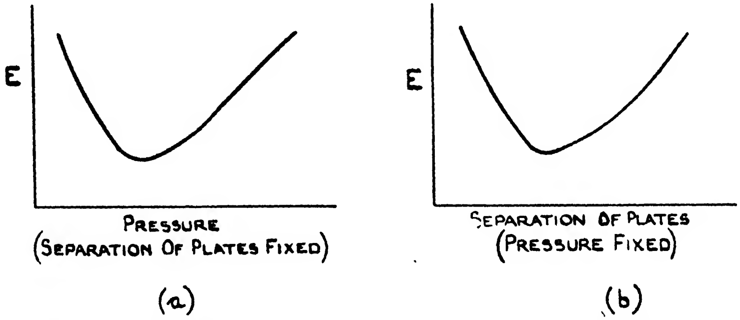


Fig. 2/2. Showing the relation between sparking potential, pressure and electrode separation in a gas.

and do not gain enough energy to cause further ionisation unless the field strength is high. At some intermediate pressure, ignition occurs with a minimum voltage across the tube (fig. 2/2a). Similar reasoning accounts for the shape of fig. 2/2b.

The results set out graphically in fig. 2/2 may be combined in the form of *Paschen's law,* which states that in a given gas at constant temperature, the breakdown voltage between plane electrodes depends only on the product of the pressure and the electrode separation. The number of collisions made by an ionising particle in the inter-electrode space is proportional to the concentration of gas molecules, that is, to the pressure. It is also proportional to the distance between the electrodes. Thus if the pressure is increased and the separation is reduced in the same ratio, the product remaining constant, the number ot collisions made by a particle traversing the inter-electrode space remains the same. Breakdown occurs when the rate of ionisation by collision reaches a certain value sufficient to maintain the discharge and the ignition voltage thus depends on the product of pressure and electrode separation.

The glow discharge. At *C* in fig. 2/1 the current begins to rise as the potential difference across the tube is increased and at *D* it rises very steeply to a value of the order of one microampere. The potential difference across the tube in these circumstances remains almost constant over a wide range of current until at *E* in fig. 2/1 it begins to rise again and, if the external circuit is of low resistance, the discharge may change into an arc. In the region *DE* the discharge is spoken of as a normal glow, and between *E* and *F* as an abnormal glow. In both cases the discharge is self maintained, that is, it does not depend on the presence of an external agency for the production of ions.

Characteristics of the glow discharge. The principal characteristic of the glow discharge is the emission of light, the colour depending on the nature of the gas in which the discharge occurs ; for air it is pink. This luminosity does not fill the whole of the tube and the inter-electrode space has the appearance shown in fig. 2/3 which refers to mercury vapour at a pressure of about one millimetre. It will be observed that there is a large potentential drop—as much as 300–400 volts in some cases—in the cathode dark space, and a fairly constant potential gradient of about 20 volts per cm. in the positive column, which fills most of the tube. There is a small potential drop (10–20 volts) in the dark space near the anode. If either electrode is moved, the discharge remains in the same position relative to the cathode, and the length of the positive column alters accordingly. Close to the cathode is a thin luminous layer known as the cathode glow. At low currents, this luminosity covers only part of the cathode surface and its area increases as the current increases. The current density at the cathode therefore remains at the same value and the cathode fall of potential, that is, the voltage drop between the cathode and the beginning of the positive column, remains constant. The discharge under these conditions is spoken of as a normal glow. When the glow covers the whole of the cathode, its area can no longer be proportional to the current, and the current density therefore increases. The discharge is then known as an abnormal glow. In this régime, the cathode

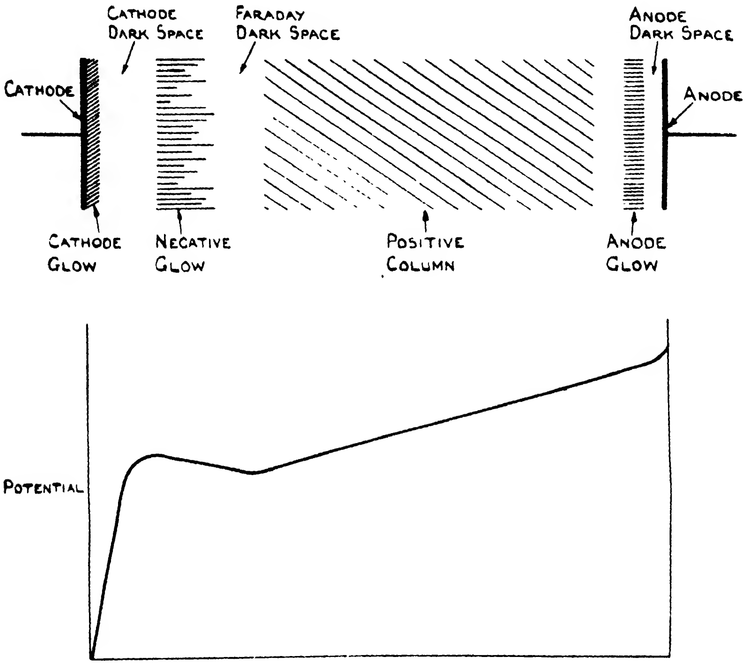


Fig. 2/3. General appearance of a glow discharge in a gas or vapour. Below—variation of potential along the discharge.

fall of potential increases—though only slowly—with increase of current. The potential drop in the positive column decreases gradually as the current increases, and the total voltage drop across the tube remains almost constant for wide variations of current. This property of the discharge is used in voltage stabilisers.

To explain the production of the discharge, it is necessary to postulate the presence of a small number of ions in the tube before any potential difference is applied. This is a reasonable assumption, since the air is always ionised to a small extent by radiation from the sun. When a potential difference is applied across the tube the positive ions are accelerated towards the cathode, where the bombardment causes the emission of electrons which move towards the anode, gaining considerable energy in the region near the cathode, on account of the large potential drop in this part of the tube. They produce ionisation by collision with the molecules in their path and there is a cumulative increase in the number of ions present. As the ions and electrons move about, some recombination takes place and eventually the rate at which electrons are removed in this way becomes equal to the rate at which they are produced by collision, so that the discharge becomes stable—a process occupying only a small fraction of a second.

Luminous nature of the discharge. The luminous regions of the discharge have been the subject of considerable discussion, of which it is hardly possible to give an adequate account here, but a summary of the more important processes involved will be useful. Light may be produced when an ion and an electron recombine. In this case, some of their kinetic energy is changed into radiation and if the amount of energy taking part in this change is E ergs, then one quantum of radiant energy is produced, with a frequency ν, given by the relation

$$E = h\nu \quad \ldots\ldots 2{\cdot}1$$

where h is Planck's constant, equal to

$$6{\cdot}6 \times 10^{-27} \text{ erg. sec. approximately.}$$

The quantum theory which was formulated by Planck in 1901 indicates that the energy in light or other forms of electromagnetic radiation is not uniformly distributed, but occurs in packets, or *quanta*, of $h\nu$. Emission of light consists in the production of one or more quanta, and absorption similarly involves the transformation of individual quanta into some other form of energy.

A number of important processes resulting in the emission of radiation may occur if an atom is *excited*. In an excited atom, one electron is lifted from its normal state to one of higher energy ; in the case which we are considering, this change may occur as the result of a collision. The electron returns almost immediately to its original state, sometimes in two or more stages, and the energy thus released appears as radiation. In gas and vapour discharge tubes this radiation is usually in the visible or ultra-violet region of the spectrum. If the electron proceeds from the excited state directly to its normal energy level, it emits resonance radiation ; if the return takes place by way of intermediate energy levels, the frequency of the radiation emitted in each step is smaller and its wavelength is therefore greater than that of the resonance radiation. Another possibility is a *collision of the second kind*, in which the energy of excitation is transferred as kinetic energy to another electron, without emission of any radiation. Glow discharge tubes are important as sources of illumination and are discussed further in chapter XI.

If the current in a glow discharge is allowed to increase there occurs at one point, in a manner not fully understood, an abrupt change to a new régime, that of the arc discharge, which is characterised by large current—several amperes—and low voltage drop across the tube, perhaps ten volts or less. The onset of the arc is accompanied by changes at the cathode, where current is drawn chiefly from one small area rather than from the whole surface as in the glow discharge. The cathode thus becomes heated to a considerable extent and melting or vaporisation may occur. The voltage drop across an arc diminishes if the current is allowed to increase, and

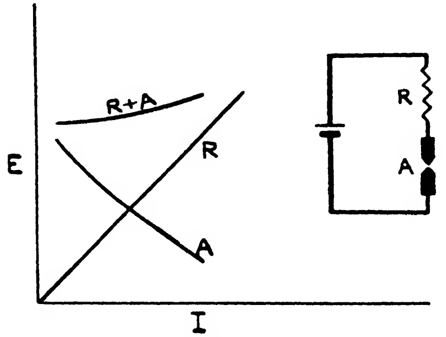


Fig. 2/4. The potential difference across an arc decreases if the current through it increases and for stable operation a resistance is connected in series with the arc.

the current which is drawn may rise to a value limited only by the resistance in the external circuit. To secure stable operation, an arc (A, fig. 2/4) must always be connected in series with a resistance *R* of suitable value. Fig. 2/4 shows the potential: current relationships in each part of the circuit. Applications of the arc discharge are discussed more fully in chapter XI.

CHAPTER III

THERMIONIC AND PHOTO-ELECTRIC EMISSION; X-RAYS

ELECTRONS *may be removed from the surface of a body in several ways, of which the most important are thermionic emission and photo-electric emission.*

Thermionic emission. Electrons may be made to leave the surface of a body most easily by the application of heat. This is the thermionic effect, discovered by Edison in 1885. A filament of metal, or a coating of a metallic salt on a wire, emits positive ions and electrons if it is heated. At temperatures above 1000° C. the electrons form the greater part of the emission.

To give a satisfactory explanation of thermionic emission it is necessary to introduce a slightly more complicated picture of the atom than that put forward on page 9. There are theoretical reasons for believing that in conductors some of the electrons are not permanently associated with any particular atom, but are free to move as they do, for example, under the influence of a temperature difference (thermal conduction) or a potential difference (electrical conduction). A poor conductor on this view is one in which most of the electrons are firmly bound to the atoms and the movement of electrons which constitutes a current is therefore difficult. The *conduction electrons*, as they are called, behave in very much the same way as the molecules of a gas or a liquid—thus they move about at random inside the metal with high speeds—and thermionic emission may aptly be compared with the evaporation of a liquid.

A liquid evaporates when some of its molecules break through the surface into the atmosphere outside. At low temperatures the energy of the molecules is not very great, and only a small proportion of them can overcome the surface forces which prevent them from escaping. If energy in the form of heat is supplied to the liquid, a greater proportion of its molecules are able to break out and the rate of evaporation is increased until, at the boiling point, all the molecules have enough energy to escape and the liquid changes completely to vapour. If a metal is heated, some of the energy supplied is distributed among the conduction electrons, some of which are enabled to escape. To break out of the metal, an electron must overcome a potential barrier at the surface, corresponding to the latent heat in a liquid, and as the temperature is increased a greater number *of electrons are able to do this. If no arrangements are made to remove the emitted electrons they form a cloud near the* wire; individual electrons still move about as though they were molecules of a gas and after a short time the rate at which electrons are returned to the wire by this random motion becomes equal to the rate at which they are emitted, so that the number of electrons surrounding the wire remains more or less constant. This cloud is referred to as a *space charge.* If an electrode at a higher potential than the emitting surface is placed near, electrons are attracted to it from the space charge and the negative electricity which they carry reduces the potential of this electrode, so that the flow eventually stops. By using a bulb from which most of the air has been evacuated in an arrangement of the kind sketched in fig. 3/1 the electrons captured by the positive electrode, or anode, are returned to the emitter, or cathode, and current flows continuously. If the anode is at a negative potential with respect to the cathode, no current flows and if an alternating e.m.f. is used instead of the battery of fig. 3/1 then current flows during the half of each cycle in which the anode is positive but not during the other half. Thus an alternating current may be turned into a unidirectional current and as we shall see later, it is not difficult to obtain a steady current. On account of this property a vacuum tube

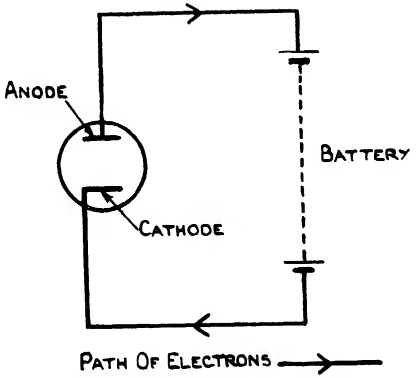


Fig. 3/1. In the diode valve, a heated metallic filament emits electrons which are attracted by the positively charged anode and return through the external circuit.

of the kind described is referred to as a ***valve.*** The first two-electrode or ***diode*** valve was made by Fleming in 1904. The emitting wire, known as the ***filament*** of a valve is heated by passing a current through it and in order to avoid oxidation the electrodes must be enclosed in a highly evacuated glass or metal bulb.

The laws governing thermionic emission, which are based on the experiments of O. W. Richardson and others, may be summarised as follows. The three variable quantities in a diode valve are :

(1) The cathode temperature, which depends on the filament current.

(2) The potential difference between anode and cathode.

(3) The current *flowing through the valve.*

Variation of current with cathode temperature. The total number of electrons emitted by a filament is related to the surface temperature by the relation

$$I = A\,T^2 e^{-\frac{b}{T}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3{\cdot}1$$

where I = quantity of electricity emitted in coulombs/sq.cm/sec.
T = absolute temperature of the surface.
A and b are constants characteristic of the emitting substance.

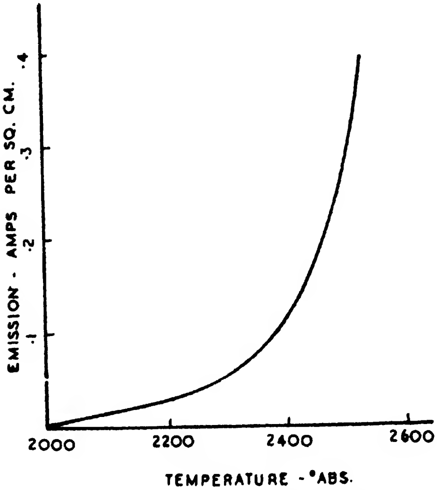


Fig. 3/2. Showing the relation between temperature and emission for a typical metallic filament.

A graph of this function against T is shown in fig. 3/2. If the anode voltage is fixed, the graph of current against filament temperature takes the form shown in fig. 3/3. At low filament temperatures the anode attracts all the electrons emitted and the current is therefore equal to the total emission. As T is increased the space charge increases and complications ensue ;

(1), the space charge shields the emitting surface from the influence of the anode so that the anode draws electrons out of the space charge rather than from the cathode itself, and,

(2), for electrons to be attracted in this way the electrostatic field produced by the anode must be great enough to overcome the mutual repulsion of the electrons in the space charge.

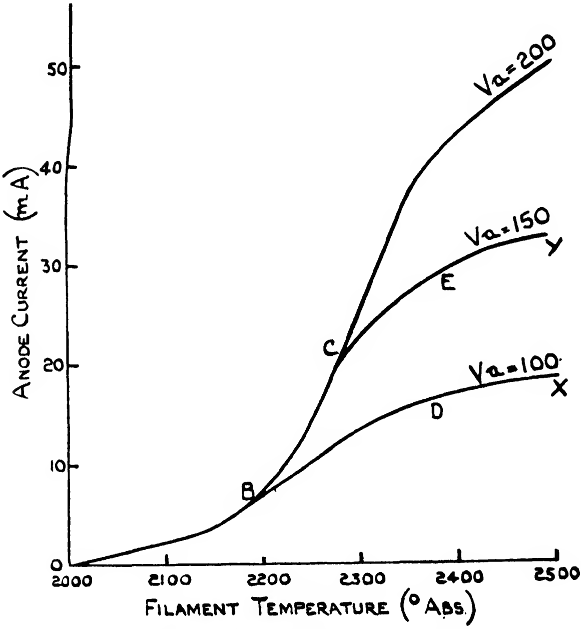


Fig. 3/3. Showing the relation between filament temperature and anode current for various values of anode potential in a typical diode.

For a given value of anode voltage this latter condition will be satisfied as long as the space charge is below a certain size. If the filament temperature is increased beyond this point, so that the space charge becomes larger, the anode is unable to take advantage of the increased emission and the anode current becomes nearly constant—*DX* in fig. 3/3. If the anode voltage is increased the current reaches a higher value before becoming stationary—*EY* in fig. 3/3. The portion *ABC* of fig. 3/3 is a portion of the emission curve shown in fig. 3/2. Valves are usually operated on this part of the graph in order that the current may be limited by the anode voltage and not by the filament temperature, which is liable to variation from such causes as fluctuation in the operating current and alteration in the external temperature; along *ABC* the current flowing is

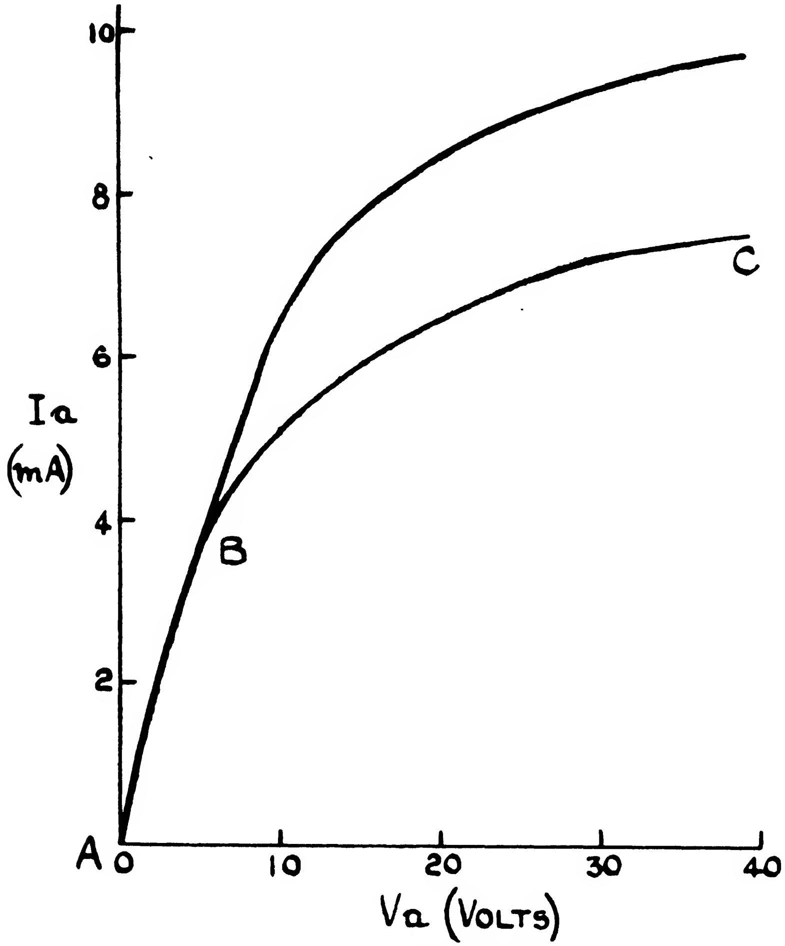


Fig. 3/4. The lower curve, *ABC* shows the variation of anode current with anode potential for a typical diode. If the filament temperature is increased, the available electron emission becomes greater and the upper curve is obtained.

not influenced significantly by small changes in filament temperature. In the regions DX and EY, the valve is said to be saturated ; this state is spoken of more exactly as *temperature saturation* or *space-charge saturation.*

The relation between current and anode voltage with filament temperature held constant is shown in fig. 3/4. The portion AB of the graph follows Child's Law

$$I \propto V^{3/2}$$

where I is the current and the constant of proportionality depends on the size, shape and separation of the electrodes. At a certain value of anode voltage saturation occurs, because the anode is attracting all the electrons emitted by the filament, and further increase in current can be obtained only by increasing the cathode temperature. Beyond C a state of voltage saturation begins and valves are not normally operated in this condition.

Cathode materials. The constant b in equation 3/1 is equal to $\frac{\phi e}{k}$ where k is the Boltzmann constant (or gas constant for one molecule), e is the electronic charge and ϕ is the *thermionic work function* representing the potential barrier which the electron must overcome at the surface to escape. It is apparent that to obtain a copious emission of electrons, T should be high and ϕ should be small. In practice T is limited to a value a few hundred degrees below the melting point of the cathode and increased emission can more easily be obtained by the use of materials with a low work function. The emitting surfaces now used fall into three main groups.

(1) Pure metals. Tungsten is the only metal extensively used ; its work function is 4·5 volts, which is rather high, but its melting point is about 3300°C. and, in fact, tungsten filaments can be operated at temperatures up to 2200°C. A large emission may thus be obtained, though the power needed to raise the filament to its working temperature is sometimes inconveniently great.

(2) Oxide-coated filaments. The process of manufacture is complicated and, to a certain extent, secret, but it may be summarised as follows. The filament has a core of nickel or of an alloy on which is coated the active material—a mixture of barium and strontium oxides with a little of the pure metals.

In the first place, a paste of strontium and barium carbonates with a little carbon is spread on to the core and when the valve has been assembled and evacuated the filament is heated, by the passage of a current, to about 1200°C., when the carbonates change into oxides and small amounts of the metals themselves are formed by reduction. The carbon dioxide produced is removed by further evacuation and the temperature is then reduced to about 1000°C. In the final stage a voltage is applied to the anode for a time, during which some of the oxide migrates to the surface where it forms a very thin layer with metallic particles distributed in it. Emission takes place principally from these particles, though the processes involved in the formation of the cathode surface and the emission of electrons from it are not fully understood. Oxide-coated cathodes which are widely used in small valves for radio receivers give a satisfactory emission at a temperature of about 700°C. and therefore do not require much power for heating; in this respect they have an advantage over tungsten filaments. The work function of oxide-coated cathodes is about 1 volt.

(3) Thoriated tungsten and caesiated tungsten filaments are also used though not to such a large extent as the oxide-coated type. In the former, a little thorium oxide is mixed with the tungsten during the manufacturing processes and the filament is operated for a short time at about 2200°C., when some of the oxide is reduced to thorium which forms a thin layer over the tungsten surface after a further few minutes' heating at about 1900°C. The normal operating temperatures for cathodes of this kind is about 1750°C. and the work function is 2·6 volts—less than that of pure thorium. Caesiated tungsten filaments are made by depositing a thin layer of caesium on a tungsten wire and more efficient than any of the types so far mentioned, having a work function of only 0·7 volts. Filaments of this kind have, however, several disadvantages which make them unsuitable for use in commercially-produced valves.

Cathode construction. The cathode of a valve may be heated directly or indirectly. In the former type, heating current is passed through the wire from which emission is to take place. Cathodes of this kind are usually made in one of the shapes shown in fig. 3/5. The emitting surface may alternatively be indirectly heated by a separate coil, insulated from it as is shown in fig. 3/6. The advantage of this method of construction is that alternating current may be used to supply the heating coil. If a filamentary (directly heated) cathode is supplied with alternating current the potential difference between anode and cathode has an alternating component superimposed on it and this is generally undesirable. It is often useful to be able to operate apparatus such as radio receivers entirely from alternating current mains, and in this case a low alternating voltage for heating the cathodes may be obtained by the use of a transformer: direct current heating would be inconvenient in these circumstances. Directly heated valves of British manufacture usually require a filament current of 0·1 amp. and are so designed that this current is obtained with 2 volts across the filament. Indirectly heated cathodes are usually operated at a current of 1 amp. which requires 4 volts across the heater, or at 0·3 amp., 6·3 volts. American valves have a wider range of heater currents and voltages.

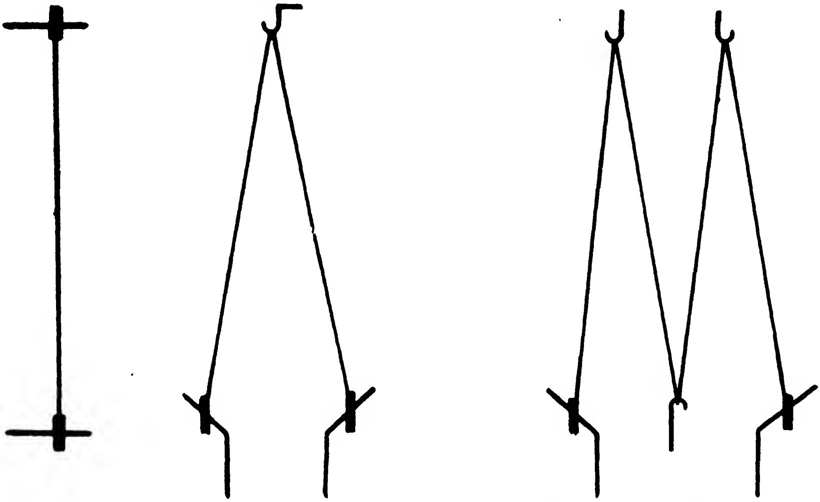


Fig. 3/5. Some common filament shapes in directly-heated valves.

Photo-electric emission. A phenomenon similar in many ways to thermionic emission is photo-electric emission in which electrons are dislodged from the atoms of a substance by the incidence of light or other electromagnetic radiation. In metals the conduction electrons (page 20) abstract energy from the radiation and some of them are then able to overcome the surface forces. As in the case of thermionic emission, this occurs if an electron has at least an initial velocity v_0 given by

$$\tfrac{1}{2}mv_0^2 = \phi e \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ 3{\cdot}2$$

The factor ϕ represents the potential barrier at the surface of the metal and, as might be expected, is identical with the thermionic work function.

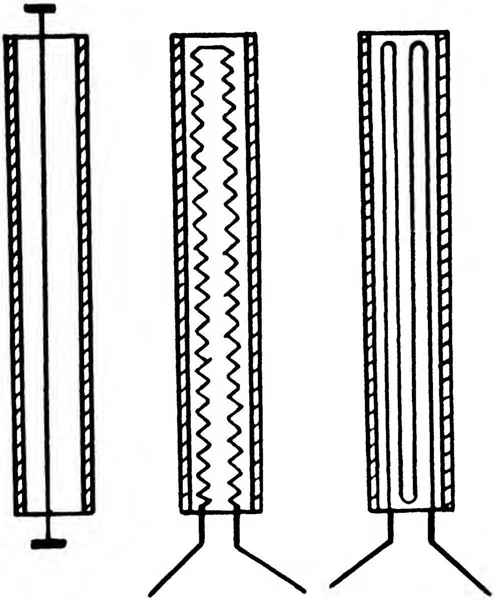


Fig. 3/6. Some common types of indirectly-heated cathode. The heater wire is inside a porcelain or metal tube which is coated on its outer surface with a layer of emitting material.

Many metals display the photo-electric effect in ultra-violet light, and some, such as sodium and potassium, in visible light. Experiments on photo-electricity have led to the following conclusions.

(1) The number of *photo-electrons* emitted in unit time is proportional to the intensity of the incident radiation.

(2) The velocity with which the electrons are emitted does not depend on the intensity of the radiation. On ejection the photo-electrons are found to have velocities distributed between zero and a sharply defined upper limit which depends on the frequency of the radiation used.

(3) For a given metal, no photo-electric emission is caused by radiation, however intense, if its frequency is less than a certain value.

The photo-electric effect was explained by Einstein in 1906 on the basis of the quantum hypothesis put forward by Planck in 1901. (cf. p. 17) The Einstein theory of the photo-electric effect is as follows :—

(1) Each electron may absorb one quantum of energy from the radiation.

(2) If the energy $h\nu$ obtained in this way is less than ϕe, the energy needed to overcome the surface forces, then the electron cannot escape.

(3) If $h\nu$ is greater than ϕe, the electron escapes from the surface with a velocity which can have any value up to v max where

$$\tfrac{1}{2}mv^2_{\max} = h\nu - \phi e \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3{\cdot}3$$

The velocities of photo-electrons may be measured by deflection experiments in electric and magnetic fields and are found to be in agreement with Einstein's law. Another verification of the law is made possible by considering that if in equation 3.3 above $h\nu = \phi e$ no electrons can escape from the surface. Thus for a given value of ϕ, radiation of frequency less than

$$\nu_{\min} = \frac{\phi e}{ch} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.4$$

will not excite any photo-electric emission, c being the velocity of electromagnetic radiation. This conclusion has been verified experimentally for several metals.

At this stage it is necessary to amplify the picture of the atom given on page 9. The electrons surrounding the atom are considered to be arranged in groups or shells according to the energy they possess. The energy of an electron depends on the orbit in which it rotates, some orbits corresponding to higher energy levels than others. Only certain orbits are permissible and to each there is a definite quota of electrons. When this quota has been established for one orbit new elements are obtained by the addition of electrons to the next shell until this too is filled and so on. The first two electrons have orbits nearest to the nucleus and, in the heavier elements, are followed by two groups of eight electrons, two of eighteen, and one of thirty-two. Thus Uranium ($Z = 92$) has its ninety-two electrons arranged as follows :—

first (K) shell	2
second (L) shell	8
third (M) shell	18
fourth (N) shell	32
fifth (O) shell	18
sixth (P) shell	12

The remaining two electrons are in an uncompleted shell.

Energy levels. To move an electron from an inner orbit to one further out, or to detach it from the atom altogether, requires the expenditure of energy and, the nearer the electron is to the nucleus, the greater the amount of energy required to move it further out. Each shell of electrons thus has an energy value associated with it, those nearest to the outside of the atom having the highest energy.

In a metal, the atoms are tightly packed in a regular array and are, indeed, so close to one another that their outer electrons may be regarded as belonging equally to several neighbouring atoms. These are the conduction electrons which may be made to move through the material, in the space between the atoms, by the application of an electric field or a temperature difference, thus producing electrical or thermal conduction. In metals there is one conduction electron (sometimes two or three) per atom; in non-conducting materials there are normally no electrons in the conduction band.

The photo-electric effect takes two forms.

(1) When, in metals and other conducting substances, conduction electrons occupying the highest energy levels are ejected, this is the *outer photo-electric effect.*

(2) Non-conductors may display photo-electricity in a rather different form. Electrons may be moved from low energy levels to higher energy levels, and, for example into the conduction band. This process increases the conductivity of the material but does not lead to emission and is known as the *inner photo-electric effect.* Devices employing the photo-electric effect, which will be described in chapter X, may be classified as follows.

(1) Those using the outer photo-electric effect are known as *emission cells.* The emission cell is essentially a diode valve in which the cathode has a surface sensitive to light from which electrons are emitted in the presence of radiation of suitable frequency. A metal anode kept at a potential higher than that of the cathode attracts these electrons which set up a current in the external circuit when they return to the cathode. The anode is usually a ring or grid so that light can pass through it to reach the cathode, and a potential of about 100 volts is usually applied to it. The sensitivity of an emission cell is greatly

needed to accelerate the electrons sufficiently, the tube is operated under conditions of voltage saturation (page 25). The supply of electrons is normally obtained from a directly heated cathode of tungsten wire. To obtain an intense beam of X-rays from the target it is necessary to concentrate the electron emission to some extent, and for this purpose a cylindrical shield is built round the cathode.

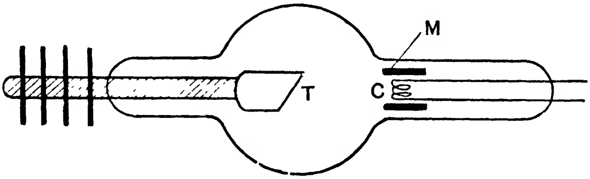


Fig. 3/8. In this diagrammatic representation of a simple X-ray tube, C is the cathode and T the anode. A metallic cylinder, M, surrounds the cathode and is connected to one end of it. This shield serves to concentrate the electron emission into a narrow beam.–(From Noakes: Text-book of Electricity and Magnetism, by courtesy of Macmillan and Company, Ltd.)

One end of the filament is joined to the negative terminal of the high tension supply and to the shield, which is thus at a negative potential with respect to the whole of the cathode and repels any electrons which approach it. The anode, which is also the target, consists of a substantial block of tungsten or some other metal with a high melting point. Most of the kinetic energy given up by the electrons on impact is changed into heat, and it is therefore necessary for the anode to be a good conductor and to have a large thermal capacity.

The anode current of a moderately large X-ray tube, such as is used extensively in medical practice, may be 15 mA at 200 Kv, though tubes operating at anode voltages of two million volts have been built. The high tension is usually supplied by valve rectifiers, though an unrectified supply may be used. In this case the tube acts as a rectifier and X-rays are produced in alternate half cycles.

In the simplest arrangement, the tube is connected in series with a diode valve and the secondary winding of a transformer, from which current is drawn for half of each cycle. As will be explained below, the nature and properties of the radiation emitted depend largely on the anode voltage, and it is necessary in many cases to keep a constant potential difference across the tube while X-ray emission is in progress. When half-wave rectification is employed, only the radiation produced near the peak of each cycle is useful and it is generally more satisfactory to adopt a constant-potential rectifying system. In full-wave rectification it is a common practice to use four valves, arranged as in fig. 3/9. Current flows through two of them during each half cycle and the output voltage may therefore be twice as great as that which a single valve could handle. The rectified potential difference is smoothed by a condenser and the tube voltage has the waveform shown.

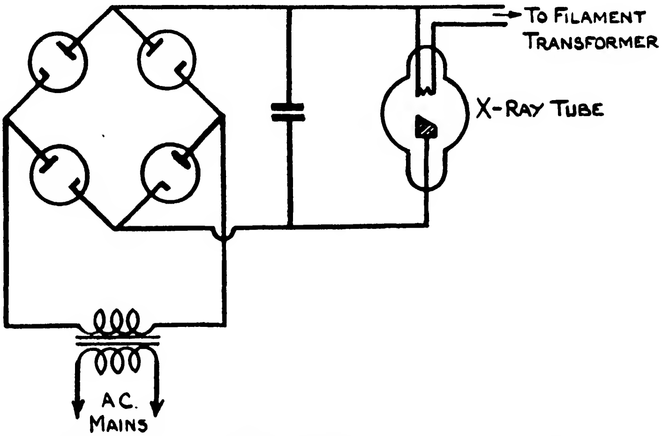


Fig. 3/9. Full-wave rectifying circuit for X-ray tube.

The Greinacher circuit, fig. 3/10, using two valves and two condensers, supplies the tube with a voltage equal to approximately twice the transformer peak output. The operation of this circuit is as follows. During the first half-cycle the condenser, C_1, is charged almost to the peak of the transformer secondary voltage, current flowing along the path $DCBA$. In the next half-cycle C_2 is similarly charged. Then, as C_1, C_2 and the X-ray tube are in series, the potential difference across the tube is approximately twice that of the transformer output. C_1 and C_2 have values such that they lose only a small portion of their charge during the time occupied by one cycle of the supply. Each condenser is charged once in every cycle and discharges through the tube for the remainder of the period. The tube voltage has the waveform shown in fig. 3/11.

Anode Design. The problem of anode cooling is an important factor in the design of X-ray tubes. Since an intense beam of radiation is desirable in most applications, only a small area of the target should be bombarded and it is therefore necessary to concentrate the electron stream into a fine pencil.

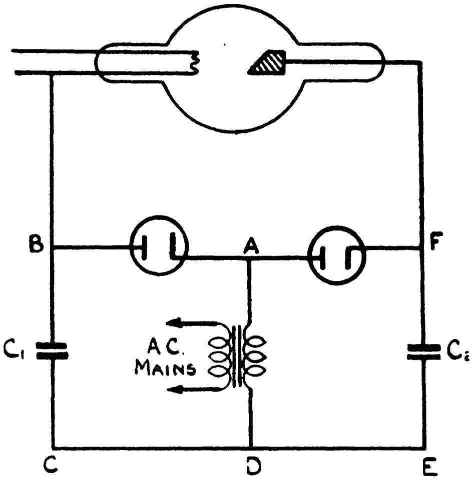


Fig. 3/10. The Greinacher circuit for high tension supply to an X-ray tube. A rectified potential difference equal to twice that of the transformer secondary voltage is obtained.

The focal spot, that is, the portion of the target which is heated by electron impact, is raised to a high temperature and the power which can be released as X-radiation is limited by the possibility of melting the anode. In tubes of moderate power rating the target may be a block of tungsten embedded in a large copper cylinder, with air or water cooling. If greater power is required the problem becomes more acute, for the temperature of the target can be kept down only by using a broader pencil of electrons and if this is done the tube is of little use for photographic work, as the sharpness of definition is impaired. The difficulty may be resolved by the use of a rotating anode (fig. 3/12). Here the target is a frustrum of a cone and is rotated by an induction motor with one winding outside the tube and one

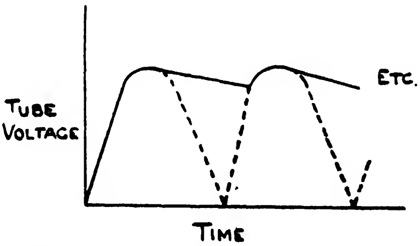


Fig. 3/11. Output waveform of Greinacher circuit. The charge on the condensers is replenished at the peaks of the transformer output voltage and they discharge slowly during the rest of the cycle, keeping a fairly steady potential difference across the tube.

inside. Thus the focal spot is not in one place, and the heat generated is distributed over a wide area, so that a large amount of X-radiation may be produced without overheating the anode.

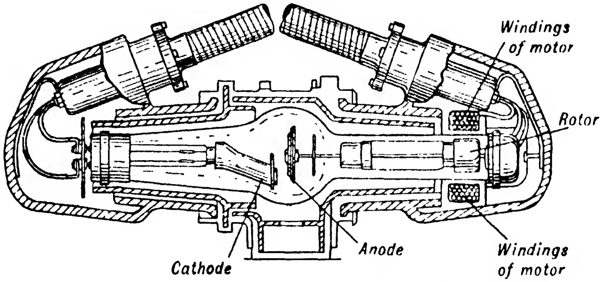


Fig. 3/12. Rotating anode X-ray tube—(From Noakes: Text-book of Electricity and Magnetism, by courtesy of Macmillan and Company, Ltd.)

The remarks on page 58 with reference to high vacuum diodes apply equally to X-ray tubes, which are commonly made of metal, a small window of some transparent material permitting the radiation to escape. The metal casing absorbs radiation which does not leave by the window and thus affords protection to persons operating the tube from the harmful effects of the radiation. The tube is often enclosed in another metal case, which is earthed, thus eliminating the danger of electric shock from the high tension system. The space between the tube and the outer casing is usually filled with oil which assists in electrical insulation and in the conduction of heat from the tube.

Theory of X-ray emission. The X-rays emitted from tubes of kind described are not monochromatic—that is to say, they are not confined to a single wavelength. The radiation has two constituents :

1. *Continuous radiation,* in which the energy is distributed continuously over a broad band of wavelengths, with a sharply defined lower limit. The minimum wavelength produced is found to be inversely proportional to the potential difference across the tube. At the long wavelength end, the intensity of the radiation falls off gradually. This component of the X-radiation may be compared to the continuous optical spectrum produced by white light, and it depends only to a small extent on the target substance.

2. *Characteristic radiation.* This consists of a number of separate wavelengths, all greater than the lower limit mentioned above. The wavelengths present depend on the target substance and this component of the emission may be compared to a line spectrum in optics.

X-ray emission is explained as follows :—

1. The kinetic energy of an electron striking the target is in general changed partly into heat and partly into X-radiation.

2. The kinetic energy of each electron can be changed into one quantum of radiation with frequency ν given by

$$\tfrac{1}{2}mv^2 = h\nu \quad (cf.\ 3.2)$$

If all of the kinetic energy is changed in this way, the frequency of the radiation is

$$\nu_{\max} = \frac{mv^2}{2h} \qquad \text{.........} 3.4$$

Now, a particle of charge e accelerated through a potential difference of X acquires kinetic energy equal to Xe and the maximum frequency of the radiation emitted from a tube with anode voltage V is therefore

$$\nu_{\max} = \frac{Xe}{h} \qquad \text{.........} 3.5$$

The corresponding minimum wavelength is

$$\lambda_{\min} = \frac{ch}{Xe} \qquad \text{.........} 3.6$$

In general, some of the energy appears as heat and the frequency of the X-radiation can have any value up to $\nu_{\max}$. This is the origin of the continuous radiation.

3. The energy of an incident electron may be used to ionise one atom of the target substance by removing an electron from one of the inner shells. The vacant space is filled by an electron falling from an outer orbit and this is in turn replaced by one from another shell still further out. Thus the ionisation produced by one electron impact sets up a chain of transitions from one shell to another. In each jump, the electron loses energy which is released as radiation with frequency given by

$$\nu = \frac{E_1 - E_2}{h} \qquad \text{.........} 3.7$$

where E_1 and E_2 are the initial and final energies of the electron making the jump.

The changes which may occur are shown in fig. 3/13. Electrons are arranged round the nucleus in shells, labelled *K,L,M,N,* and so on. A *K* electron is first removed by ionisation. Its place may then be filled by an electron from any of the outer shells, and if a number of atoms are involved, wavelengths corresponding to transitions from *L* to *K*, *M* to *K* and so on are all emitted. These constitute the *K* wavelengths of the characteristic radiation. If an electron from the *L* shell falls into the *K* level, its place may be filled from the *M, N,* or outer shells, and a set of *L* lines is produced. The characteristic radiation thus consists of a number of wavelengths, corresponding to the *K,L,M,*. . . shells.

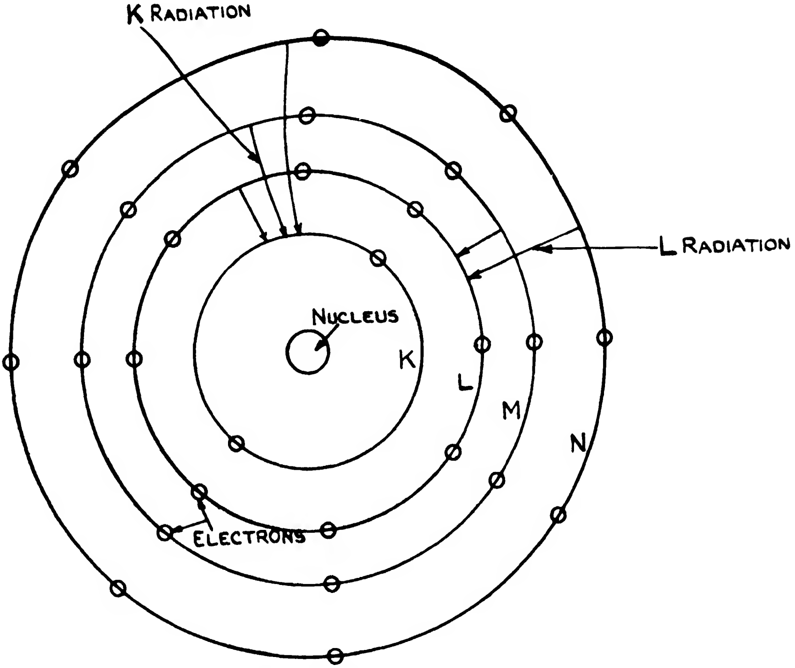


Fig. 3/13. Diagrammatic representation of the arrangement of electrons in a heavy atom, showing the electron transitions resulting in the various lines of the X-ray emission spectrum.

CHAPTER IV

ALTERNATING CURRENT THEORY

The behaviour of steady currents in conductors depends on certain simple laws with which the reader is assumed to be familiar. In this book we shall be concerned frequently with currents varying both in magnitude and in direction. Before studying the properties of such currents, we shall recall the laws of electromagnetic induction.

Electromagnetic induction. A conductor through which a current is flowing behaves like a magnet, that is to say, lines of magnetic force are set up by the passage of currents. Conversely, a variation of the magnetic field near a conductor induces a flow of current in it. The first law of electromagnetic induction states that:

Any change in the number of lines of flux cutting a circuit causes an electromotive force to be established in it.

The second law states that:

The magnitude of the induced electromotive force is proportional to the rate of change of the number of lines of flux cutting the circuit.

The third law, usually known as Lenz's law, states that:

The direction of the induced electromotive force is always such as to oppose the change which causes it.

These three laws may be expressed in the single relation

$$E = -\frac{dN}{dt} \qquad \text{..........................4.1}$$

where E = induced e.m.f., N = number of lines of flux cutting the circuit and t = time. (E will be in electromagnetic units if N is in lines. 10^8 e.m.u. = 1 volt.)

Consider a rectangular coil of wire rotating in a uniform magnetic field. Let p = the angular velocity of the coil in radians per second, and N_0 = the number of lines of flux cutting the coil when it is in the vertical position (fig. 4/1). After a time t, the coil has moved through an angle pt radians, and the number of lines passing through it is

$$N = N_0 \cos pt \qquad \ldots\ldots\ldots\ldots 4.2$$

As the coil rotates N changes continuously and therefore an *e.m.f. is induced in the coil as long as it continues to rotate.*

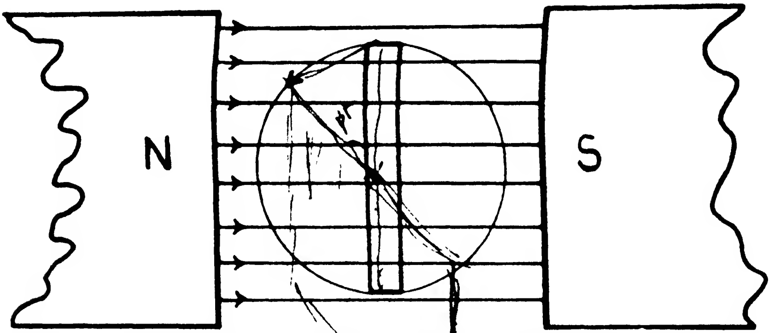


Fig. 4/1. Generation of an alternating current. A coil is rotated between the poles of a magnet. As the coil moves, the number of lines of magnetic flux through it rises and falls resulting in the appearance of an induced e.m.f.

The magnitude ot this e.m.f. is, from equation 4.1 above

$$E = -\frac{dN}{dT} = pN_0 \sin pt \qquad \ldots\ldots\ldots\ldots 4.3$$

This e.m.f. has a maximum value pN_0 and calling this E_0, write

$$E = E_0 \sin pt \qquad \ldots\ldots\ldots\ldots 4.4$$

A graph of this function (fig. 4/2) shows that E changes its magnitude continually and its direction at intervals of pt equal to π radians, that is, in every half revolution of the coil. The portion AB of the alternation after which it repeats itself, is called a cycle and the number of cycles executed in a second of the frequency. It will be observed that p is 2π times the frequency ; this quantity is sometimes called the ***pulsatance*** of the alternation.

Alternating e.m.f. in a circuit containing resistance. If an e.m.f. $E_0 \sin pt$ is applied to a circuit containing resistance R the current flowing is simply

$$\frac{E_0 \sin pt}{R} \cdot$$

Since the maximum value I_0 of this current is E_0/R, we may write

$$I = I_0 \sin pt \qquad \ldots\ldots\ldots\ldots 4.5$$

In this circuit the current and e.m.f. have maxima and minima at the same instants and are in step or in phase throughout each cycle (fig. 4/3).

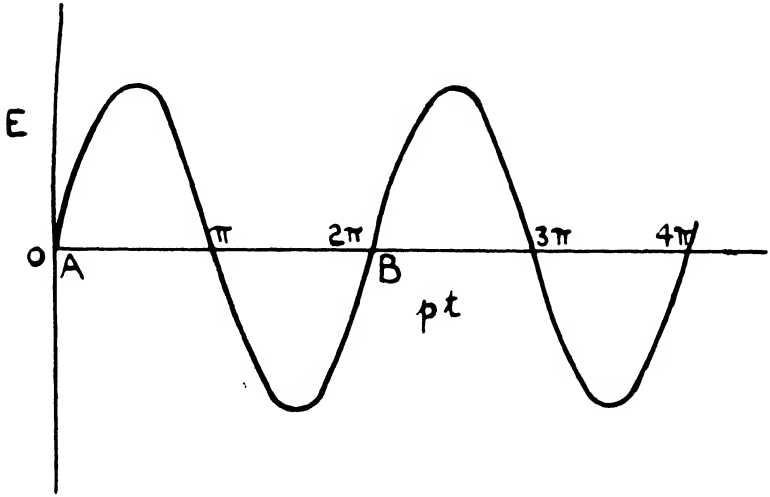


Fig. 4/2. Waveform of a sinusoidal alternating e.m.f. The alternation repeats itself after a period T given by $pT = 2\pi$.

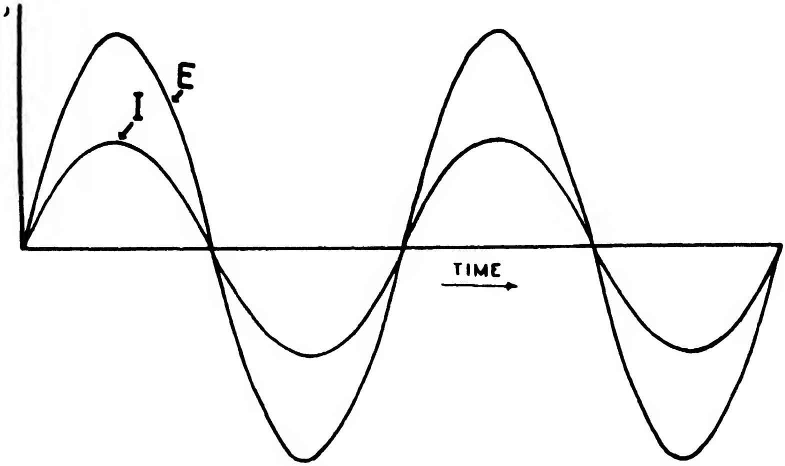


Fig. 4/3. In a circuit containing only resistance, the current and the e.m.f. are always in phase, passing through maxima and minima at the same times.

The average value of a sinusoidal current, taken over a whole number of cycles is obviously zero. Several important effects, however, depend on the square of the current and, as a graph of I^2 against t lies wholly on the positive side of the axis, the mean value of I^2 is not zero; it may be found as follows, referring to fig. 4/4.

$$\overline{I^2} = \frac{\text{Area under } I^2 \text{ curve in one cycle}}{\text{time occupied by one cycle}}$$

$$= \frac{I}{T}\int_0^T I^2\,dt$$

$$(T = \text{time occupied by one cycle})$$

$$= \frac{I}{T}\int_0^T I_0^2 \sin^2 pt\,dt$$

$$= \frac{I}{pT}\int_0^{2\pi} I_0{}^2 \sin^2 pt\,d\,(pt)$$

$$= \frac{I}{2\pi}\int_0^{2\pi} I_0^2\,\frac{1-\cos 2pt}{2}\,d\,(pt)$$

$$\left(\text{since } T = \frac{2\pi}{p}\right)$$

$$= \frac{I_0^2}{2\pi}\left[\frac{pt - \frac{1}{2}\sin 2pt}{2}\right]_0^{2\pi}$$

$$= \frac{I_0^2}{2\pi}\,.\,\pi$$

$$\overline{I^2} = \frac{I_0^2}{2}$$

The root mean square or (R.M.S.) value of the current is the square root of $\overline{I^2}$ and $\overline{I} = I_0/\sqrt{2}$ 4.6

$\overline{I}$ is the direct current which would, for example, produce the same heating effect as the alternating current $I_0 \sin pt$.

The power dissipated by a direct current I flowing through a resistance R is I^2R ; in an alternating current circuit, the power is $\overline{I^2}R$ which is equal to $\overline{IE}$ and to $\overline{E^2}/R$. Thus a circuit containing an alternating e.m.f. in series with a resistance behaves in the same way as a direct current circuit with current $\overline{I}$ and e.m.f. $\overline{E}$. Most alternating current circuits, unfortunately, do not behave as simply as this because every conductor has, in addition to resistance, the properties of inductance and capacitance, which do not affect the passage of direct currents, but are important where alternating currents are concerned.

Inductance. When a current flows in a conductor, lines of magnetic flux are set up around it. If N lines are produced around a conductor by the passage of a current I, the inductance of the conductor is defined by the relation

$$N = LI \qquad \text{.........................4.7}$$

L is the *coefficient of self-induction* or the inductance of the conductor. (The practical unit of inductance is the Henry, which is equal to 10^8 lines per ampere.) The inductance of a system containing one or more conductors depends on their size, shape and relative position. Thus the inductance of a straight length of wire is less than that of the same material wound into a coil. The coil is in fact the most satisfactory arrangement for obtaining inductance and except for very small inductances is invariably used for this purpose.

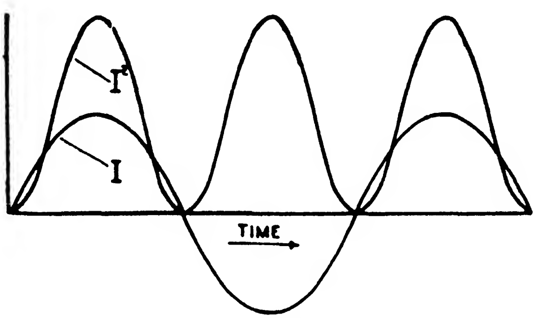


Fig. 4/4. The mean value of a sinusoidally alternating current is zero, but I^2 is always positive and its mean value can be evaluated by an integration.

Alternating current in an inductive circuit. The behaviour of an inductance in an alternating current circuit may be examined by finding the potential difference across a coil of inductance L when a current $I = I_0 \sin pt$ flows through it. The back e.m.f. set up by electromagnetic induction is equal to

$$-\frac{dN}{dt}$$

and if the potential difference of the supply is E, then the net e.m.f. available for driving current through the coil is

$$E - \frac{dN}{dt} \quad \text{or } E - L\frac{dI}{dt}$$

In general the current flowing is related to the supply voltage by the relation

$$E - L\frac{dI}{dt} = RI \qquad \ldots\ldots 4.8$$

where R is the resistance of the coil. For the present we consider the case in which R is negligible, so that the potential difference across the coil must be enough to balance the induced e.m.f. but no larger. Then

$$\begin{aligned} E &= L\frac{dI}{dt} \\ &= LpI_0 \cos pt \\ &= LpI_0 \sin (pt + 90^\circ) \qquad \ldots\ldots 4.9 \end{aligned}$$

In this case the phase of E is 90°, or one-quarter of a cycle, ahead of the phase of I.

Reactance. From above.

$$E_0 = LpI_0 \qquad \ldots\ldots 4.10$$

Comparing this with the corresponding expression $E = RI$ for a direct current we see that E_0 and I_0 are connected by a relation similar to Ohm's Law and that the ratio between them is Lp. This is the *reactance* of the coil, and is measured in ohms. It should be noted that the maximum value of E does not occur at the same time as the maximum value of I, so that the resemblance to Ohm's Law is only superficial. From the preceding paragraph it follows also that $\overline{E} = Lp\overline{I}$.

If the current flowing through an inductance L is $I_0 \sin pt$, the potential drop across it is $LpI_0 \sin pt$ and the rate of dissipation of energy at time t – that is, the power of the circuit is

$$LpI_0^2 \sin pt \cos pt$$

The average power during a whole number of cycles is therefore

$$\begin{aligned} &\frac{I}{T}\int_0^T LpI_0^2 \sin pt . \cos pt . \, dt \\ &= \frac{I}{pT}\int_0^{2\pi} LpI_0^2 \frac{\sin 2pt}{2} \, d(pt) \\ &= \frac{LpI_0^2}{2pT}\left[- \frac{\cos 2pt}{4} \right]_0^{2\pi} \\ &= 0 \end{aligned}$$

This curious result is a consequence of the 90° phase difference between E and I. Physically it means that the circuit absorbs energy from the supply during one half cycle and gives it up during the next. In practice an inductance coil has resistance also and some energy is dissipated, the power being I^2R ; in general this is much smaller than EI. An inductance may thus be used to regulate the current in an A.C. circuit without the expenditure of energy which is necessary if a resistance is placed in the circuit. Coils used in this way are often referred to as *chokes*.

Capacitance. The capacitance of a body may be regarded as its ability to store electric charge and is defined as follows :—

If a charge Q raises the potential of a body relative to its surroundings by an amount V then $C = Q/V$ where C is the capacitance of the body.

Any arrangement of conductors possesses inductance, capacitance and resistance ; one in which capacitance predominates is called a condenser and in the simplest form consists of two metal plates separated by air or some other insulating medium. (The units are farads, microfarads (μF) or micro-microfarads ($\mu\mu F$)). If a condenser of this kind is connected to a battery as in fig. 4/5 electrons flow from the negative pole to the plate B —and no further. At the same time the accumulation of electrons on B is compensated by a withdrawal of electrons from A, which therefore becomes positively charged, while B acquires a negative charge. Once the initial surge of current necessary for these operations has occurred there can be no further movement of electrons. Suppose now that the battery is replaced by a generator of alternating current. If at first the current flows counter-clockwise, the plate A takes on a positive charge and B becomes negative—that is, electrons flow from A to the generator and from the generator to B. In the next half cycle the e.m.f. is reversed and the electrons gathered on B flow back to A through the generator. As this process continues, electrons flow from B to A and back again in each cycle. Thus although no electronic current flows in the insulating material between the plates, electrons move

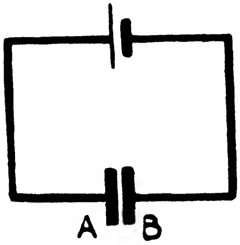


Fig. 4/5.

everywhere else in the circuit and a recording instrument placed at any point would record the passage of an alternating current. To investigate this result mathematically, we have

$$Q = CV \qquad \text{.........................4.11}$$

Now I, the current flowing, is the rate at which charge passes a given point, that is to say

$$I = \frac{dQ}{dt} = \frac{CdV}{dt} \qquad \text{....................4.12}$$

V is the e.m.f. supplied by the generator and is equal to $E_0 \sin pt$, so that

$$\begin{aligned} I &= CpE_0 \cos pt \\ &= CpE_0 \sin (pt + 90^\circ) \qquad \text{...............4.13} \end{aligned}$$

This result is similar to that (cf. 4.9) which was obtained for an inductance, but now I leads E by one quarter of a cycle. The reactance of the condenser is E_0/I_0, that is, $1/Cp$ and is therefore, inversely proportional to the frequency.

Alternating current in a circuit containing inductance and resistance. This is a more practical problem than that involving inductance alone. A coil connected to a source of alternating current may be represented by the circuit of fig. 4/6. Since R and L are in series the same current I must flow through each. The potential difference across R is RI and is in phase with the current, while that across L is LpI and leads I in phase by 90°. The total potential drop across the combination is the sum of these two components, but since the voltages are not in the same phase, they cannot be compounded simply by addition. The solution is most easily obtained if the alternating voltages are regarded as two vectors. Thus in fig. 4/7 the potential difference across R is represented by a line of length RI units pointing in the west-east direction and that across L by a line of length LpI units pointing to the north. (The reader is assumed to be familiar with this method of representing simple harmonic motions: if the vector is rotated anti-clockwise with a period of revolution equal to $2\pi/p$, its projection on a diameter of the circle thus traced out is equal to the value of the alternating

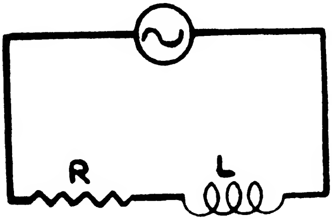


Fig. 4/6

quantity at each instant.) These two vectors may now be compounded by the parallelogram rule, from which it is apparent that the resultant E is $I\sqrt{R^2 + L^2p^2}$. The direction of the resultant makes an angle ϕ with the RI vector, given by $\tan\phi = Lp/R$, and the potential difference across the coil therefore leads the current in phase by ϕ degrees.

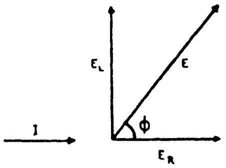
We may now write

$$E = \sqrt{R^2 + L^2p^2}\; I_0 \sin(pt + \phi)$$

It follows that $E_0/I_0 = \sqrt{R^2 + L^2p^2}$ 4.14

This quantity corresponds to resistance in a D.C. circuit and to reactance in a circuit containing inductance or capacitance. It is referred to as the *impedance* of the circuit and is usually represented by the symbol Z; in the circuit under discussion, Z may be regarded as the vector resultant of R and Lp.

Fig. 4/7. Vector diagram for the circuit of fig. 4/6. The potential difference across the resistance is in phase with the current and that across the inductance is 90° ahead. The total potential difference is the resultant of these two and is found by the parallelogram rule for the combination of two vectors. The tangent of the phase angle ϕ is Lp/R.



Circuit containing capacitance and resistance in series. This is an arrangement which is widely used in radio.

(*a*) Direct current. When the switch is closed in the circuit of fig. 4/8 the condenser C becomes charged through the resistance R. The rate of charging depends on the values of C and R in the following way.

$E = E_C + E_R$ 4.15

where E is the potential difference between the terminals of the battery and E_C and E_R are the potential differences across the condenser and the resistance respectively. If the current flowing at any instant is I, then $E_R = RI$ and if at the same instant the charge on the condenser is Q, then $E_C = Q/C$, so that

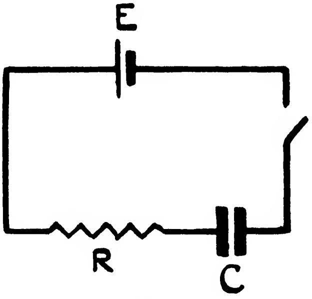


Fig. 4/8

$$E = \frac{Q}{C} + R\frac{dQ}{dt} \qquad \text{......................4.16}$$

Putting EC, which is the ultimate charge on the condenser, equal to Q_0, we have :

$$Q_0 = Q + CR\frac{dQ}{dt} \qquad \text{.........................4.17}$$

and the solution of this differential equation is

$$Q = Q_0\left(1 - e^{-\frac{1}{CR}t}\right) \qquad \text{.........................4.18}$$

Thus after an infinite time, $Q = Q_0$; a graph of the charging process is shown in fig. 4/9.

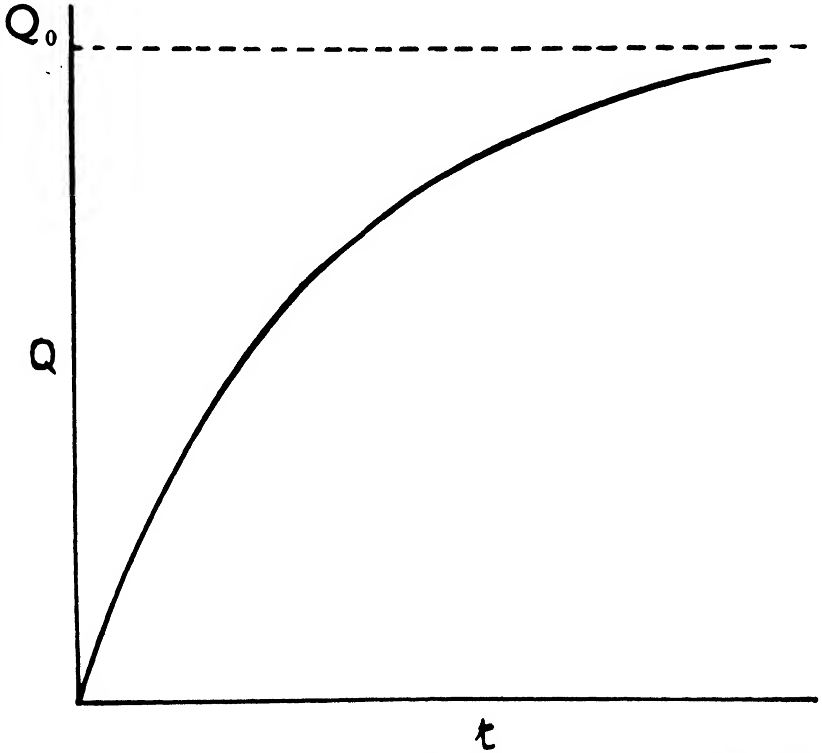


Fig. 4/9. The charge on the condenser in the circuit of fig. 4/8 rises exponentially to its maximum value Q_0.

In the case of a condenser discharging through a resistance, the initial equation is

$$E_C + E_R = 0 \qquad \text{.........................4.19}$$

and the solution eventually obtained is

$$Q = Q_0\, e^{-\frac{1}{CR}t} \qquad \text{.........................4.20}$$

After a time CR seconds, $Q = Q_0 e^{-1}$ and the charge has fallen to $1/e$ of its initial value, that is, the condenser has lost about 63 per cent of its original charge. In the case previously considered, CR seconds is the time taken to charge up to 63 per cent. of the ultimate value. This product therefore indicates roughly the speed at which a condenser will charge or discharge through a resistance ; it is called the *time constant* of the circuit.

(*b*) Alternating current supply.

Now we have

$$E_C + E_R = E_0 \sin pt \qquad \ldots\ldots\ldots\ldots\ldots\ldots 4.21$$

$$\frac{Q}{C} + R\frac{dQ}{dt} = E_0 \sin pt \qquad \ldots\ldots\ldots\ldots\ldots\ldots 4.22$$

$$\frac{dI}{dt} + CR\frac{d^2I}{dt^2} = CpE_0 \sin pt \qquad \ldots\ldots\ldots\ldots\ldots\ldots 4.23$$

The solution of this problem is more simply performed by using the vector method, illustrated by fig. 4/10.

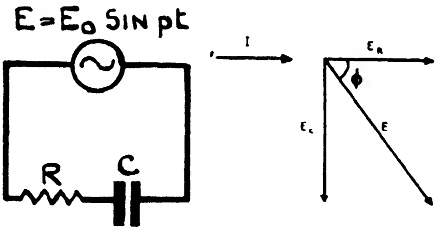


Fig. 4/10. In a series circuit consisting of resistance and capacitance, E_C is 90° behind E_R and the phase angle between the resultant potential difference and the current is given by $\tan\phi = I/CpR$.

✓**Circuit containing capacitance, inductance and resistance in series.** The vector method is appropriate here and as usual we take I for the standard vector, fig. 4/11. E_L and E_C are out of phase by 180° and may be replaced by a single voltage E_{LC} which leads or lags behind I by 90° according as $E_L >$ or $< E_C$. L and C together contribute a reactance $X = Lp \backsim 1/Cp$.

The total impedance of the circuit is given by

$$Z^2 = R^2 + \left(Lp \backsim \frac{1}{Cp}\right)^2 \qquad \ldots\ldots\ldots\ldots\ldots\ldots 4.24$$

$$\text{and } \tan\phi = \frac{Lp \backsim \dfrac{1}{cp}}{R} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 4.25$$

For given values of L and C there is one frequency, f_0 for which the reactance vanishes. This frequency is given by

$$L\,p_0 = \frac{1}{Cp_0} \qquad \text{..............................4.26}$$

$$\text{i.e. } f_0 = \frac{1}{2\pi\sqrt{LC}} \qquad \text{...................4.27}$$

A circuit containing resistance inductance and capacitance is said to be resonant when its reactance vanishes or, in other words, when the current flowing through it is in phase with the potential difference across it.

Properties of the series resonant circuit. If an electromotive force of constant amplitude and variable frequency is applied to the circuit of fig. 4/11, then, at resonance,

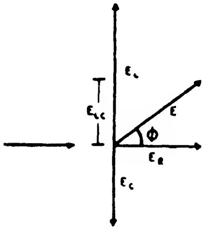
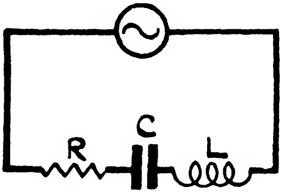


Fig. 4/11. In a series circuit of L, C and R the vector diagram has the form shown by this diagram. E_C and E_L are separated by a phase difference of 180° and may be reduced to a single vector which is then combined with E_R in the usual way.



(1) Impedance $= R$. This is the smallest value which Z can have and the current flowing in the circuit is therefore greatest at the resonant frequency.

(2)

$$E_L = LpI = E_C \qquad \text{........................4.28}$$

$$E_R = \text{supply e.m.f.} = E \qquad \text{..................4.29}$$

$$\frac{E_L}{E} = \frac{E_C}{E} = \frac{Lp}{R} = \frac{1}{CpR} \qquad \text{................4.30}$$

In any practical case the ratio Lp/R is usually quite large and the potential difference developed across L or C is therefore greater than the e.m.f. of the supply. The ratio $Lp/R = 1/CpR$ is called the 'Q' of the circuit.

Resonance curves are obtained for a series circuit by plotting the current I against the frequency, using the relation

$$I = \frac{E}{\sqrt{R^2 + \left(Lp - \frac{1}{Cp}\right)^2}} \qquad \text{..............4.31}$$

The graph of this function is of the shape shown in fig. 4/12 from which it will be seen that there are two frequencies corresponding to any given value of I—one on each side of the resonant frequency. It may be shown that if f_1 and f_2 are the frequencies at which the current is $1/\sqrt{2}$ of its value at resonance, then

$$\frac{f_0}{f_2 - f_1} = Q \qquad \dots\dots\dots\dots\dots\dots 4.32$$

Thus if Q is large the current falls away rapidly and the impedance therefore rises on each side of the resonant frequency, giving a sharply peaked curve, while if Q is small the resonance curve is fairly flat and the impedance changes little as the frequency alters.

The series LCR circuit offers a path of low impedance at its resonant frequency and has higher impedance at all other frequencies. On account of this property it may be used to separate one frequency from several others and is called an *acceptor circuit.*

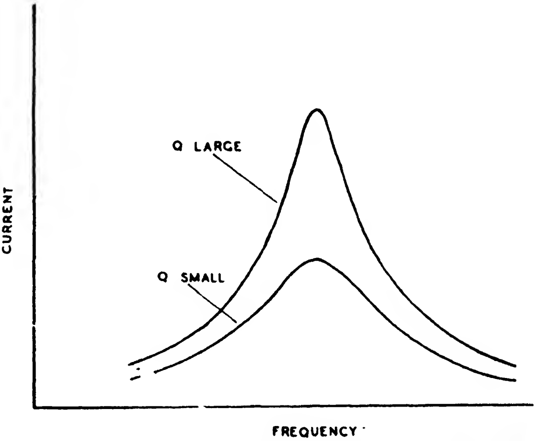


Fig. 4/12. In an acceptor circuit the reasonance curve is sharply-peaked when Q is large.

Circuit containing Resistance and Inductance in parallel with Capacitance. Another important circuit is that shown in fig. 4/13 and is obtained in practice by connecting a coil across a condenser. The vector diagram is *also* shown in fig. 4/13. Since the same difference of potential exists across both branches E is taken as the reference vector and we find the vector sum of the two currents.

The total current consists of two components :

(i) An in-phase component equal to $I_{LR} \cos \phi$

(*ii*) A reactive or out-of-phase component equal to

$$ECp - I_{LR} \sin \phi$$

Analysis of this circuit for the general case is tedious and not particularly profitable and we shall confine our attention to its properties at resonance. This state occurs when the reactance is zero, that is, when the current flowing in the circuit is in phase with the potential difference across it. For the reactance to be zero

$$ECp = \frac{E}{\sqrt{R^2 + L^2p^2}} \; sin \; \phi$$

$$= \frac{ELp}{R^2 + L^2p^2}$$

$$\left(\text{since } sin \; \phi = \frac{Lp}{\sqrt{R^2 + L^2p^2}}\right)$$

$$\therefore \; p^2 = \frac{1}{LC} - \frac{R^2}{L^2} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 4.33$$

The second term in this expression is usually much smaller than the first and may be neglected without serious error. In practice therefore the parallel circuit resonates at the same frequency as a series circuit having the same components.

The total current at resonance is

$$\frac{E \; cos \; \phi}{\sqrt{R^2 + L^2p^2}}$$

Substituting for *cos* ϕ and p^2 reduces this to ECR/L and the impedance is therefore L/CR. It will be observed that the current drawn from the supply at resonance is less than the current flowing in either branch of the circuit and it can be shown that the supply current is a minimum almost at resonance. For practical purposes, since ϕ is nearly 90° the difference between the resonant frequency and the frequency for minimum current is negligible. The graph of Z against p has the same form as that of I against p for the series circuit. When ϕ is almost 90° the currents in the two branches are approximately equal and approximately 180° out of phase. The ratio of either current to the total current is Q.

The currents flowing in the two branches are equivalent to a single circulating current QI flowing through L, C and R in series, as indicated by the arrow in fig. 4/13 and the circuit may be regarded as giving current magnification corresponding to the voltage magnification of the series circuit. In most of its applications the parallel circuit is used as an impedance which has a maximum value at the resonant frequency, and is then called a *rejector circuit.*

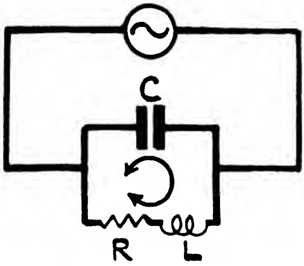
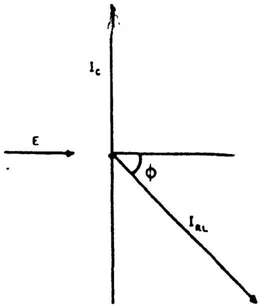


Fig. 4/13. Vector diagram for the parallel resonant circuit. $I_C = ECp$ & $I_{RL} = \frac{E}{\sqrt{R^2+L^2p^2}}$



Coupled circuits. It is often necessary to transfer energy from one circuit to another. In direct current working this can be done efficiently by direct connection but for alternating currents particularly at high frequencies the transfer must be made indirectly. Two circuits are said to be coupled if energy can be exchanged between them and this coupling necessitates the presence of an impedance common to both circuits. This impedance may be made up of resistance, inductance or capacitance, but is most often a mutual inductance (fig. 4/14). Mutual

CAPACITIVE COUPLING INDUCTIVE COUPLING

RESISTIVE COUPLING COUPLING BY MUTUAL INDUCTANCE

Fig. 4/14. Various methods of coupling between two circuits.

inductance is a circuit element resembling self inductance in many of its properties, and arises in the following way. If an alternating current is sent through a coil A an alternating magnetic field is established around the coil. The lines of force so produced pass through any coil B placed near A and an e.m.f. will therefore be established in the circuit containing B. The magnitude of this e.m.f. depends on the frequency and number of lines set up by the current in A and on the proportion of these lines which cut B. In other words the e.m.f. induced in the secondary circuit (B) depends on the current flowing in the primary circuit (A), and on the relative position of the two circuits. A mutual inductance exists between any two circuits so placed that the lines of force produced around one can intersect the other. The e.m.f. induced by a current I across a mutual inductance M is given by

$$E = MpI \qquad \text{.........................4.34}$$

If this equation is compared with the corresponding result for self inductance, ($E = LpI$, cf. 4.10), it is seen that M has the same dimensions as L and may be measured in the same units. The degree of coupling between two circuits of self inductances L_1 and L_2 is expressed by a *coefficient of coupling* k defined by the relation

$$k = \frac{M}{\sqrt{L_1 L_2}} \qquad \text{.........................4.35}$$

M being the mutual inductance ; the greatest value k can have is unity when all of the lines of force produced by one coil pass through the other. In audio-frequency transformers, this value may be approached, but in radio-frequency circuits when the primary and secondary are usually at or near resonance, the coefficient is normally very much less than unity, for the following reasons.

Referring to eq. 4.34, it is apparent that the e.m.f. induced in the secondary circuit is equivalent to a generator of output MpI_1 which sets up a current $I_2 = MpI_1/Z_2$.

The lines of flux produced by the secondary current cut the primary circuit and produce in it an e.m.f. of value MpI_2, which is equal to

$$\frac{M^2p^2I_1}{Z_2}$$

The generator in the primary circuit must overcome this e.m.f. which, as might be expected from Lenz's Law, is in opposition to the initial primary e.m.f., and must also produce a potential difference Z_1I_1 across the impedance of the primary circuit. The primary current is therefore given by

$$E_1 = Z_1I_1 + \frac{M^2p^2I_1}{Z_2} \qquad \text{.......................4.36}$$

and is smaller than it would be if no secondary circuit were coupled to the primary. The e.m.f. induced in the primary by the secondary current has the same effect as an additional impedance. M^2p^2/Z_2; this is called the *reflected impedance.* As Z_2 has its minimum value at the resonant frequency of the secondary circuit the reflected impedance is greatest at that frequency and the resonance curve of the primary becomes flatter.

We shall consider briefly what happens when two identical circuits of L, C and R, one containing a generator of variable frequency and constant e.m.f. are coupled by mutual inductance.

1. Coupling coefficient very small. The secondary e.m.f. and therefore the secondary current are small at all frequencies. The reflected impedance is negligible and both circuits exhibit the normal resonance phenomena.

2. Coupling increased. The secondary e.m.f. becomes greater as M increases and the secondary current therefore rises. The reflected impedance increases and since it is greatest near resonance the primary resonance curve becomes flattened.

3. Coupling increased further. As this process continues the primary current near resonance becomes noticeably smaller and the secondary e.m.f. therefore diminishes producing a reduction in the secondary current. Thus if the coupling coefficient is increased beyond a certain point the primary and secondary currents at the resonant frequency are both reduced. The value ot k for which the secondary current at resonance has a maximum value is the coefficient of critical coupling and is equal to $1/Q$. If k is less than this value the secondary current is reduced because the secondary e.m.f. is smaller and if k is larger than $k_{\text{crit.}}$ the reflected impedance becomes great enough to reduce the primary current and therefore the secondary current and e.m.f. are again diminished.

4. Coupling greater than the critical value. It may be shown that in this case the primary circuit exhibits resonance at two frequencies, one on each side of its normal resonance point, and the secondary current, which depends on the primary current also has two maximum values. These changes are shown in fig. 4/15. In radio frequency work, coupled circuits are seldom used with k greater than $k_{\text{crit.}}$ At critical coupling the secondary current-frequency curve begins to flatten near the peak and for coupling slightly greater than this the top of the curve becomes flat over a small range of frequencies. This characteristic is used when it is desired to select a band of frequencies rather than a single frequency and coupled circuits so adjusted are spoken of as a *band pass* arrangement.

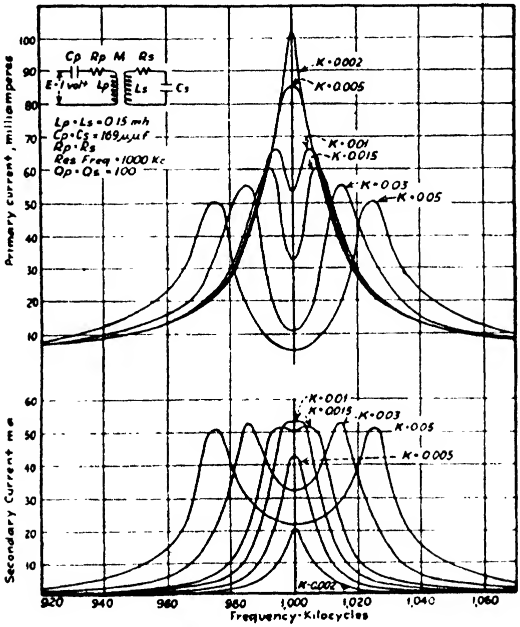


Fig. 4/15. Resonance curves for coupled circuits. (From Terman: Radio Engineers' Handbook, by courtesy of McGraw-Hill Publishing Company).

CHAPTER V

VALVE RECTIFIERS

Types of rectifiers. A rectifier is a device so constructed that an alternating potential difference applied between its ends causes a unidirectional current to flow through it. We have already seen how a diode may be used in this way (page 21). Though the basic theory of valve rectification is simple, many interesting problems arise in its practical applications. Diode rectifiers are of two types ;

(1) high vacuum valves in which the pressure is less than 10^{-6} mm. of mercury and

(2) gas-filled tubes, which contain helium, argon or mercury vapour at pressures ranging from 0·05 to 0·001 mm.

Before discussing the characteristics of these valves we shall consider the behaviour of a diode rectifier in general terms, having applications to both high-vacuum and to gas-filled tubes.

Rectification. The simplest circuit for diode rectification is that of fig. 5/1. During one half of each cycle the anode of the valve is negative and no current flows ; the valve may then be regarded as an infinite resistance. As the load offers a constant finite resistance to the current, the whole voltage of the generator is applied across the valve during this half cycle. In the other half cycle when the anode is positive, the valve offers only a small resistance and a portion of the total e.m.f. appears as a voltage drop across it, most of the generator voltage being now applied to the load. The potential drop across the valve is responsible for accelerating the electrons to the anode, thus maintaining the current. During the positive half cycle electrons arrive at the anode with a velocity v given by

$$Xe = \tfrac{1}{2}mv^2 \qquad \text{.............................5.1}$$

where e is the charge on each electron, m is the mass of an electron and X is the potential difference across the valve, which varies as the generator voltage rises and falls.

Characteristics of high vacuum rectifiers. During the negative half cycle, when the cathode is positive with regard to the anode, the valve offers an infinite impedance and the whole of the generator e.m.f. is therefore set up across it. There is then a danger of arcing between the electrodes, which does not arise during the positive half cycle. The maximum peak voltage which can be applied between cathode and anode without causing a discharge to occur is determined mainly by the distance between the electrodes and is called the maximum peak inverse voltage.

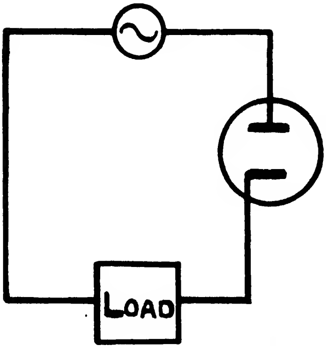


Fig. 5/1. If an alternating e.m.f. is applied to a diode valve, current flows in the half-cycles during which the anode of the valve is positive.

Some of the electrons emitted by the cathode may travel to the walls of the tube instead of to the anode. The negative charge so produced on the walls reduces the current and has the effect of increasing the impedance of the valve by increasing the voltage drop across it. This *wall effect* may be overcome by having the cathode in the centre of a cylindrical anode which may, in fact, form the outer wall of the valve. This construction gives a greater current, for no electrons are lost to the anode, and facilitates cooling, as the anode is exposed directly to the air (Frontispiece).

Bombardment of the anode by the electrons may produce the following effects :

(*a*) heating. Most of the kinetic energy of the electrons is changed into heat and the temperature of the anode therefore rises.

(*b*) thermionic emission. If the temperature rises to a very high value thermionic emission may take place from the surface of the anode.

(*c*) secondary emission. Some of the energy of the electrons arriving at the anode is used to eject other electrons.

(*d*) X-radiation. Retardation of the electrons which strike the anode may lead to the emission of X-rays—electromagnetic radiation of short wavelength (see page 32).

In a diode, effects (*b*) and (*c*) above are not very important, for electrons emitted by the anode are immediately attracted back to it by the positive charge. The heating effect and the X-ray effect can both be reduced by keeping the potential difference across the valve at a low value, that is, by designing the valve so that it offers a low resistance to the passage of current through it. The simplest way in which the impedance of a diode can be reduced is to reduce the separation between the anode and the cathode. This is possible only up to a point, for if the electrodes are too close together a considerable amount of heat is transferred by radiation to the anode and we have already seen that this is undesirable. A more important consideration is that, if the electrodes are close together, the peak inverse voltage may be reduced to such an extent that an arc discharge from cathode to anode is set up during the negative half cycle, when the whole generator voltage is applied to the valve. When the electrodes have been brought together as close as safety allows, the impedance can then be further reduced by increasing the length of the anode and of the cathode. The length of the anode is decided by the size of the tube, but the cathode, if it is of the filamentary kind, may be folded several times (fig. 3/5) thus gaining a considerable increase in effective length. With an indirectly heated cathode such folding is difficult to accomplish and rectifier valves of the high vacuum kind therefore usually have directly heated cathodes. Dissipation of heat at the anode may be considerable and arrange-

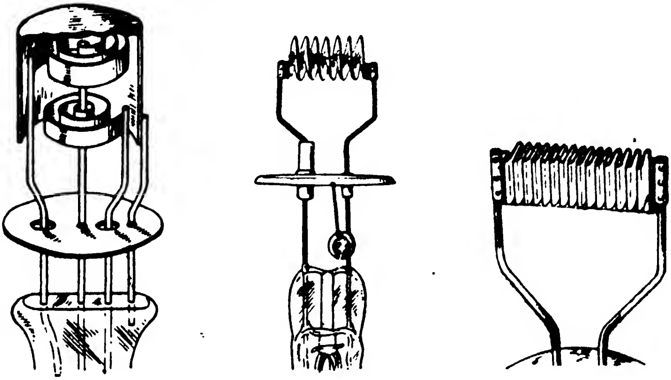


Fig. 5/2. Cathode efficiency is an important consideration in gas-filled valves designed for heavy currents and directly-heated cathodes of the kinds shown in this diagram are commonly used. Most of the heat radiated from the surface is absorbed by adjacent portions of the filament and is not lost.
(From Reich: Theory and Applications of Electron Tubes, by courtesy of McGraw-Hill Publishing Company).

ments to deal with this usually take the form of air cooling by a system of fins around the anode end of the valve, or water cooling in which a stream of water circulates continuously around the outside of the tube.

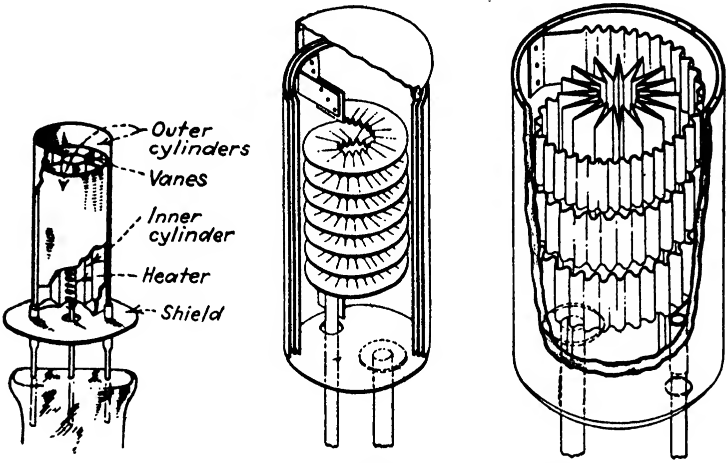


Fig. 5/3. Folded cathodes for indirect heating.
(From Reich: Theory and Applications of Electron Tubes, by courtesy of McGraw-Hill Publishing Company).

We have seen that the alternating voltage which can be rectified by a diode is limited by the danger of an arc discharge between cathode and anode during the negative half cycle. The potential difference which a tube will safely rectify may be raised by increasing the separation of the electrodes but this change increases the impedance of the valve and therefore the voltage drop across it. This being so, the total voltage to be rectified is never applied to the load and the efficiency is less than the ideal by the proportion of the generator voltage lost in this potential drop across the tube. As the valve offers a constant resistance during positive half cycles the potential drop across it varies as the current drawn by the load is altered.

Limitation of current. It is desirable that the potential drop across a diode rectifier should be kept fairly small and the operating conditions are therefore chosen so that the current through the valve is limited by the space charge (page 25). An increased current may be obtained by increasing the temperature of the filament, thus producing a larger space charge. This however reduces the life of the filament, which is normally operated at only a few hundred degrees below its melting point, and for applications requiring heavy currents the high vacuum diode is not suitable. It is still satisfactory where moderate currents (up to a few amps.) are required and is used in this way for the rectification of very high voltages—diodes have been made to deliver one amp. at 100,000 volts.

Characteristics of gas-filled rectifiers. Several of the limitations of the high vacuum diode may be overcome by the use of gas-filled valves. In valves of this kind the electrons attracted to the anode produce ionisation (page 10) in colliding with the gas molecules. The electrons so formed add to the anode current and the positive ions move towards the cathode under the influence of the electrostatic field between the electrodes. The ions and electrons suffer further collisions and there is a cumulative increase in ionisation which finally becomes very considerable. The positive and negative particles which are formed have the same numerical charge and therefore acquire the same amount of kinetic energy in moving across the valve. As the positive ions are much more massive than the electrons they move more slowly for the same kinetic energy and therefore remain in the inter-electrode space for a longer time. Thus at any instant there is a considerable number of positive ions inside the valve. These eventually reach the cathode and in so doing they reduce the effect of the negative space charge. If enough ionisation takes place, the effect of the space charge can be eliminated completely, so that all the electrons emitted by the cathode, together with those produced by ionisation, are drawn to the anode. Gas filled valves are therefore capable of giving much larger currents than high vacuum valves. The gas-filled rectifier has certain other advantages over the high vacuum type. The voltage drop across it is small—less than 50 volts—and constant. This voltage drop depends on the ionising potential of the gas or vapour in the tube and on the temperature. It does not depend on the shape of the electrodes or the distance between them and, for this reason, indirectly heated cathodes may be used. The power required to maintain the cathode at a suitable temperature may be greatly reduced by using a folded construction as shown in figs. 5/2 and 5/3. Here electrons are emitted from the whole surface, though loss of heat by radiation occurs only from a portion of the total emitting area. In high vacuum tubes this form of cathode is not practicable, because the space charge would concentrate in the holes and slots, turning back any electrons emitted from these regions ; in gas tubes the trouble does not arise, since the space charge is neutralised by positive ions. Folded cathodes can be designed to give the required emission with only a twentieth ot the heating power required by a filamentary cathode.

The operating characteristics of the mercury-filled valve depend on the vapour pressure which in turn depends on the temperature. Valves of this kind are therefore susceptible to changes in temperature of their surroundings ; thus for one type of mercury rectifier the maximum peak inverse voltage is 25,000 volts at 30°C. and only 5,000 volts at 90°C. This consideration arises only in vapour filled tubes ; if gases, such as argon or helium are used, the variation in pressure with changes in temperature is not so great as to upset the operation of the valve.

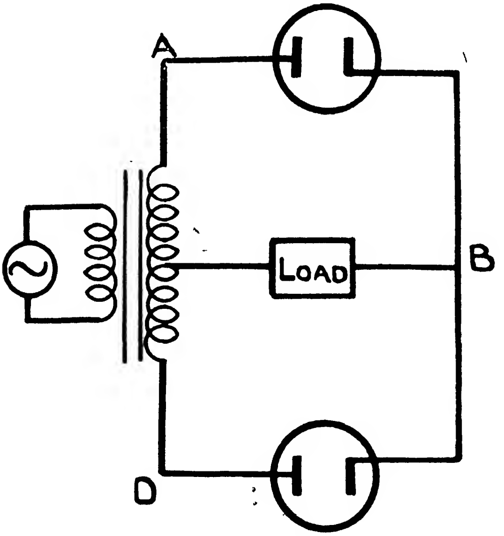


Fig. 5/4. Full-wave rectifying circuit. Current flows through each valve in alternate half-cycles giving an output waveform of the kind shown in fig. 5/5.

The indirectly heated cathodes normally used in gas-filled rectifiers do not reach their operating temperature immediately and it is necessary to allow an interval, sometimes of several minutes, before switching on the high tension supply to the anode. If the tube is connected to a high voltage generator with the cathode cold, no current flows and there is no voltage drop in the load. The whole of the generator e.m.f. is thus applied to the valve, where there is a certain amount of ionisation.* The

---

*See page 13.

electrons so produced move to the anode at high speed under the influence of the potential difference across the valve and the positive ions similarly travel towards the cathode. Since the positive ions are much heavier than the electrons and since the cathode is necessarily less robustly made than the anode, this process may lead to disruption of the emitting surface of the cathode. It is therefore essential to delay application of the high tension to the anode until the cathode has reached its operating temperature and time delay switches or circuits are often used for this purpose.

Direct current from an alternating source. A problem of frequent occurrence in electronics is that of obtaining a supply of direct current from the alternating current mains or from an A.C. generator. We have already seen that a diode rectifier produces a unidirectional current by suppressing alternate half cycles of the supply. In full wave rectification both valves of the supply cycle may be used (fig. 5/5). Here the generator voltage is applied to the circuit by means of a transformer.

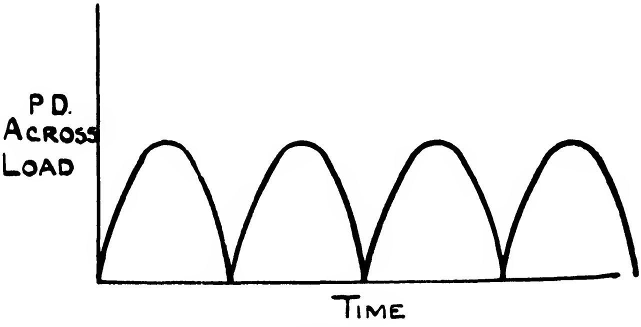


Fig. 5/5

The centre of the secondary winding is connected through the load to the cathodes of the two diodes and the ends of the winding are joined one to each of the two anodes (fig. 5/4). The operation of this arrangement is as follows. An alternating potential difference is set up between *A* and *D*. During one half cycle *A* is positive and *D* is negative with respect to the centre of the secondary winding so that the current flows in the circuit *ABC*. In the next half cycle *A* is at a lower potential than *C* and no current flows in *ABC*, but since *D* is now positive with respect to C, current flows in the circuit DBC. Thus in each half of the cycle current flows through the load in the direction BC and the potential difference across it has the form shown in fig. 5/5. Full wave rectification may also be accomplished by the use of a single valve with two anodes and a common cathode as shown in fig. 5/6.

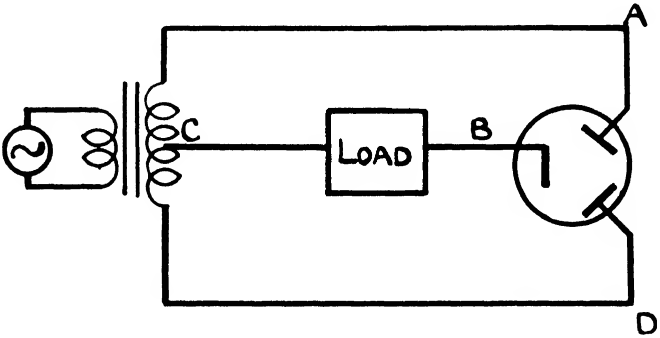


Fig. 5/6. The two diodes of fig. 5/5 may be combined in one valve with a common cathode and two anodes.

The pulsating voltage obtained from the rectifying valve must now be changed into one of constant amplitude by a process known as smoothing, which involves the use of a filter circuit. (fig. 5/7). During the first quarter cycle (AB on graph) current flows through the load by way of L and at the same time the reservoir condenser C_1 becomes charged to a potential equal to the peak e.m.f. of the generator less the voltage drop across the valve. The generator voltage then decreases to zero and immediately begins to rise again (BCD). While the generator e.m.f. is less than the potential difference across C_1, this condenser discharges through L, thus maintaining a flow of current through the load. At the point D where the rising generator e.m.f. becomes equal to the potential difference across the condenser, C_1 acquires an additional charge (DE). The potential difference across C_1, which has a waveform represented by the line $BDEF$ on the graph, is applied to L and C_2 in series. This potential difference may be resolved into a steady component, equal to its mean value, and several alternating components. For the steady component, C_2 behaves as an infinite resistance and if no current is drawn by the load the whole of this component is developed across C_1. If a current I is taken by the load, the voltage across C_2 drops by RI, where R is the resistance of the choke. L offers a large impedance to the alternating component of the output current so that only a small alternating voltage appears across the load. In a typical circuit L may be 40 henries and C_2 4 microfarads. At 100 c/s, which is the fundamental ripple frequency for rectification from the usual 50 c/s supply, the impedance of L is about 60 times that of C_2 and the potential difference developed across C_2 is substantially constant. If further smoothing is necessary additional chokes in series and condensers in parallel may be added to the filter.

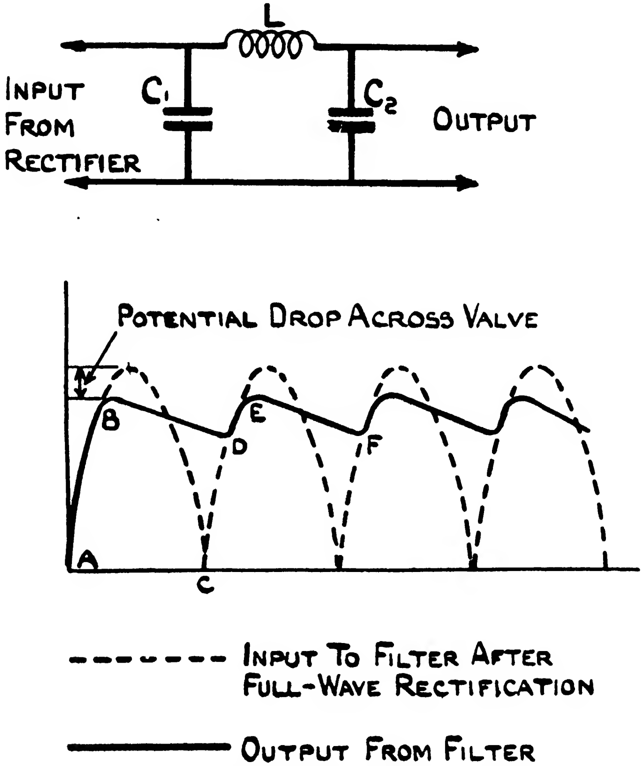


Fig. 5/7. A smoothing network is necessary to convert the output from a rectifier circuit into a steady potential difference.

Reservoir condenser. It will be observed that current is drawn from the valve for only a small part of each cycle, during which the condenser replenishes its charge. The momentary overheating caused in the valve by these surges limits the current which may be drawn. The size of the current pulses depends on the extent to which the potential difference across C_1 falls between the charges, that is, on the current drawn by the load and on the capacitance of C_1. To indicate this limitation valve manufacturers specify the maximum rectified current which may safely be drawn and the maximum capacitance of the reservoir condenser if overheating is to be avoided.

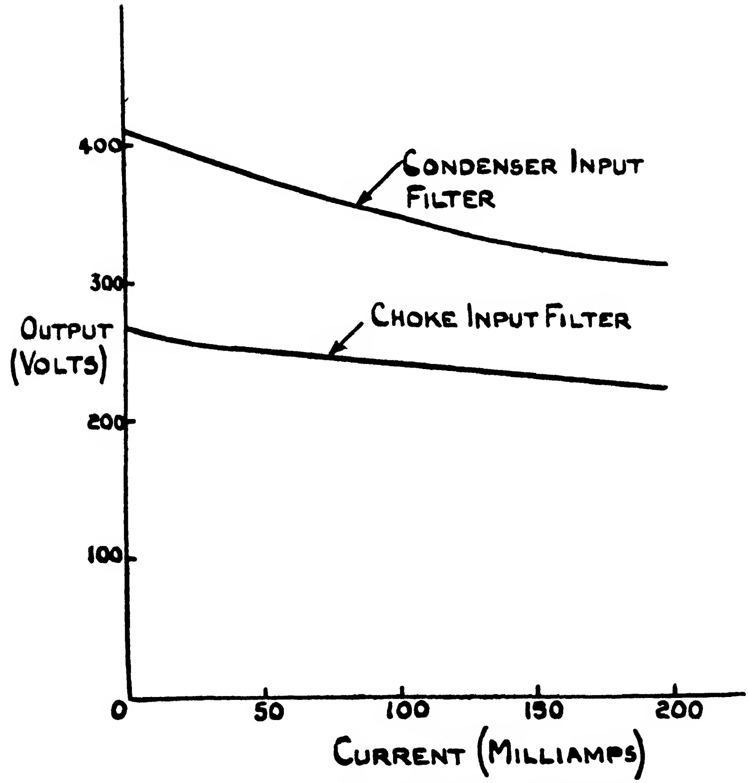


Fig. 5/8. Relation between output voltage and current drawn by load for condenser-input and choke-input filter circuits.

Regulation. In small diode rectifiers for use at voltages up to 5000 the impedance cannot be reduced below a few hundred ohms, on account of the necessity for separating the electrodes by a distance large enough to prevent sparking. Thus a large current cannot be drawn without increasing the potential drop across the tube and therefore reducing the voltage delivered to the smoothing network. In general, valves of this type will not deliver a rectified current greater than 250 milliamperes.

As the load current is increased the rectified voltage delivered at the output of the filter is reduced on two counts. Firstly, the impedance of the valve is fairly constant and the voltage drop across it increases as the current is made larger. Secondly, with a resistive load across C_2, the voltage lost as a drop in potential across the resistance of L is proportional to the current flowing. Curves illustrating these effects are set out in fig. 5/8. The constancy of the potential difference supplied by a rectifying device as the current varies is known as its ***regulation*** and the regulation of a diode rectifier is, as we have seen, inherently poor. Mercury vapour rectifiers may be used for greater currents, as the voltage drop across a valve of this type is small (less than 50 volts) but cannot, on the other hand, be used for the rectification of high voltages because the presence of ionised vapour lowers the insulation between anode and cathode. The mercury vapour diode is also easily damaged by momentary overheating. Smoothing circuits for it therefore usually take the form shown in fig. 5/9 in which the current surges to the reservoir condenser are limited by the choke L_1. With a choke input filter, the voltage delivered to the reservoir condenser is less than would be the case with a condenser input filter, such as that previously discussed, and the rectified output voltage is correspondingly less.

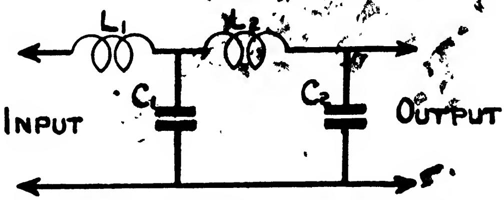


Fig. 5/9. Heavy-current surges, tending to overheat the rectifier cathode, may be avoided by using the choke-input filter circuit illustrated here.

CHAPTER VI

THERMIONIC VALVES

✓**Triodes.** We have seen that the diode is used principally as a rectifier. Other uses of thermionic valves were made possible by the production in 1907 of the three electrode or *triode* valve, which contains, between the anode and cathode, a *grid.* The grid is usually a spiral wire and is nearer to the cathode than to the anode (fig. 6/1).

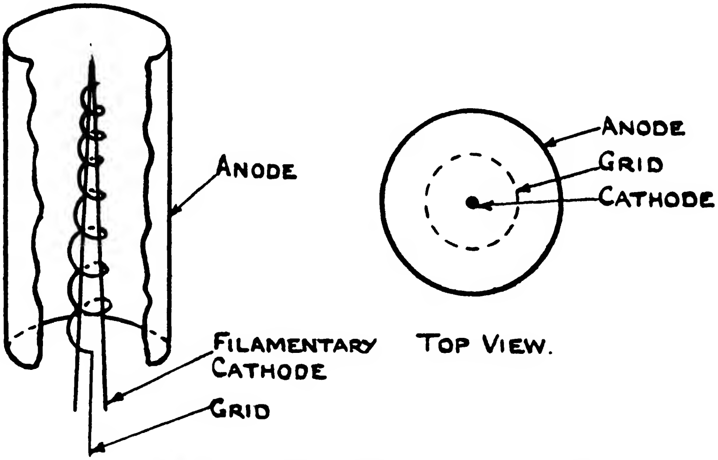


Fig. 6/1. Construction of the triode valve.

Valves are usually operated at anode voltages well below the saturation value and as explained on p. 23 the anode draws electrons from the space charge. The effect of interposing a grid is as follows.

1. If no potential difference is applied externally between the grid and cathode, conditions are the same as in a diode, except that a few electrons moving towards the anode strike the wires of the grid and make it negatively charged, unless there is an external path back to the cathode.

2. If the grid is kept at a higher potential than the cathode by means of a battery, the positive charge attracts electrons from

the space charge. Most of these pass through the grid and reach the anode. Some strike the grid with the result mentioned above. As the grid is close to the cathode, the current flowing through the valve is greatly increased by the application of a positive grid potential. Triodes are usually constructed so that, with the normal anode voltage, saturation sets in when the grid is a few volts positive and most of the useful working range lies in the region of negative grid voltage.

The anode current I_a flowing in a triode depends on the anode voltage V_a and the grid voltage V_g. The relations existing among these three variables are usually set out graphically in the form of

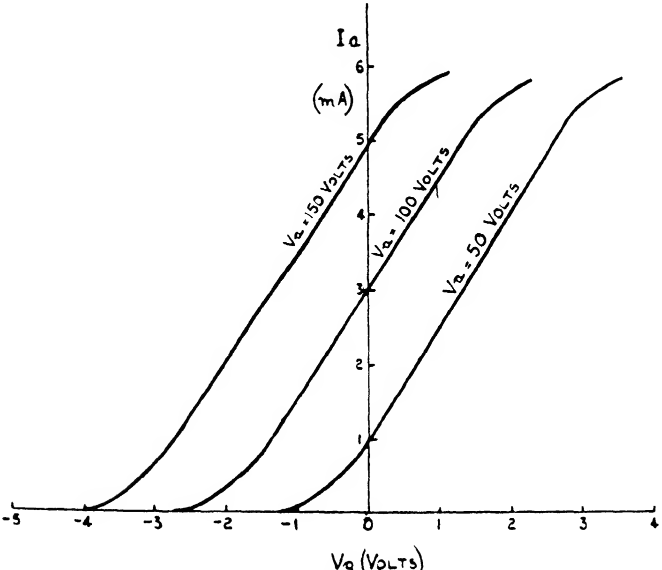


Fig. 6/2 (*a*). Grid characteristics for a typical triode valve, showing the relation between anode current and grid potential, for various anode potentials.

characteristic curves (figs. 6/2 (*a*), 6/2 (*b*)). The anode characteristics show the relation between I_a and V_a for various fixed values of V_g. The curve corresponding to $V_g=0$ is of the same shape as the anode characteristic of a diode (p. 24). The anode characteristics for negative values of V_g are straight and parallel for most of their length. The anode a.c. resistance of the valve is defined as

$$\rho = \frac{\partial V_a}{\partial I_a} \quad \dots\dots\dots\dots\dots\dots 6.1$$

and is seen to be the reciprocal of the slope of the anode characteristic. For a triode ρ may be about 20,000 ohms.

As we have seen, the anode current is influenced more by the grid voltage than by the anode voltage and grid characteristics, relating I_a to V_g, are often more useful than the anode characteristics. Curves of this kind, drawn for various values of V_a are shown in fig. 6/2 (*a*), and are seen to be approximately straight and parallel for most of their length. These graphs contain the same information as those of fig. 6/2 (*b*), and both may be regarded as sections of a three-dimensional characteristic surface. The slope of the grid characteristics is known as the ***mutual conductance*** of the valve and is given the symbol *g*. Thus

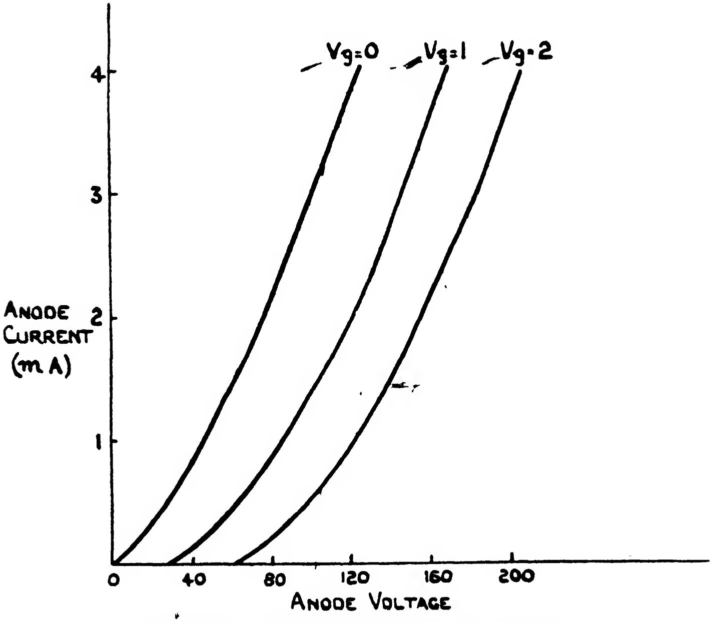


Fig. 6/2*b*. Anode characteristics for a triode.

$$g = \frac{\partial I_a}{\partial V_g} \qquad \text{..............................6.2}$$

Mutual conductance is measured in milliamperes per volt or microamperes per volt (the term micromho is sometimes used to denote a mutual conductance of one microampere per volt). For a triode *g* may be 1·5 mA/V.

A third quantity related to ρ and g is the amplification factor μ defined as

$$\mu = \frac{\partial V_a}{\partial V_g} = \rho g$$

$$\text{Thus } \mu = \frac{\text{change in anode voltage to produce a given change in anode current}}{\text{change in grid voltage to produce the same change in anode current.}}$$

μ indicates the relative effectiveness of the grid and anode potentials in controlling the anode current. As the grid is nearer to the cathode than to the anode, μ is greater than unity and may be about 30 for a typical triode. μ, ρ and g are known as the *characteristic constants* of the valve though actually they are only constants if the grid and anode characteristics are parallel straight lines. V_g and V_a are usually adjusted so that the valve is working on the straight portion of the characteristics and the values quoted for g and ρ refer to these conditions; it may be shown that $\mu = g\rho$. It is apparent that μ and ρ depend largely on the size, shape and separation of the electrodes. μ is entirely a geometrical property of the valve; ρ and g depend also on the emission characteristics of the cathode.

Triode as an amplifier. A triode may be used in such a way that the application of a small difference of potential between the grid and cathode causes a large difference of potential to be set up elsewhere in the circuit. This process is known as *voltage amplification* and the operation of an amplifier is as follows.

Consider the grid characteristics shown in fig. 6/2. When $V_g = 0$ and $V_a = 100$ volts, $I_a = 3$ mA. If V_g is increased to $+1$ volt, I_a becomes 4·5 mA. and if V_g is reduced to -1 volt I_a falls to 1·5 mA. Thus if an alternating potential difference of peak value 1 volt is applied between grid and cathode, an alternation of amplitude 1·5 mA. is superimposed on the steady anode current which flows when $V_g = 0$. This is shown in the graphs of V_g against time and I_a against time, drawn with the same scales as the characteristic curve in fig. 6/3. If now a resistance R is placed in series with the anode battery, the alternating component of I_a flowing through it produces an alternating potential difference and, as will be shown, this may be much larger than that originally applied to the grid.

Grid bias. If the output potential difference is to have the same wave form as the input, the change in anode current must be proportional to the change in grid voltage, that is, the valve must be operated on the straight portion of its characteristic. To do this it is usually necessary to apply a small negative potential to the grid, in series with the alternating input voltage. This is known as the *grid bias* and in the case of the valve to which fig. 6/2 refers the operating point is in the centre of the straight part of the 150 volt characteristic when $V_g = -1{\cdot}5$ volts.

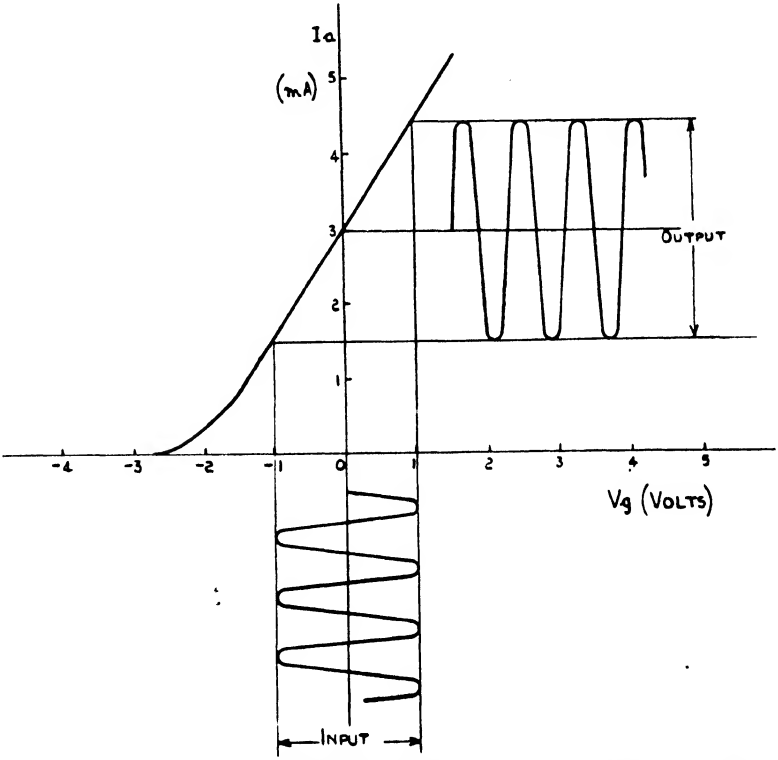


Fig. 6/3. When an alternating potential difference is applied between the grid and cathode of a triode valve, the anode current alternates at the same frequency.

Effect of load resistance. The resistance R is referred to as the load of the valve. As the output voltage is equal to R multiplied by the alternating component of I_a, it might be supposed that the amplification could be increased indefinitely by increasing R; this, however, is not the case. The anode current consists of a steady component (depending on the anode and grid potentials), which flows through the load whether or not an alternating voltage is applied to the grid together with an alternating component which depends on the input between grid and cathode. The steady component flowing through the load produces a voltage drop across it and the potential difference between anode and cathode is therefore less than that of the anode supply battery. As *R* is increased the anode voltage becomes less and the effectiveness of the circuit as an amplifier therefore diminishes.

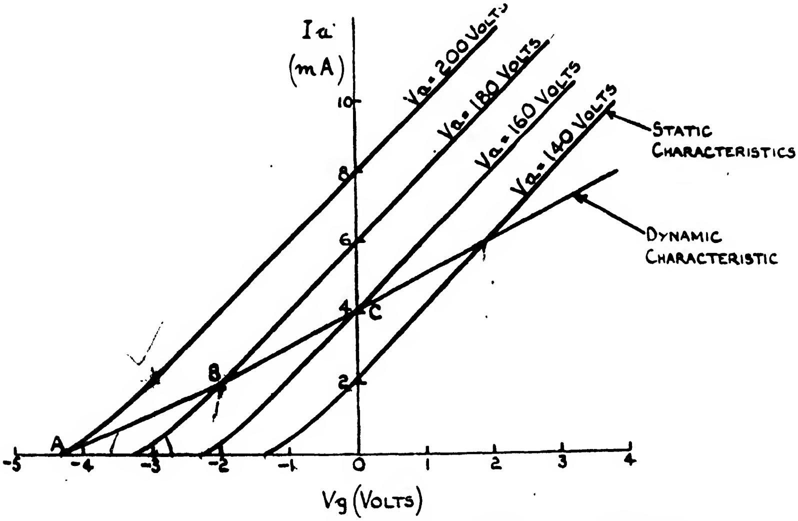


Fig. 6/4. Static and dynamic characteristics for a triode with a resistive anode load.

Reference to fig. 6/4 illustrates this point. If the high tension supply is at 200 volts, and no load is connected, $I_a = 2$ mA. when $V_g = -3$ v. With a load of 10,000 ohms and an anode current of 2 mA. the potential difference between anode and cathode is only 180 v. for the current flowing through *R* causes a potential difference of 20 v. between its ends. With this reduced anode voltage, a current of 2 mA. corresponds to a grid bias voltage of -2 v. Thus the point *B* corresponding to a current of 2 mA. lies on the characteristic curve for which $V_A = 180$ v. Similarly the point *C* for $I_A = 4$ mA. lies on the curve drawn for $V_A = 160$ v. The line *ABC* indicates the true relation between V_g and I_A, allowing for the potential fall in the load resistance, and is the *dynamic characteristic* curve of the valve plus resistance. Its slope depends on the value of the load, but is always less than the slope of the curves for zero load—referred to as *static characteristic* curves. To calculate the behaviour of an amplifying circuit the dynamic characteristic must be used, though a rough idea of the processes involved may be obtained from the static curves.

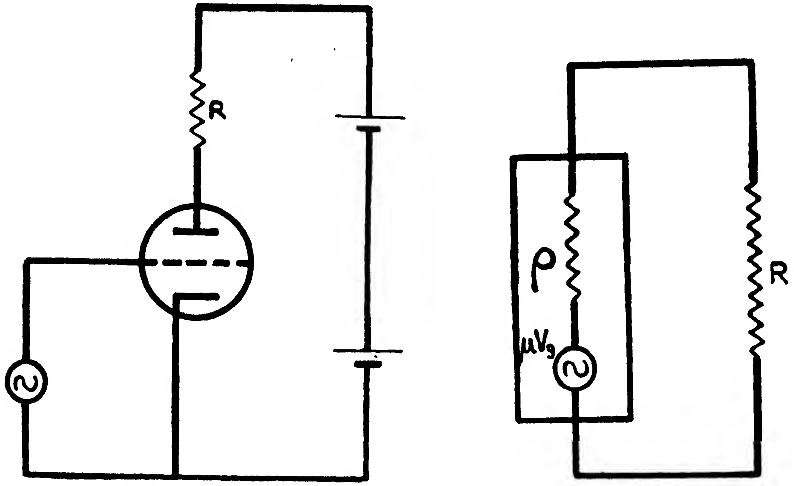


Fig. 6/5. Left—basic circuit for amplification with a triode valve. Right—equivalent circuit, used in calculating the amplification ratio and other properties of the amplifier.

Equivalent circuit. The circuit of fig. 6/5 is generally used for the amplification of alternating voltages and in these circumstances its behaviour may be studied by means of a simpler equivalent circuit in which the valve is replaced by a generator giving an output of μv_g volts in series with a resistance ρ ohms. v_g is here the alternating potential difference between grid and cathode. The high tension supply need not be included in the equivalent circuit since, if it is assumed to have negligible resistance, it does not affect the alternating component of I_a. In the equivalent circuit the alternating current is $\frac{\mu v_g}{R+\rho}$ and the potential difference developed across R is therefore $\frac{R\mu v_g}{R+\rho}$. As the input is v_g volts, the voltage amplification ratio is $\frac{\mu R}{R+\rho}$. Thus the gain may apparently be increased up to the theoretical limit of μ by increasing the load resistance R. As has already been explained, if R is increased too far and the high tension supply remains the same, the anode voltage falls off to such an extent that the circuit ceases to amplify effectively. The gain may be made to approach μ closely if the high tension voltage is increased as R is increased, but it is more economical to use a moderate supply voltage and accept an amplification somewhat less than μ. For triodes R is commonly about twice as great as ρ, giving a gain of $2\mu/3$.

It is important to observe that, with a sinusoidal input voltage, the potential difference developed across the load resistance is out of phase with that applied to the grid by 180 degrees. This relation may be verified by reasoning as follows. If the grid becomes more negative, the anode current falls and the potential difference across the load is reduced, with the result that the anode becomes more positive. If, however, the grid is made less negative, the anode current rises, and, with it, the potential difference across the load; consequently, that between anode and cathode is reduced, that is, the anode potential falls. If the input voltage is not sinusoidal, the output voltage is not, strictly speaking, in the opposite phase, but its waveform is a mirror image of the input waveform (fig. 6/6). Thus if an alternating voltage is applied to the grid, the anode potential rises and falls with the same frequency, but the anode voltage is a maximum when the grid voltage is a minimum and vice versa.

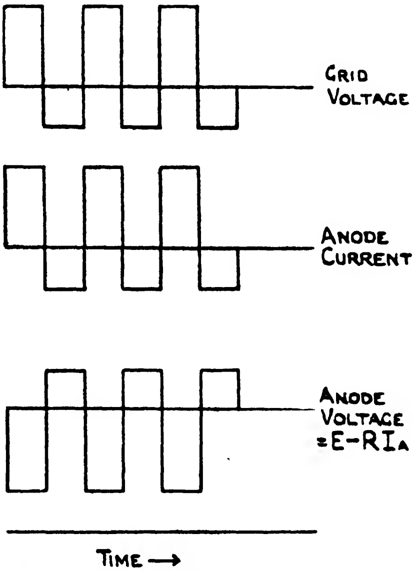
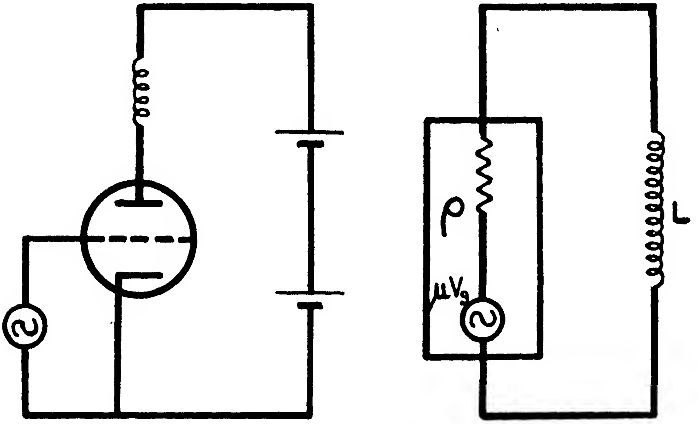


Fig. 6/6. This diagram shows the relation between input and output voltages for a triode amplifier circuit.

Inductive load. The reduction in V_a caused by the steady component of anode current flowing through the load may be much reduced by using a choke in place of a resistance in the anode circuit. A coil of inductance 20 henries has a reactance of about 12,500 ohms at 100 c./sec., but its resistance to direct current may be no more than 100 ohms. The equivalent a.c. circuit is as shown in fig. 6/7. Ignoring the resistance of the choke, the gain at pulsatance p is $\frac{\mu L p}{\sqrt{\rho^2+L^2p^2}}$ The steady voltage drop across the load is that corresponding to the resistance of the coil and is negligible, so that almost the whole supply voltage is applied to the anode. This arrangement is useful when a single frequency is to be amplified—if the input contains a wide range of frequencies the choke must have a very large inductance, since otherwise the gain will vary with frequency.



(i) (ii)

Fig. 6/7. Left—basic circuit for choke-loaded triode amplifier. Right—equivalent a.c. circuit.

The steady component of anode current flows through the choke and the core is therefore permanently magnetised to an extent which depends on the size of this current. The inductance of the coil depends largely on the magnetic intensity in the core, which follows the variations in current flowing through the coil only up to a point, after which further increase in current produces little additional magnetisation; the core is then magnetically saturated. Though the steady component of anode current is not usually sufficient to produce this effect, the total current when the alternating component is added may rise at the positive peaks to a value at which saturation occurs and the voltage across the choke then has not the same waveform as the current flowing through it. This form of distortion is objectionable in amplifiers for the audible range of frequencies and to overcome it the choke must have a large core.

Limitations to triodes as amplifiers. It is difficult for several reasons, to design a triode with an amplification factor greater than about fifty. This can be increased if the grid-cathode distance is reduced, but if the separation is made too small, thermal radiation to the grid may cause emission of electrons from it and other undesirable effects. Improvement is possible by increasing the anode-grid distance, but this method is not satisfactory because valves of large size are inconvenient in apparatus such as radio receivers. The voltage gain which can be obtained with a triode is thus seen to be inherently low, and there is another serious disadvantage of the triode which becomes apparent at high frequencies; it arises from the *inter-electrode capacitances* of the valve. The electrodes are made of conducting materials, and there are small capacitances between them. Thus the anode and grid may be regarded as the two plates of a small condenser, which may form part of an alternating current circuit. Similar capacitances exist between each pair of electrodes as well as between the pins of the valve base and between the leads from these pins to the electrodes themselves. The inter-electrode capacitances are, except at the highest frequencies, the only ones which need be considered. At radio frequencies or even at high audio frequencies, they offer conducting paths between the electrodes for alternating current energy. Thus the alternating component of anode current in a valve may return through the high tension supply to the cathode, or may return by way of the anode-grid capacitance to the input circuit. As we have seen, the anode alternating voltage is out of step with the grid voltage, and the effect of feedback is therefore to reduce the useful input and the gain of the stage in which the valve occurs. This phenomenon which is known as *negative feedback*, or *degeneration*, becomes noticeable at the higher audio frequencies with triode amplifiers and limits their usefulness in this region. The anode-grid capacitance may be neutralised by the use of various circuits, which are now mainly of historical interest, but may be more easily eliminated by means of an electrostatic shield between the grid and the anode. The shield usually takes the shape of a spiral of wire and is known as the *screen grid*. If this grid is connected to the cathode, which is at zero potential, the anode is shielded from the other electrodes, and the anode-grid capacitance is almost eliminated, but another difficulty arises because the electrons moving to the anode are retarded in the low-potential region and many of them do not reach this electrode. This effect again may be overcome by giving the screen grid a positive potential. The screen then accelerates the electrons and the anode current is maintained; if the positive potential is obtained from a battery of negligible internal resistance, the screen grid is at earth potential for alternating currents, and no alternating voltage is fed back to the input circuit across the anode-grid capacitance.

Characteristic curves of the screen grid tetrode. In a tetrode valve of the kind described above, the anode is shielded from the other electrodes and the potential applied to it therefore has little influence on the current flowing through the valve. Electrons are drawn from the space charge by the electrostatic field set up by the screen grid potential; some of these electrons strike the screen grid and the remainder pass through to the anode. Those which are intercepted constitute a screen current, I_s, and return to the cathode. The total current flowing through the valve, known as the space current, is $I_a + I_s$, and depends mainly on V_s and V_g. The way in which the space current divides between the screen and the anode depends on V_a and V_s.

Anode characteristics for a screen grid tetrode are of the form sketched in Fig. 6/8, and the reason for their curious shape is as follows. The curves are drawn for fixed values of V_g and V_s, so that the space current rises only slightly as V_a is increased. If the anode is at zero potential, very few electrons reach it and the screen current is therefore large. As V_a is increased, I_a increases and there is a corresponding reduction in I_s. The velocity of the electrons which reach the anode depends on V_a, and as this voltage rises, some of them acquire sufficient energy to eject other electrons from the anode surface, thus producing secondary emission. In the case of the valve to which fig. 6/8 refers, secondary emission begins when V_a is 20 volts, and the secondary electrons are therefore attracted to the screen grid which is at a higher potential than the anode. The result of this migration is to increase the screen current at the expense of the anode current. It is possible for one electron arriving at the anode to eject several others and I_a may therefore even have a negative value. As V_a continues to increase, some of the secondary electrons are returned to the anode and when V_a is greater than V_s, secondary emission produces no effect on the screen or anode currents. Thus in Fig. 6/8 the characteristic curve is fairly straight for values of V_a greater than about 100 volts. The following points should be noted :

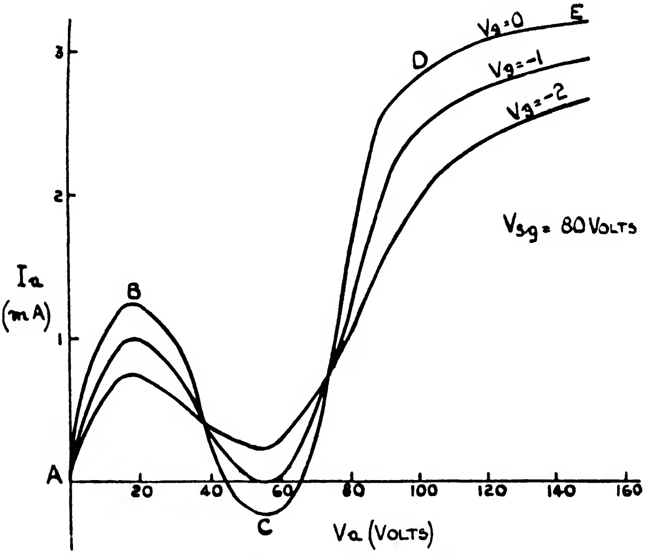


Fig. 6/8. Anode characteristics for a typical screen-grid tetrode.

(1) In the region *BC* of fig. 6/8, the anode current decreases as the anode voltage increases and the valve may be regarded as having a negative resistance. A valve operated in this way is known as a *dynatron* and is sometimes used in oscillator circuits.

(2) Secondary emission occurs in triodes, but the secondary electrons return to the anode, because it is at a higher potential than either of the other electrodes and the anode current is therefore not affected.

(3) The screen tetrode is normally operated with V_a greater than V_s and the anode characteristic in this region is almost straight. The slope of the working part of the curve—*DE* in

fig. 6/8—is small and ρ is therefore large; 250,000 ohms is a typical value. The mutual conductance is of the same order as in a triode.

(4) The screen grid tetrode has a large amplification factor, because the anode voltage has little effect on the anode current. μ which is $\frac{\partial V_a}{\partial V_g}$ may be about 300. This property makes the valve more effective than the triode for voltage amplification and, as the inter-electrode capacitance is reduced by the presence of the screen, amplification at high frequencies is possible. The anode-grid capacitance for a screen grid valve may be about 0·1 micro-microfarad—fifty times less than that of a typical triode. At high radio frequencies—above one megacycle per second—even this small capacitance becomes objectionable and to reduce it further another grid is necessary.

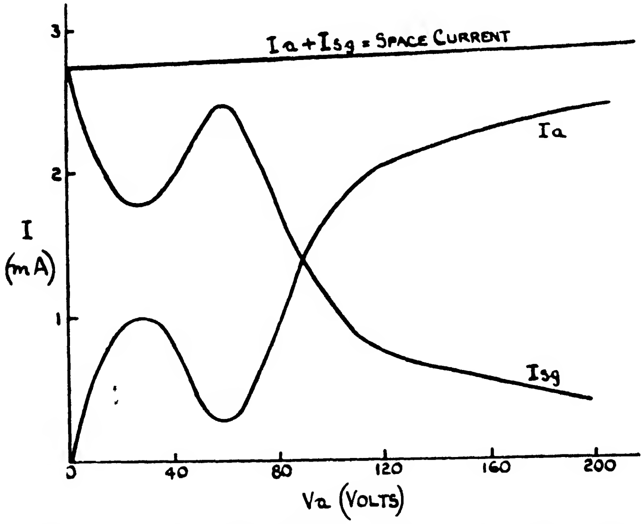


Fig. 6/9. In a screen-grid tetrode, the sum of the anode current and the screen current is approximately constant, and the two quantities vary with anode voltage as shown in this diagram.

The pentode valve. The pentode has a third grid, the suppressor, between the screen grid and the anode. The suppressor is a loosely-wound spiral wire, usually connected internally to the cathode. Its effect is to slow down the secondary electrons emitted by the anode, so that they do not reach the screen. The suppressor grid does not appreciably retard the normal grid-to-anode current, as these electrons are moving much faster than those ejected from the anode. In the pentode, secondary emission effects are almost eliminated and the anode current rises to its full value when V_a is quite small; anode characteristics for a typical pentode are shown in fig. 6/10. Since the screening of the anode from the other electrodes is even more complete than in the tetrode, μ and ρ are both very large. ρ is usually of the order of a megohm and μ may be as high as 2,000. The anode-grid capacitance is normally about ·001 micro-microfarad—a value low enough to permit a satisfactory amplification at frequencies of several megacycles per second.

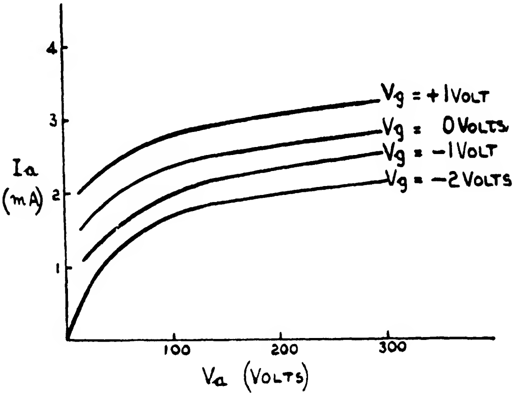


Fig. 6/10. Anode characteristics for a typical pentode valve.

For voltage amplification the screen grid tetrode is now almost obsolete, though a modification—the *beam tetrode*—which will be discussed later, is widely used for power amplification. It is not practicable to obtain from a pentode valve an amplification ratio approaching μ, since to do so a load of several megohms must be used and the steady anode voltage is therefore reduced to a low value (see p. 73). Common practice in the design of pentode voltage amplifiers is to use a load resistance of 250,000 ohms, which usually permits a gain of 100-200 in the audio-frequency range. The anode current is almost constant over the useful portion of the characteristic curve and the amplification ratio depends mainly on the extent of the variation in I_a which can be brought about by alternations of V_g. Thus the mutual conductance, g, really determines the performance of the pentode as a voltage amplifier; its value may be 1-2 mA./volt.

Voltage supplies to electrodes. It will be convenient at this stage to consider some of the methods used for supplying positive and negative potentials to the various electrodes in a valve.

(1) Anode. The high tension supply is usually connected to the anode through a load resistance and may be derived from batteries or from the alternating current mains, with a rectifying circuit of the kind described previously (page 63).

(2) Screen grid. If batteries are used, it is convenient to connect the screen grid to a tapping at the appropriate potential, which is usually less than that of the anode. In this case, the screen grid is effectively at earth potential for alternating current, since there is no source of alternating e.m.f. between it and the cathode.* In apparatus driven from the mains, it is necessary to obtain the anode and screen voltages from a single high tension supply. Since the screen grid draws a small steady current, on account of the electrons which strike it, its potential may be adjusted by means of a resistance R_2 connected to the positive terminal of the supply (fig. 6/11). The value of this resistance is calculated as shown. When an alternating voltage is applied to the grid, an alternating component is introduced into the screen current and this electrode is no longer at earth potential for alternating currents—the screen supply resistance now functioning as a generator of alternating e.m.f. The difficulty is overcome by connecting between the screen and earth a condenser of large capacitance, which offers a low-impedance path for the alternating component of the screen current, though not permitting the passage of direct current.

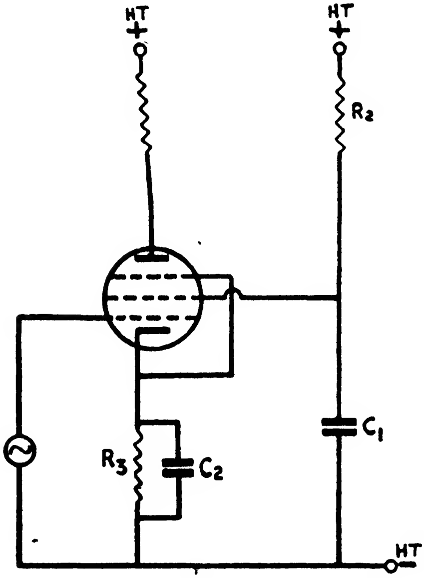


Fig. 6/11. By the arrangement shown in this circuit, potentials required for the anode, screen grid and control grid may all be derived from a single source of supply.

(3) Control grid. The control grid (that nearest to the cathode)

---

* This statement is valid only if the internal resistance of the battery is small. Otherwise, the alternating component of the screen current produces an appreciable alternating voltage in flowing through the battery.

may be given a negative potential with respect to the cathode by means of a resistance in the cathode lead (R_3 in fig. 6/11). The steady component of the space current flows through the valve and through R_3 to reach the negative terminal of the supply, thus setting up a difference of potential across R_3 which makes the grid end more negative than the cathode end. If the lower end of R_3 is connected directly or through the input circuit to the grid, the potential drop across R_3 is applied between the grid and cathode. The space current has also an alternating component and if R_3 alone is connected between A and B, an alternating potential difference is established between grid and cathode. As has been explained (p. 75), this voltage is out of step with the input and therefore has the effect of reducing the gain of the circuit. To eliminate this difficulty, a condenser C_2 is connected across R_3. The combination of R_3 and C_2 then offers a resistance R_3 ohms to direct current and an impedance made up of R_3 and $1/C_2p$ in parallel to alternating current. The capacitance of C_2 is made very large so that the a.c. impedance of the combination is low at all the frequencies to be amplified and no appreciable degeneration occurs.

A numerical example will serve to illustrate the foregoing remarks. Consider a pentode valve which it is desired to operate under the following conditions :

I_a = 1 mA. I_s = 0·5 mA. V_s = 150 volts.
V_g = - 3 volts. High tension supply—350 volts.

In this case the drop of potential across the screen resistor must be 350—150 = 200 volts and the value of R_2 must therefore be 200 volts/0·5 mA = 400,000 ohms. The potential difference across the bias resistance R_3 is 3 volts and the current flowing through it is 1·5 mA ($I_a + I_s$). The value required for R_3 is thus 2000 ohms. The by pass condenser C_2 must have a reactance much below 2000 ohms at all the frequencies to be amplified. A 25 μF condenser which has a reactance of 126 ohms at 50 c/s will serve ; the reduction in gain at lower fequencies is not likely to be important.

(4) Filament or heater. In the case of battery valves, a D.C. supply is necessary for the filament, which is a directly-heated cathode. Mains valves have indirectly heated cathodes and the heaters may be connected to a special winding on the transformer which supplies the rectifying and smoothing circuit.

CHAPTER VII

VALVE AMPLIFIERS

Multi-stage audio-frequency amplifiers. If amplification greater than that obtainable with a single stage is required, two or more valves may be connected in cascade, the output of one forming the input of the next. Such circuits are used for amplification at audio-frequencies, and to a lesser extent at radio frequencies.

Coupling between stages in audio-frequency amplification. Energy may be transferred from one stage to the next in a variety of ways and some of the more common methods will be described.

(a) Resistance-capacitance coupling. In fig. 7/1 the output voltage of the first stage is developed across the load resistance R_1. The point B, which is connected to the cathode through the high tension supply, is at earth potential with respect to the alternating current, but the potential of the point A has an alternating component, and it is this which must be impressed on the grid of the next valve, V_2. It is not advisable to join A to the next valve directly, as the anode of V_1 is at a high steady potential, apart from any alternating voltage which may be present, and a coupling condenser C_1 is connected in the position shown by the diagram. The point B is connected to the cathode of V_2 by the common negative lead of the high tension supply and the alternating voltage developed across R_1 is thus applied between the grid and cathode of V_2. A number of electrons strike the grid of V_2 and it is necessary to include a grid leak resistance, R_2 to provide them with a path back to the cathode. If R_2 is omitted the grid becomes negatively charged and eventually current ceases to flow through the valve.

Equivalent A.C. circuit. The circuit equivalent to that of fig. 7/1 for alternating currents is shown in fig. 7/2 and the following conclusions may be drawn from it.

(1) The output of V_1 is applied to C_1 and R_2 in series and the potential difference developed across R_2 constitutes the input for V_2. That developed across C_1 is not amplified further, and the

reactance of C_1 should therefore be small in comparison with the resistance of **R₂** at all the frequencies to be amplified.

(2) The anode load of V_1 consists of R_1 with C_1 and R_2 connected across it, and is therefore less than the resistance of R_1 alone. The gain of the stage depends on the load and to avoid reducing it, R_2 should be several times larger than R_1. Customary values for a triode are C_1 = 0·01—0·05 μF and R_2 = 250,000—500,000 ohms.

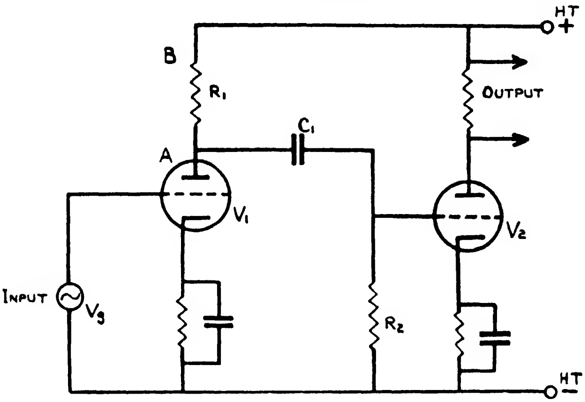


Fig. 7/1. Resistance-capacitance coupling in an amplifier.

(3) At low frequencies the reactance of C_1 becomes large and a smaller proportion of the output from V_1 is applied to V_2. The gain therefore falls off.

(4) At high frequencies the effect of capacitances between the electrodes and of other stray capacitances becomes noticeable. Across the load of V_1 there are (*a*) the anode-cathode capacitance of V_1 (5 – 10 $\mu\mu F$) (*b*) the grid-cathode capacitance of V_2 (about 5 $\mu\mu F$) and various smaller capacitances due to valve-holders and other components. At low frequencies the shunting effect produced in this way is not important, but at high frequencies—above 10,000 c/s—the reactances become small enough to reduce the gain considerably. Fig. 7/3 shows a typical response curve

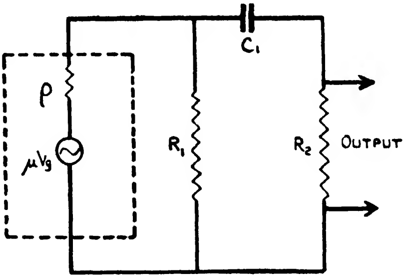


Fig. 7/2. Equivalent a.c. circuit for the amplifier shown in fig. 7/1.

for a resistance-capacitance coupled amplifier, which gives almost uniform gain from 100 c/s to 10,000 c/s, a range which is satisfactory for most purposes. In apparatus such as radio receivers the fidelity of reproduction is determined largely by the frequency response characteristics of the loud speaker.

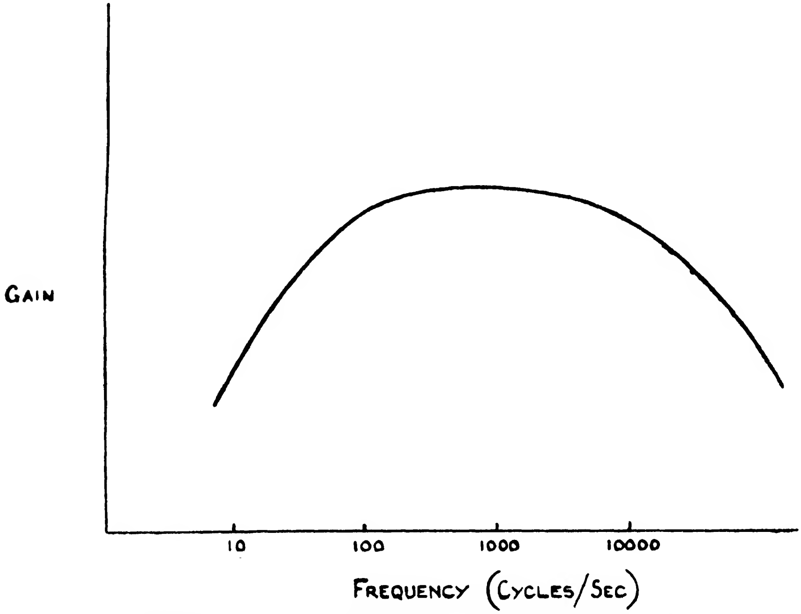


Fig. 7/3. The resistance-capacitance coupled amplifier gives a gain which is approximately constant between 100 c/s and 10,000 c/s, falling off at each end of the frequency scale.

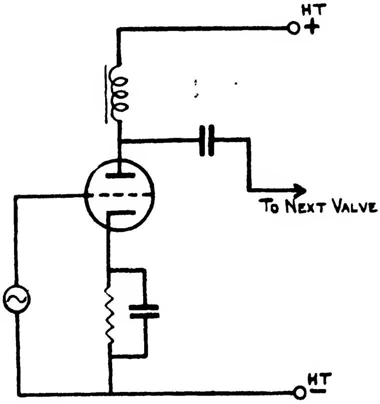


Fig. 7/4. Choke-capacitance coupling.

(b) Choke-capacitance coupling. The use of an inductance as the anode load of a valve has already been mentioned (page 75) and a choke-loaded valve may be coupled to the succeeding stage by a condenser (fig. 7/4). The advantage of this arrangement is that there is only a small steady voltage drop across the load and the high tension voltage can therefore be made smaller than for a resistance loaded valve. If a fairly uniform frequency response is to be obtained the inductance of the coil must be so large that its impedance is much greater than ρ at all the frequencies to be amplified. By using a large inductance (50—100 henries) a satisfactory low-frequency response may be obtained, but the following considerations then arise. The self capacitance of the choke windings become appreciable and may cause resonance at a frequency within, or near the audible range, thus increasing the gain. (The self capacitance is effectively in parallel with the inductance, and the impedance of the circuit has a maximum value at the resonant frequency.) A choke-loaded amplifying stage usually has a frequency response which rises at the high-frequency end, and may even show a peak at an audible frequency ; in any case, the response is not so uniform as for a resistance-loaded stage.

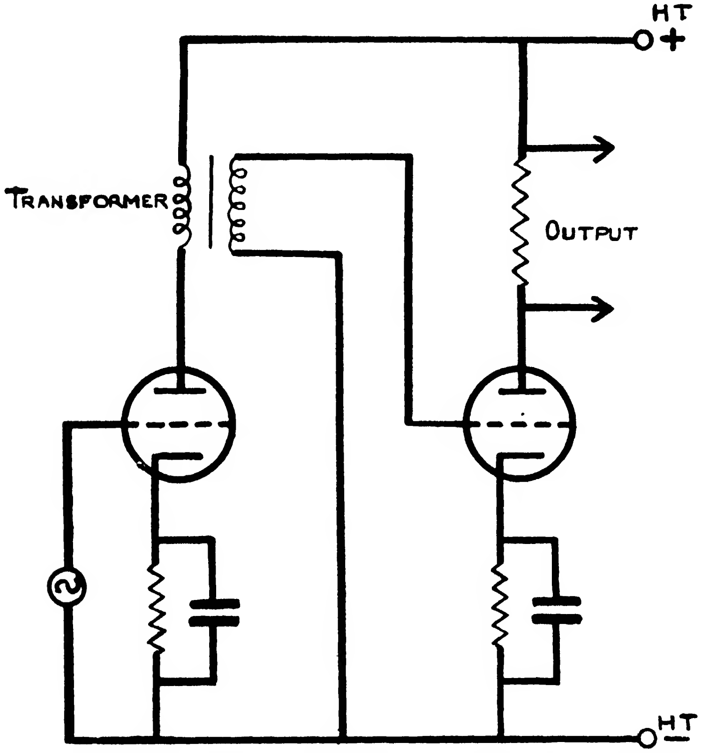


Fig. 7/5. Two stage amplifier with transformer coupling.

✓(c) Transformer coupling. This method is widely used and is a development of the arrangement just described. The output from the first stage is fed to the next valve by mutual inductance between the windings of a transformer (fig. 7/5). In this way, additional gain may be obtained if the secondary winding of the transformer has more turns than the primary ; step-up ratios of 3 : 1 or 4 : 1 are commonly used. The frequency response is affected by the self capacitance of the windings and the capacitance between them but, by taking advantage of resonances produced by these stray capacitances, it is possible to design a transformer which has a fairly uniform response from 100 c/s to about 8000 c/s. Unless a large core is used the steady component of anode current flowing through the primary winding may lead to magnetic saturation which has the effect of flattening the peaks of the secondary current (fig. 7/6). This effect may be avoided by the use of parallel-feed coupling (fig. 7/7). In this circuit the anode current of V_1 has two alternative paths. It may flow through the choke L_1 and the high tension system to reach the cathode, or it may return by way of the condenser C and the primary winding of the transformer T. If L_1 has a large induc-

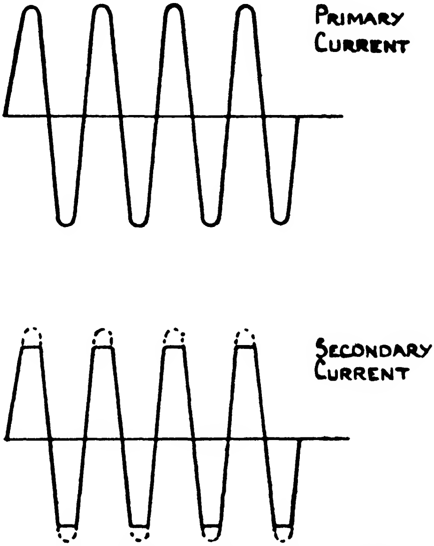


Fig. 7/6. Distortion may be produced in a transformer by magnetic saturation of the core.

tance most of the alternating component of the anode current flows through the transformer, but the direct current component is prevented by the condenser *C* from taking this path and the possibility of saturation is therefore reduced.

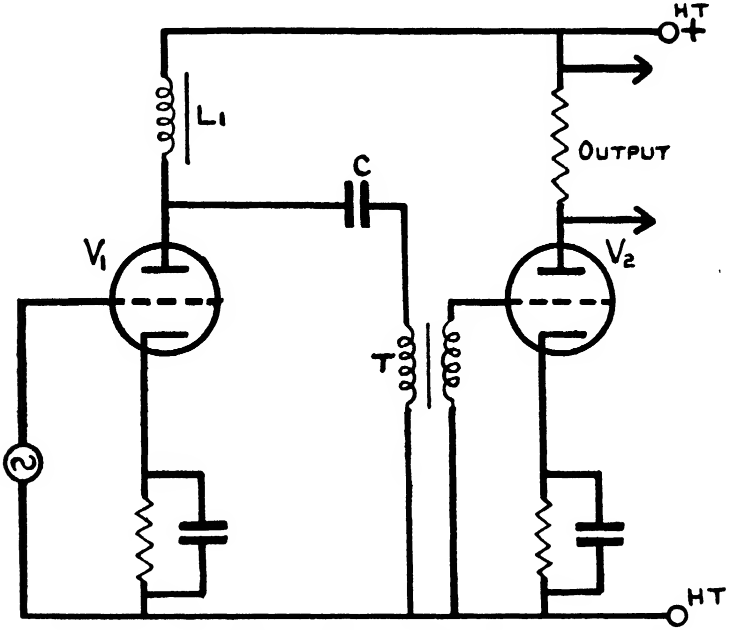


Fig. 7/7. Two stage amplifier with parallel-feed coupling.

Power amplifiers. The amplifying circuits so far described operate most effectively when the output is developed as an alternating difference of potential across a large resistance. The current flowing through the load resistance of a voltage amplifier is generally very small and the power in the output circuit is also small. The ultimate object of most amplifiers is to produce energy which at audio-frequencies may be converted into sound by a loud-speaker, and at radio frequencies may be converted into electromagnetic radiation by means of an aerial system. Each electron arrives at the anode with a maximum kinetic energy of eV_a and the greatest power which can be dissipated in the anode circuit is therefore $I_a V_a$. Much of this is used in heating the anode and the useful power is considerably below the maximum but it is apparent that to increase the energy output it is necessary to increase I_a or V_a or both. In large power amplifiers V_a may be several thousand volts, but in small valves it is limited by insulation difficulties to a few hundred volts and I_a must therefore be large. To obtain a large anode current with a moderate anode voltage μ must be small so that the anode can produce a considerable electrostatic field in the region of the space charge. It is also necessary to use a cathode which gives a large electron emission. In the case of filamentary cathodes this is achieved by using a long wire folded several times and with indirectly heated cathodes the area of the emitting surface must be increased. The effect of these modifications is to produce a valve which differs considerably from the types so far discussed. For a small power triode ρ may be 10000 ohms and g as much as 8 mA/volt, giving an amplification factor of 8. A triode used for power amplification always has a small amplification factor and a large input is therefore necessary to develop the maximum output. Power pentodes are more satisfactory in this respect, as may be seen from the data set out in table 7/1. Though μ and ρ are smaller than for a pentode voltage amplifier, the amplification factor is still fairly large and a small input to the grid is sufficient to produce the full power output.

TABLE 7/1

CHARACTERISTICS AND OPERATING CONDITIONS OF AMPLIFYING VALVES.

Type	Description	V_a volts	I_a mA	V_s volts	I_s mA	μ	g mA/v	ρ ohms	Power outp't watts
6o8	triode	250	3·1	—	—	38	1·45	26000	—
6A3	power triode	250	60	—	—	4·2	5·2	800	3·5
6D6	pentode	250	8·2	100	2·0	1280	1·6	0·8 meg.	—
42	power pentode	250	34	250	6·5	220	2·2	0·1 meg.	3·0

✓**Distortion.** In voltage amplifiers the amplitude of the input voltage seldom exceeds one or two volts and by adjustment of the grid bias a portion of the characteristic which is very nearly straight over this range may be found. In the case of power amplifiers a considerable input voltage may be needed to obtain the maximum energy output and often it is necessary to use almost the whole of the grid characteristic. Distortion produced by curvature of the characteristic then becomes possible (fig. 7/8). A distorted sine wave may be resolved into a sinusoidal component at the original or fundamental frequency and a number of components known as harmonics having frequencies which are multiples of the fundamental frequency. The percentage distortion of an amplifier is

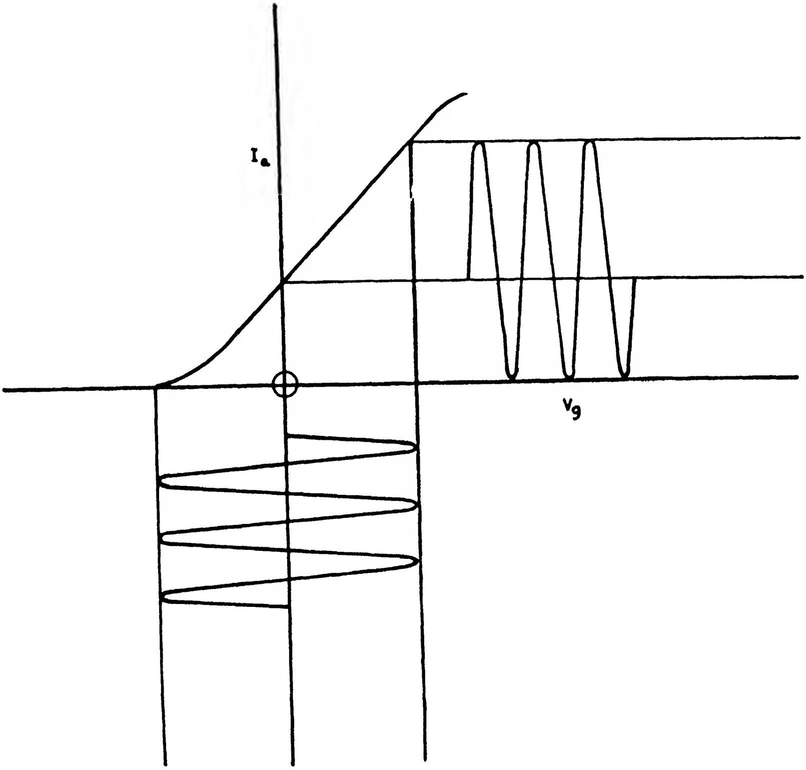


Fig. 7/8. Distortion may be produced by working on the curved portion of a valve characteristic. In the case illustrated here, the upper half of each cycle is amplified more than the lower half-cycle.

$$100 \times \frac{\text{power developed at harmonic frequencies}}{\text{power developed at fundamental frequency}}$$

Some distortion is always present since the dynamic characteristic of any valve is at least slightly curved; with triodes the distortion consists mainly of second and other even harmonics, while pentodes generate odd harmonics. Fortunately a distortion of less than about 5 per cent is not objectionable to the ear and amplifiers for speech or music are designed on this basis; the term ***undistorted power output*** indicates the output when the harmonic distortion does not exceed 5 per cent. or some other specified figure.

Impedance matching. Fig. 7/9 shows the relation between load impedance and power output for a typical pentode and it is apparent that there is an optimum value for the load, which should be adhered to fairly closely if a large output is to be obtained. For triodes the optimum load is usually about 2ρ and for pentodes about 0·1ρ. In audio-frequency amplifiers the loud-speaker may be connected as the anode load if its impedance is approximately correct, but it is more usual to use a coupling transformer which enables the impedance to be matched more exactly. A transformer is essential for the moving coil speakers now commonly used, which have an impedance of only a few ohms. If a transformer has a secondary/primary turns ratio of n and an impedance of Z_s is connected across the secondary winding, then the apparent impedance between the ends of the primary winding is Z_s/n^2—a statement which may be proved by using the theory of coupled circuits (cf. p. 53). Thus a pentode with an optimum load of 7500 ohms may be coupled to a loud speaker with an impedance of 3 ohms by using a step-down transformer with a turns ratio of $\sqrt{7500/3}=50:1$.

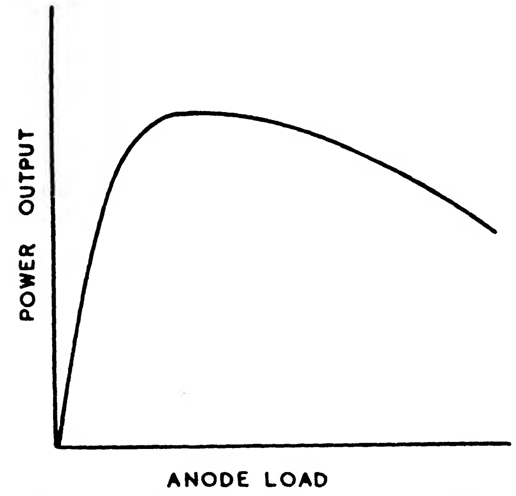
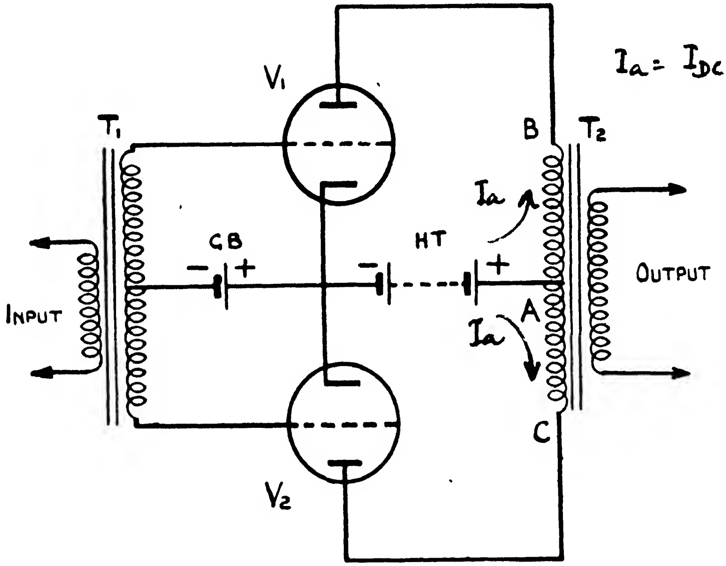


Fig. 7/9. Relation between power output and anode load for a typical triode.

Push-pull amplification. The push-pull circuit uses two valves and gives an undistorted power output greater than twice that of a single valve. The circuit is shown in fig. 7/10, and it operates as follows. The grids of V_1 and V_2 are supplied with equal voltages, 180 degrees out of phase, by connecting them to opposite ends of the secondary winding of the input transformer T_1. The anode currents of the two valves are therefore also in opposite phase. The current drawn by V_1 from the supply consists of a steady component with an alternating component superimposed on it (fig. 7/11) and flows through the primary winding of the output transformer T_2 in the direction AB. The anode current of V_2 consists of the same steady component with an alternating component of the same amplitude as in V_1 but in the opposite phase. This current flows through the transformer primary in the direction AC. Thus the two anode currents flow through the primary winding of T_2 in opposite directions, and their steady components cancel out, leaving no direct current in this winding. The alternating current flowing is the difference between the two alternating components and, since these are in opposite phase to start with, their combined effect is twice that of either of them alone. The



$I_a = I_{DC} + I_{AC}$

Fig. 7/10. The push-pull circuit.

disappearance of the steady current in the output transformer enables a core of moderate size to be used without the danger of magnetic saturation, and it may also be shown that second and other even harmonics, produced by the curvature of the valve characteristics, are eliminated. Any harmonic distortion present in the input to the push-pull stage is, of course,

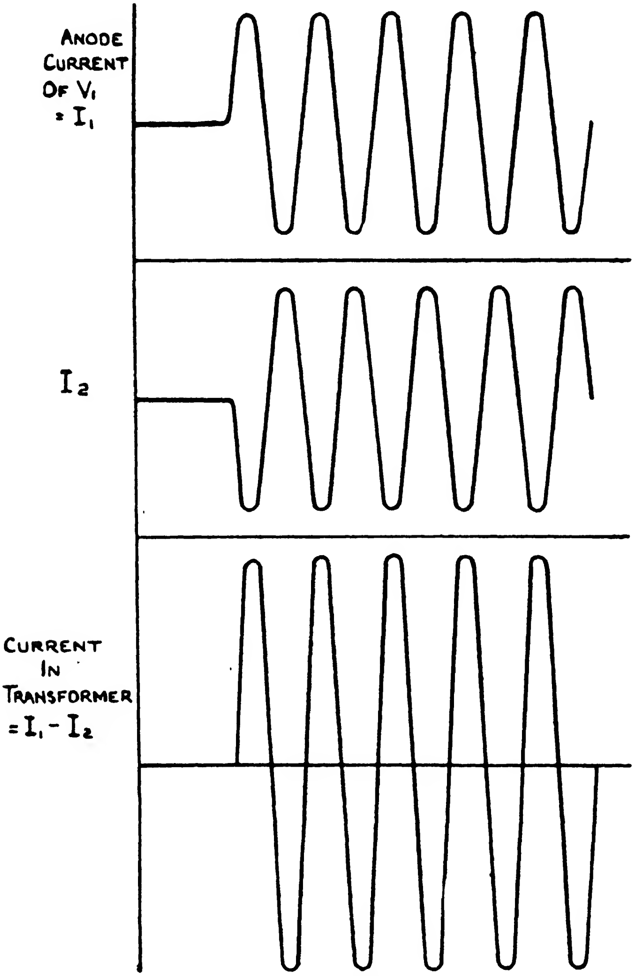


Fig. 7/11. The anode currents of the two valves in a push-pull circuit combine in the transformer primary winding and their d.c. components cancel out, while the a.c. components are compounded to give twice the output of a single valve.

amplified and appears in the output. If triodes are used in the push-pull arrangement, the harmonic distortion which they generate is nearly all in the even harmonics and it is therefore possible to range over a greater portion of the valve characteristic, obtaining greater power output, before distortion becomes objectionable. As has been explained, pentodes are more sensitive and thus often more suitable as power amplifiers, but as the distortion they produce lies chiefly in the odd harmonics, the push-pull arrangement does not eliminate it. The beam tetrode valve is a development of the pentode and has characteristic curves so shaped as to produce even harmonic distortion, thus making the valve suitable for push-pull operation. Its construction is shown in fig. 7/12. The electron stream is concentrated into two beams by the repulsion of the beam-forming plates connected to the cathode. If the anode-screen distance is critically adjusted, a marked potential minimum appears between these electrodes (fig. 7/13). The concentration of electrons here behaves in

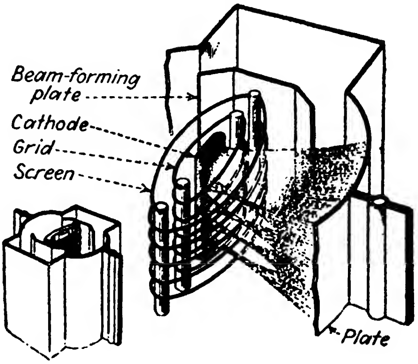


Fig. 7/12. The beam tetrode. (From Terman: Radio Engineers' Handbook, by courtesy of McGraw-Hill Publishing Company.)

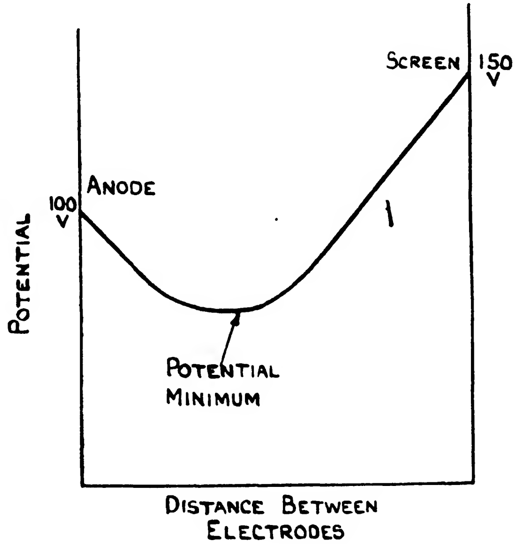


Fig. 7/13. Variation of potential between anode and screen in a beam tetrode with the distance between these electrodes adjusted to the critical value.

the same way as an additional grid at low potential and therefore serves the same purpose as a suppressor grid in returning secondary electrons to the anode. Information relating to typical beam tetrodes is set out in table 2 and anode characteristics are shown in fig. 7/14.

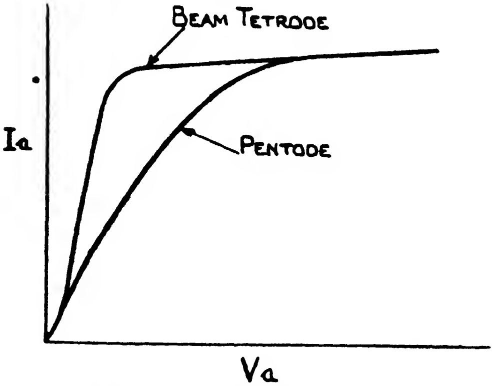


Fig. 7/14. Comparison of anode characteristics for pentode and beam tetrode valves.

Amplifier efficiency. The object of an amplifier is to use the steady electromotive force of a battery or other high tension supply to produce an alternating current in a load impedance. The steady current sent through the load is not effective from this point of view and the anode efficiency of an amplifier is defined as

$$\frac{\text{power developed by a.c. in load}}{\text{total power taken from supply}}$$

In the circuits so far considered, which amplify on the straight

TABLE 2

CHARACTERISTICS AND OPERATING CONDITIONS OF BEAM TETRODES.

Type	V_a volts	V_s volts	I_a mA	I_s mA	g mA/v	Power output watts	Optimum load ohms
6L6 (class A)	350	250	60	5	5·2	11	4200
6L6 (two valves in push-pull class AB)	360	270	90–200	5–16	—	47	3800
KT 66	250	250	85	6	6·3	7·2	2200
Pen 44	260	270	70	12	11	8	3000

portion of the characteristic, a considerable steady anode current flows and the efficiency is therefore low. A system which operates in this way is known as a *Class A amplifier* and it may be shown that its maximum anode efficiency is 50%—that is to say, only half of the energy supplied by the high tension source can be usefully employed. Actually a considerable expenditure of energy occurs in heating the anode and in sending direct current through the circuit, and the efficiency of a Class A amplifier is usually between 15% and 20%. This figure may be improved if the grid bias is reduced so that the grid becomes positive at the peaks of the input voltage. Anode current rises steeply in this region and the extent of the positive grid swing is limited by the onset of saturation. Class A amplifiers with the grid driven positive may have efficiencies up to 30%. Pentodes and beam tetrodes are more efficient, since the anode current in valves of this kind is determined by the screen voltage rather than by the anode voltage, and large current may be generated with a moderate anode potential; efficiencies of 30—35% may be obtained. Greater efficiencies may be obtained by the methods now to be described. Push-pull amplifiers may be operated in *Class AB*. Here the valves work on the curved portions of the characteristics and the anode current waveforms for a sinusoidal input voltage are shown in fig. 7/15, one half cycle producing a larger pulse of anode current than the other. The distortion so produced is elimin-

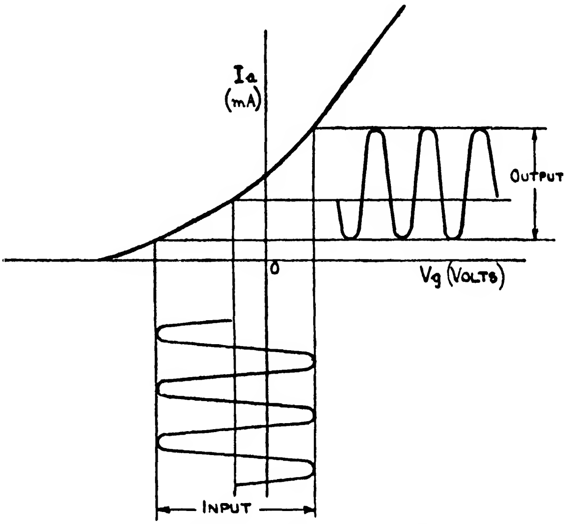


Fig. 7/15. In Class AB power amplifiers, distortion is introduced deliberately by operation on the curved portion of the valve characteristic. This distortion does not, however, appear in the final output if two valves are used in push-pull.

ated in the push-pull circuit and the output is again sinusoidal. The steady component of anode current is quite small, since a large grid bias voltage may be used and the efficiency of this arrangement is therefore greater than in Class A working. To obtain the maximum efficiency, which may be 45% or 50%, the grid bias and the input voltage should be adjusted so that the anode current stops altogether for part of each cycle and so that the grid is driven positive at the peaks of the input voltage.

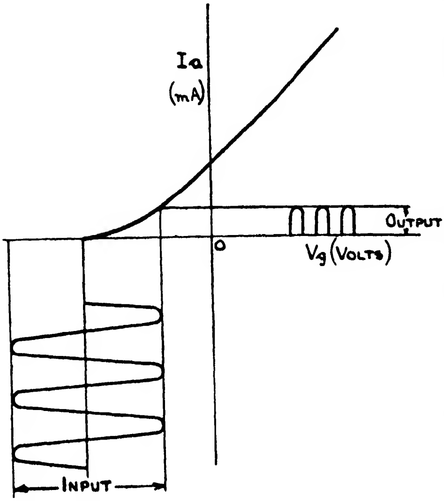


Fig. 7/16. In Class B amplification, the valve is biased to cut-off so that only half of each cycle is amplified. By the use of two valves in push-pull—one for each half of the cycle—this distortion is eliminated in the final output.

In *Class B* amplification, which is also used in push-pull circuits, the grid bias is adjusted approximately to the cut-off value and anode current flows in each valve for half of each cycle (fig. 7/16). The steady component of anode current in each valve is small and a theoretical maximum efficiency of $\pi/4 = 78{\cdot}5\%$ is attainable ; in practice, efficiencies of 60% may be reached.

Class C operation is suitable for the amplification of a single frequency or a narrow band of frequencies. One valve is used (a push-pull arrangement is also possible) and the anode load consists of a rejector circuit tuned to the frequency at which amplification is desired. The valve is biased to beyond the cut-off point and anode current flows for a small part of each cycle, forming a succession of short pulses (fig. 7/17). This current may be resolved into a steady component and a number of alternating components—one at the input frequency and others at harmonic frequencies. The anode load circuit offers a large impedance at the frequency which it is tuned to reject, but a smaller impedance at all others. Thus considerable power is developed in the load at the input frequency and negligible power at other frequencies. If the anode load circuit is tuned to the octave of the input frequency, this component of the anode current pulses is selectively amplified, thus producing frequency doubling.

For the amplification of a single frequency or a narrow band of frequencies, Class C operation is very satisfactory. As the steady component of anode current is very small, high efficiency may be attained. If the current pulses are of small amplitude the efficiency approaches 100% but the power output diminishes; with reasonable output, efficiencies of 70%-80% may be obtained.

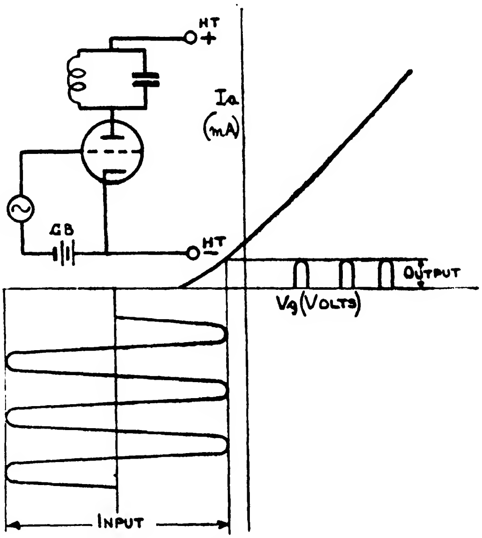


Fig. 7/17. In Class C amplification a tuned anode load is employed and the valve is biased to beyond cut-off.

In Class A amplifiers the average value of the anode current over a whole cycle is the same whether or not any alternating voltage is applied to the grid and a self-bias arrangement, such as a cathode resistor, keeps the valve at a fixed operating point. In Class AB, B, or C working the average value of I_a increases considerably when an input voltage is applied to the valve and it is not possible to obtain fixed grid bias in the same way. A separate grid battery or other source must be used in this case.

The multi-stage amplifier. The complete amplifier consists of one or more voltage amplifying stages and a power amplifying stage. All the valves are normally supplied with current from the same high tension supply and coupling between the stages may be produced on account of the resistance of the battery or other source. Fig. 7/18 is a skeleton diagram of a three stage amplifier. The internal resistance of the battery is common to all three valve circuits, which are thereby coupled. A portion of the alternating component of each valve current is therefore transferred by the common impedance to each of the other circuits. Thus some of the current generated by V_2 flows through the load resistance of V_1 and a small alternating potential is applied to the grid of V_2. This potential is in the opposite phase to the original input to V_2 and the gain of the stage is therefore reduced. A little degeneration produced in this way is not objectionable. If, however, a potential difference with the phase of the anode of V_3 is applied to the grid of V_2 via the load resistance of V_1 the gain of V_2 is increased, for the original input voltage to this stage is out of phase with the grid voltage of V_3 and therefore in the same phase as the anode potential of this valve.

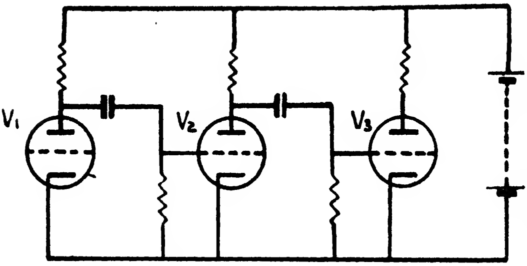


Fig. 7/18. Skeleton diagram of a multi-stage amplifier.

Regeneration is sometimes useful if it is under control, but if the coupling between output and input increases beyond a certain point the behaviour of the circuit changes as follows. When the input is increased by the feedback a larger anode current flows and a greater voltage is fed back to the grid circuit. This process is cumulative and after a short time the circuit ceases to amplify. It then generates continuous oscillations at a frequency depending on the values of the circuit elements and not on the original input; these oscillations are sometimes audible in amplifiers as a howling noise.

To eliminate *self oscillation* resulting from regeneration the alternating component of the anode current must be diverted from the high tension supply. This can be done by connecting a condenser C of large capacitance between the positive end of the anode load and the cathode (fig. 7/19) so as to offer a low impedance path to alternating currents. To make the reactance $1/Cp$ smaller than the resistance of the supply at low frequencies necessitates inconveniently large capacitances and to avoid this an additional resistance R_2 is connected in series with the supply. The steady component of I_a then flows through R_1 and R_2, while the alternating component returns to the cathode either by way of R_2 and the high tension supply or by way of C. If R_2 is 10,000 ohms and C about $1\mu F$, the reactance of C is smaller than the impedance of the alternative path at most frequencies. At very low frequencies—below about 30 c/s—it is difficult even by these means to prevent regeneration, which then becomes apparent as a "motor-boating" noise. The remedy here is to make the gain of the amplifier at low frequencies no greater than is necessary for reasonable fidelity of reproduction. In resistance-capacitance coupled amplifiers the inter-stage condenser should be of fairly small capacitance—about 0·05 μF—so that the gain is reduced at low frequencies (page 85). Reduction of amplification for frequencies below 50 c/sec. does not seriously affect the quality of reproduction for speech or music.

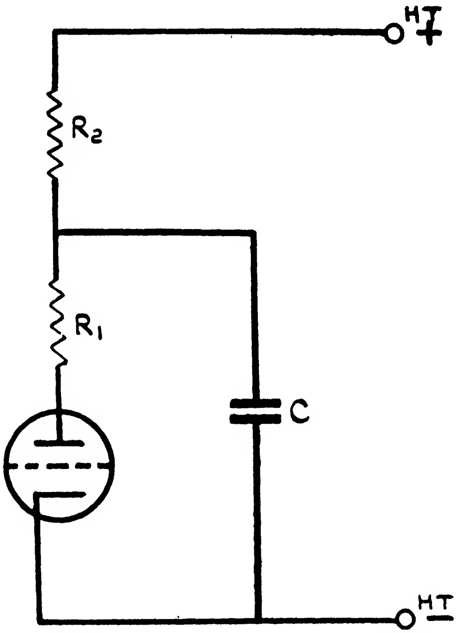


Fig. 7/19. Skeleton circuit for decoupling in a multi-stage amplifier.

Negative feedback amplifiers. It has been explained that the gain of an amplifier may be altered if some of the output is transferred to the input circuit as an additional potential difference. If the feedback is in phase with the original input, giving regeneration, the circuit becomes less stable, but amplifiers with negative feedback or degeneration are, in general, made more stable and have certain useful properties.

(1). Gain. Consider an amplifier which has a gain of M without feedback and a gain of A with feedback. If the input is 1 volt, the output is M volts without feedback and A volts with feedback. Suppose that a fraction β of the output voltage—180° out of phase with the input—is transferred to the input terminals (fig. 7/20). Then the apparent voltage input is

$$1 - \beta A$$

This is amplified M times to give an output of A volts.

Thus $M\,(1-\beta A)=A$ and $A=\dfrac{M}{1+\beta M}$ 7.1

For positive feedback the sign of β is changed and the gain is increased in ratio $1/(1-\beta M)$.

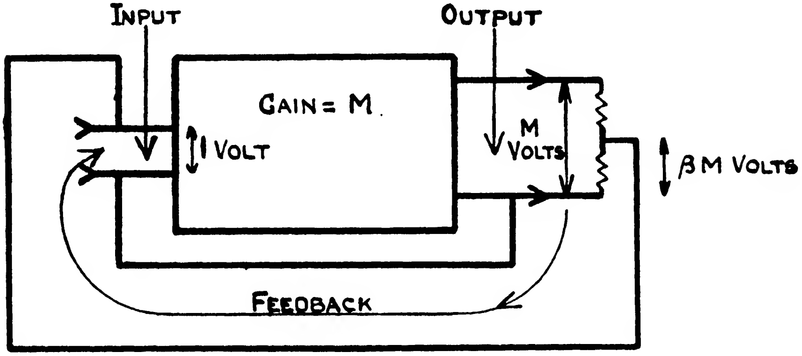


Fig. 7/20. Negative feedback or degeneration is obtained in an amplifier by tapping off part of the output and returning it to the input terminals.

(2). Frequency response. Suppose that an amplifier gives a gain of M at one frequency and N at some other frequency. Then, if negative feedback is applied, the ratio of the gains at the two frequencies is

$$\frac{M}{N}\cdot\frac{1+\beta N}{1+\beta M}$$

This ratio is always closer to unity than the ratio M/N and the gain at different frequencies is therefore more nearly uniform with feedback than without. It may also be shown that the harmonic distortion of an amplifier is reduced by the application of negative feedback.

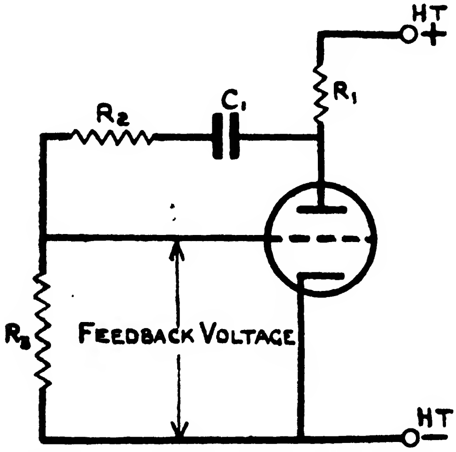


Fig. 7/21. In this circuit the alternating p.d. developed across the anode load R_1 is applied, by way of C_1. to R and R_3 in series. R_3 is in the grid circuit so that the p.d. across it is fed back to the input.

A common method of obtaining feedback is shown in fig. 7/21.

If the feedback fraction is so large that βM is much greater than unity the gain becomes approximately $1/\beta$ (eq. 7·1 above). β depends on the values of circuit elements such as resistances and capacitances and not to any appreciable extent on the properties of the valves. The characteristic constants of a valve change slightly with age and with variations in the supply voltages; values of resistances and capacitances on the other hand may be kept constant within very narrow limits. An amplifier with large negative feedback may thus be made to give a gain which, though small, does not vary with age or temperature and which is not affected if valves have to be replaced.

CHAPTER VIII

OSCILLATORS

THE effect of regeneration or positive feedback on the behaviour of an amplifier has already been noted (page 100). In general, it increases the apparent gain, since the voltage fed back by coupling from the output to the input is in the same phase as the original input. If the feedback fraction is large enough, an amplifier can supply its own input and then becomes an oscillator. For example, if in an amplifier with a gain of 100, one-hundredth of the output voltage is transferred to the input circuit, the original input is no longer necessary, and in fact ceases to control the behaviour of the circuit, which now functions as a generator of continuous oscillations. In general the feedback in an oscillatory circuit is only of the correct phase and amplitude at one frequency—if at all—so that if oscillations do occur, they are at a single frequency.

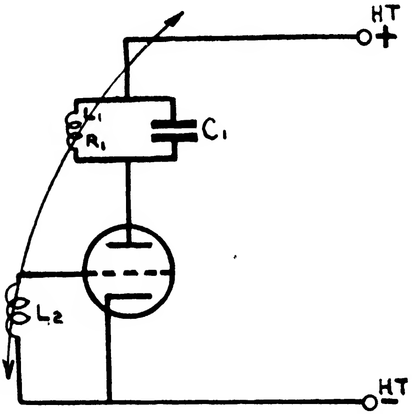


Fig. 8/1. Any amplifier will function as an oscillator if part of the output is fed back, in the correct phase, to the input. In this diagram, mutual inductance between two coils is used to supply the required coupling.

Regeneration. Regeneration in a single valve circuit may be obtained by magnetic coupling between the anode and grid circuits. In fig. 8/1 the anode load of the valve consists of a rejector circuit $L_1C_1R_1$ and an inductance L_2 is connected between grid and cathode. The operation of the circuit is briefly as follows. The current flowing through the valve is, for a variety of reasons, not quite constant. Fluctuations are produced by variations in the supply voltages and by the fact that the electrons reach the anode in small bursts rather than as a continuous stream.

The anode current therefore consists of a steady component with a ripple of complicated waveform superimposed on it. The ripple component may be resolved into a large number of sinusoidal components, of which one has a frequency equal to that at which the anode load resonates. At this frequency the rejector circuit behaves as a large resistance and the circulating current flowing through the inductance is at a maximum value (page 51). The voltage fed back to the grid coil is therefore greatest at this frequency. If the direction (clockwise or anti-clockwise) in which the coils are wound is correct, the potential difference fed back is out of phase with the anode voltage, so that after amplification it increases the anode current further, and the process continues. At frequencies other than that to which the anode load is tuned regeneration is slight because (1) the anode load behaves as a reactance, and the phase of the feedback voltage is no longer correct and (2) the amplification is smaller and the amplitude of the feedback is insufficient to maintain oscillations. The frequency of the oscillations is very nearly the resonant frequency of the anode circuit, though exact analysis shows that the characteristics of the valve are also involved.

It may seem that the oscillations should increase in amplitude until the valve becomes saturated. This is not the case, for if the amplitude increases beyond a certain point the grid becomes positive for a part of each cycle and therefore attracts electrons. When current flows in the grid circuit from this cause the effective impedance between the grid and cathode is lowered and a large impedance is reflected into the anode circuit, reducing the anode current and with it the amplitude. The amplitude of oscillation thus increases to the point at which the grid becomes positive at the peak of each cycle and no further.

Skeleton diagrams of various oscillator circuits are shown in fig. 8/2. In 8/2 (*a*) the resonant circuit is connected between grid and cathode and the anode load, consisting of an inductance only, is untuned or aperiodic. The mechanism of operation is similar to that of the tuned-anode circuit discussed above. The tuned anode-tuned grid oscillator is an obvious development of these circuits and is shown in 8/2 (*b*). The adjustment of operating voltages and circuit elements to obtain oscillation is less critical here than in circuits containing only one LC combination, and indeed, this circuit may oscillate without any coupling between the two inductances since at high frequencies there may be appreciable feedback through the inter-electrode capacitances of the valve.

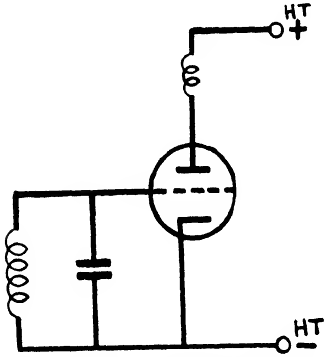


Fig. 8/2 (*a*) tuned-grid oscillator.

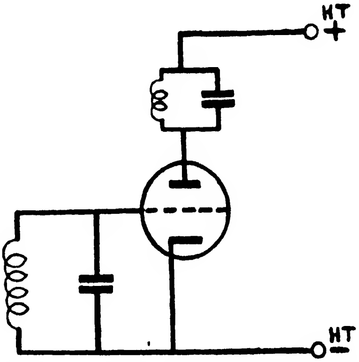


Fig. 8/2 (*b*) tuned anode-tuned grid circuit.

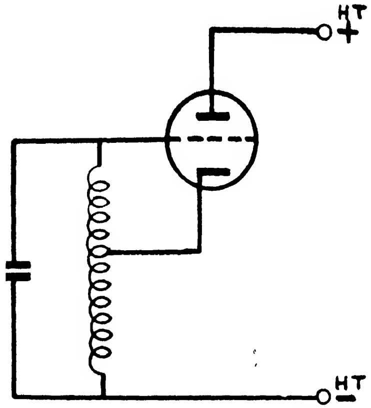


Fig. 8/2 (*c*) Hartley circuit.

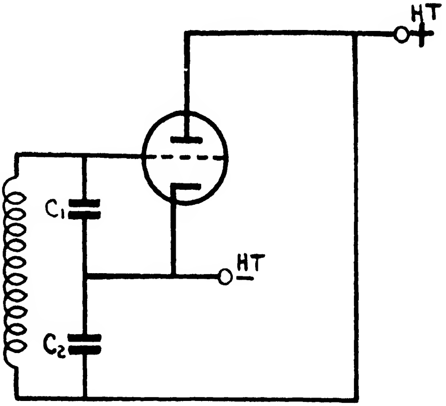


Fig. 8/2 (*d*) Colpitts circuit.

In the circuits so far discussed the amplitude of oscillation may be controlled in two ways—adjustment of the electrode voltages or alteration in the anode-grid coupling—neither of which is very convenient. From this point of view the Hartley oscillator (8/2 (*c*)) has certain advantages. Though only one tuned circuit is apparent, this arrangement in effect consists of two LC circuits, one in the negative high tension lead to the cathode and one connected to the grid. Coupling between them is provided by mutual inductance between the two portions of the coil and may be varied by adjusting the position of the tapping point. The Hartley circuit operates readily without critical adjustment and is useful at low frequencies since for a given total inductance it operates at a lower frequency than any of the circuits so far described.

The Colpitts oscillator (8/2 (*d*)), depends for its operation on feedback through capacitive coupling between the anode and grid. The condensers C_1 and C_2 act as a potential divider for this purpose. Points to be noted in connection with this circuit are:

(1) Adjustment of frequency is somewhat inconvenient since it necessitates simultaneous alteration to both condensers.

(2) The grid-anode and anode-cathode capacitances of the valve are in parallel with the tuning condensers C_1 and C_2 and the effect of inter-electrode capacitance may therefore be allowed for in calculating the constants of the circuits for any given case. For this reason the Colpitts oscillator is useful at high frequencies and it is possible to generate oscillations at very high frequencies without any external capacitance, using only the inter-electrode capacitances of the valve.

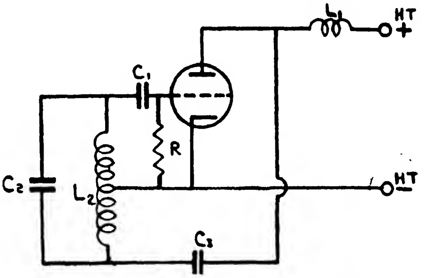


Fig. 8/3. Hartley circuit with power supply arrangements. This diagram shows a parallel-fed or shunt-fed high tension supply. The d.c. path for the valve current is from HT+ through the choke L_1, through the valve and back to HT–. The circuits in which the high-frequency currents flow are (1) from the anode through C_3 and the lower half of L_2 to the cathode (2) from the grid through the upper half of L_2 and back to the cathode. The inductances in these two circuits are tuned by the condenser. C_2, C_1 and R_1 are the grid bias network, C_3 prevents direct current from flowing in the R.F. circuit and L_1 prevents the oscillatory currents from flowing in the power supply circuit.

The circuits of fig. 8/2 become more complicated when the power supply and grid bias arrangements are included (fig. 8/3). If an oscillator is to be self-starting, the anode current must be large initially, so that the fluctuations superimposed on it are of sufficient amplitude to start cumulative regeneration, and the grid must therefore be at zero potential, or at most only slightly negative. Once the oscillations have begun, it is desirable to have a fairly large negative grid bias, so as to bring the operating point of the valve near the centre of its characteristic.

These conditions cannot be satisfied with fixed bias or with self bias and grid-leak bias is employed. This system operates in the following way (fig. 8/4). As I_a builds up, some electrons strike the grid and return to the cathode through the leak resistance R, making the grid negative with respect to the cathode and charging the condenser C at the same time. In subsequent cycles of the oscillation the grid and cathode of the valve form a diode rectifier with C and R as load, and if the values of these components are correctly chosen the output of the rectifying system, developed across R, is a fairly steady voltage, proportional to the rectifier input and therefore to the amplitude of the oscillations. Thus if for any reason the amplitude increases, the grid collects more electrons than usual and a greater negative bias is developed, with the result that the anode current falls again. Similarly if the anode current diminishes, the grid is biased less negatively. Grid leak bias maintains the amplitude of the oscillations at such a value that the peak of each cycle just drives the grid positive, and minor variations in the supply voltages or other operating characteristics do not affect the output. If fixed grid bias is used, a momentary reduction in I_a may cause a progressive diminution in the amount of feedback, resulting in the cessation of oscillations.

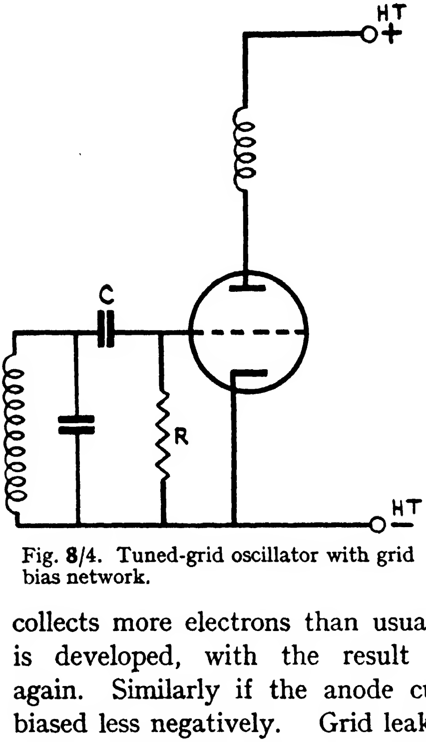


Fig. 8/4. Tuned-grid oscillator with grid bias network.

The automatic adjustment of the grid voltage depends on the discharge of C through R and the values of these components must be calculated with some care. If the time constant (CR, page 49) of the circuit is too large the charge on C cannot alter quickly enough to make up for sudden changes in amplitude and if, for example, the amplitude decreases, oscillations may cease before C has lost enough of its charge through R to reduce the bias. They will begin again when the charge on C has fallen to a sufficiently low value. This periodic interruption in the oscillations is known as *squegging* and is sometimes used intentionally when it is desired to produce short bursts of current, or pulses. If the time constant CR is too small, the grid bias voltage developed may not be sufficiently large to control the output amplitude effectively.

Frequency stability. The frequency of an oscillator varies for a number of reasons, among which may be mentioned changes in supply voltages and valve characteristics and changes in the properties of resonant circuits. Frequency drift is less likely if the tuned circuits have high Q values, which may be obtained by careful design and insulation of the inductances and condensers. The energy developed by an oscillating circuit must be delivered to a transmitting aerial or to the input of a power amplifier. This transfer is usually accomplished by inductive coupling between the circuits and if one of them has a large Q the coupling must be very loose to avoid complications caused by reflected impedance. It is therefore not possible to draw a great amount of power from an oscillator designed to give good frequency stability and it is common practice to limit the output to a few watts and to follow the oscillator with a power amplifier. In this way, small valves may be used for the oscillator, as the amount of heat to be dissipated is small and control over the temperature of the circuit elements is simplified. The frequency of an oscillator may be made almost independent of the

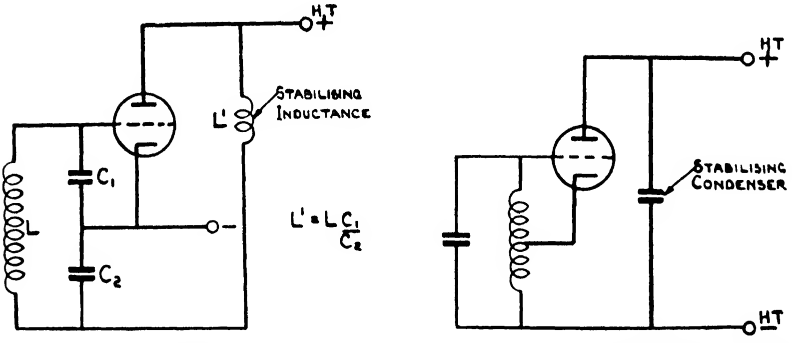


Fig. 8/5. Stabilised oscillator circuits, in which the operating frequency is independent—or almost independent—of the valve characteristics.

valve characteristics by small modifications, in the form of a stabilising inductance or capacitance, to the circuit (fig. 8/5).

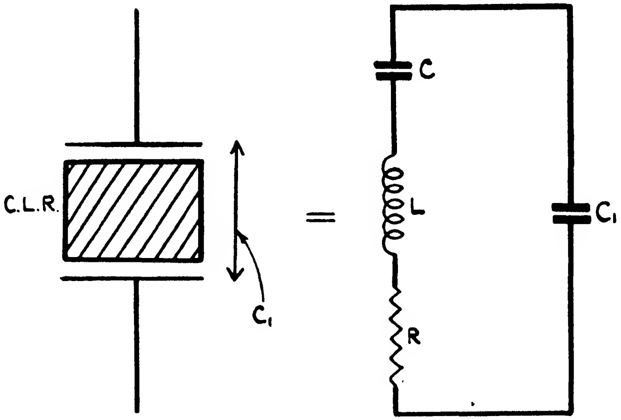


Fig. 8/6. Quartz crystal in holder, with its equivalent electrical circuit.

Piezo-electric oscillators. Crystals of quartz and certain other substances exhibit the piezo-electric effect. If a mechanical stress is set up between two opposing faces of such a crystal, a potential difference appears between another pair of faces. Conversely, application of a potential difference between one pair of faces results in a mechanical stress between another pair, and if an alternating potential difference is established between two faces, the crystal executes mechanical vibrations. Piezo-electric substances show electrical and mechanical resonance, that is to say, vibrations

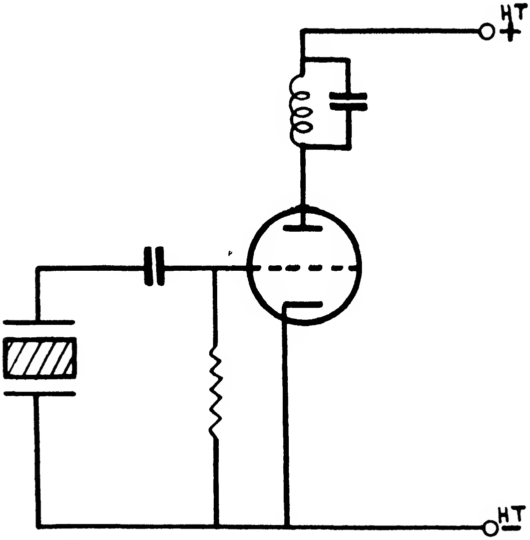


Fig. 8/7 (*a*). Tuned-anode, tuned-grid oscillator circuit using a quartz crystal as the grid resonant circuit.

are excited more readily at one frequency than at any other. A quartz crystal is electrically equivalent to an acceptor circuit (fig. 8/6) and may be used as an oscillator (fig. 8/7). In this arrangement, the mechanical oscillations of the crystal are maintained by electrical energy transferred by coupling between the crystal and the anode circuit. The anode load is usually tuned to the resonant frequency of the crystal but if frequency multiplication is desired it may be tuned to a harmonic.

The principal advantages of the crystal as compared with a conventional tuned circuit of coil and condenser are

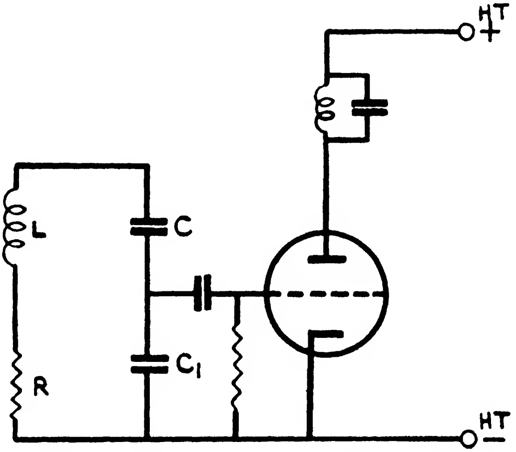


Fig. 8/7 (*b*). This circuit is equivalent to 8/7 (*a*).

(1) The electrical circuit represented by the crystal has exceptionally high Q and the frequency stability of the oscillations is therefore good.

(2) The mechanical and electrical properties of quartz vary only slightly with temperature and it is possible to make oscillators of which the frequency varies by less than two parts in 10^5 for a change in temperature of 1°C. As the temperature may be controlled by thermostats within close limits, quartz oscillators have been made with a frequency stability of the order of one part in ten million. It is impossible to obtain stability of this nature with circuits containing coils and condensers.

Power output up to 50 watts can be obtained from a crystal oscillator, but it is not advisable to draw too much energy if frequency drift is to be avoided and it is customary to take about five watts or less from the oscillating circuit.

CHAPTER IX

RADIO COMMUNICATION

It is a consequence of the electromagnetic theory developed by Maxwell that if an alternating current flows in a conductor, energy is radiated from it in the form of an electromagnetic wave. The wave consists of an alternating magnetic field together with an alternating electric field. The two fields, which alternate in intensity at the same frequency as the current in the conductor, are at right angles to one another and to the direction in which the wave moves ; electrically they are in phase. The velocity with which the wave travels is $c = 3 \times 10^{10}$ cm/sec. and the wavelength λ is related to the frequency n by the familiar relation $c = n\lambda$. For appreciable radiation to occur at a frequency f the length of the conductor must be of the same order as the corresponding wavelength, $\lambda = c/f$. At a frequency of 50 c/s $\lambda = 6 \times 10^8$ cm.—about 3700 miles and it is therefore not practicable to radiate at this frequency. At a frequency of 50 megacycles/sec. however $\lambda = 6$ metres and appreciable radiation may be obtained from a wire a metre or two in length.

Propagation of electro-magnetic waves. Radio communication is concerned with the transmission of intelligence from one point to another by means of electromagnetic radiation and frequencies above 150 kilocycles/sec. are normally used—the upper limit has not yet been reached, though broadcast communication is mainly confined to frequencies below 20 mc./sec.

Energy may be abstracted from the wave by placing a conductor in its path. The process is as follows. The alternating magnetic field which is a component of the wave surrounds the conductor and causes a periodic variation in the number of lines of flux cutting it. In accordance with the first Law of Electromagnetic Induction, an electromotive force is induced in the conductor and if there is a complete circuit an alternating current will flow, having the same frequency as the original current at the sending station. The essential requirements for transmission of signals by radio are therefore,

(1) a device for generating alternating current and,
(2) a conductor of suitable length from which energy may be radiated, and
(3) another conductor to abstract energy from the wave.

In travelling the wave spreads outwards in all directions and its amplitude is therefore progressively diminished. It is usually necessary to employ amplification circuits at the receiving station to increase the strength of the signal. The attenuation may be overcome to some extent by using *beam aerials* which concentrate the radiation in a desired direction, but as the surface of the earth is made up of conducting material a lot of energy is always dissipated by the currents induced in the ground over which the wave passes.

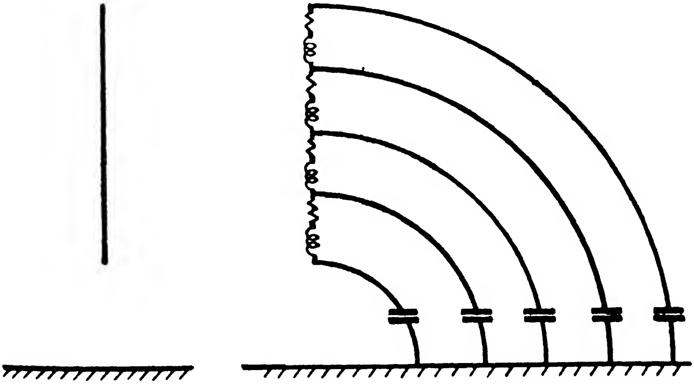


Fig. 9/1. A vertical wire supported above the ground is equivalent to a chain of LCR circuits and has a resonant frequency.

Aerials. Methods by which continuous oscillations of current may be generated have already been discussed (page 100) and it is now necessary to enquire how the energy may most efficiently be radiated. The conductor used for the transmission or reception of electromagnetic waves is called the *aerial* and usually takes the form of a wire or a rod placed above the ground. The aerial has inductance and resistance and therefore may not be connected directly to an oscillator without impairing the resonant properties of the tank circuit[*]; the transfer of energy is usually obtained by an inductive coupling between the tank circuit

---

[*] The tank circuit of an oscillator is the tuned circuit from which the output is obtained, and is usually the anode load of the oscillator valve.

coil and another coil connected to the aerial. For the maximum transference of energy by this arrangement it is necessary for the aerial to form part of a resonant circuit and it will be useful at this stage to examine its electrical properties more closely.

A wire supported above the ground has resistance and inductance, and capacitance to earth, all of which are distributed along its length. This system cannot be treated by the methods used for single circuits, in which the various circuit elements are each concentrated at one point, but an approximation to its behaviour may be obtained by considering a chain of circuits (fig. 9/1) each representing a segment of the wire. It is to be expected that an arrangement of this sort will show resonance and it is in fact found on analysis that a wire of length l remote from the ground behaves as a resonant circuit for wavelengths given by the relation

$$l = p\lambda/2, \; p = 1, 2, 3, \text{etc.}$$

In other words a single wire forms a resonant circuit if its length is $\lambda/2$, λ, $3\lambda/2$, etc., the simplest case in which the length $= \lambda/2$ is commonly used.

It is not often convenient to support an aerial several metres long so that both of its ends are remote from the ground. A system equivalent to a half-wave aerial may, however, be obtained by using a conductor $\lambda/4$ in length with one end connected to earth through the oscillator which supplies the energy to be radiated, fig. 9/2. The behaviour of this arrangement may be examined by considering the aerial to be split into small elements. Each element AB radiates in all directions and the total radiation is the sum of the effects of all the elements. Some of the radiation leaves in an upward direction and some in a downward direction; this latter component is reflected at the surface of the earth. In accordance with the laws of geometrical optics the reflected radiation appears to have come from an image element—A' B' (fig. 9/2). In this way it

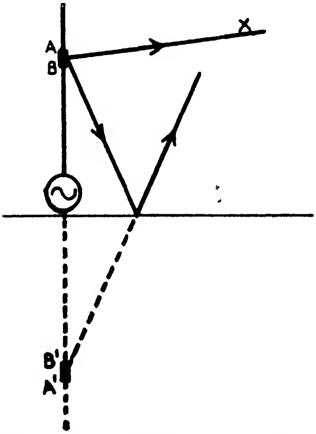


Fig. 9/2. Reflection of electromagnetic waves from the ground gives the effect of an image aerial below the surface.

becomes apparent that the total radiation is the same as that produced by a half wave aerial excited at the centre. The condition for this to occur is that the ground shall have a good reflecting surface, and in situations where the earth is not sufficiently flat it is customary to use an artificial earth made of wire netting which may be made quite flat.

Even a quarter wave aerial is sometimes inconveniently long, especially for military equipment where concealment and portability are important. It is possible to produce its electrical properties with a wire less than $\lambda/4$ in length, by adding inductance at the lower end. If the added inductance has the correct value, the aerial behaves as a resonant circuit, though the radiation from it is less than from a wire of length $\lambda/4$. The added inductance usually takes the form of a coil inserted between the aerial and the oscillatory circuit which supplies it with energy. If this inductance is variable, the resonant frequency of the aerial may be altered and a wire may therefore be used for radiating over a range of frequencies without alteration in length. The process of altering the effective length of the wire is referred to as aerial tuning and may be applied equally well to a half wave aerial.

The next process to be considered is the transference of energy from the oscillatory circuit to the aerial. This may be accomplished by direct coupling, fig. 9/3, a method used in small portable apparatus where the aerial is necessarily close to the transmitter. In general, however, the aerial should be remote from buildings and other large objects and therefore at some distance from the transmitter. If connection is made simply by two wires, radiation takes place from them and the aerial may receive very little of the energy produced by the oscillator. It is possible to transfer radio-frequency current from one point to another by means of specially designed systems of conductors in which radiation losses are almost entirely eliminated. Arrangements of this kind are known as transmission lines or feeders, and may be divided into two main types.

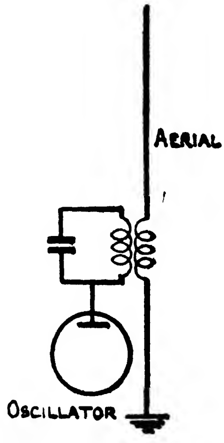


Fig. 9/3. A simple method of transferring energy from oscillator to aerial.

Tuned feeders or resonant lines. The tuned feeder consists of two parallel conductors, each having a length which is a multiple of $\lambda/4$, so that each forms a resonant circuit. The line may be connected to a half-wave aerial as shown in fig. 9/4. The currents in the two wires AB and CD are in opposite phase as A and C are joined to opposite ends of the coupling coil L', and the radiation from the wires therefore cancels out. The line may best be regarded as part of the aerial, folded double to eliminate radiation. It then becomes apparent that the inductance L' may be varied for operation at different frequencies, thus tuning the line and the aerial simultaneously. Tuned feeders are satisfactory at lengths up to 2λ or 3λ.

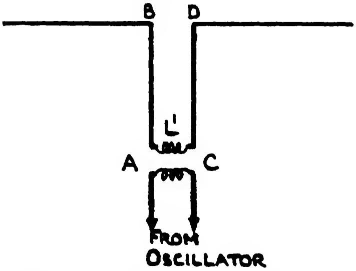


Fig. 9/4. Coupling by means of a tuned feeder.

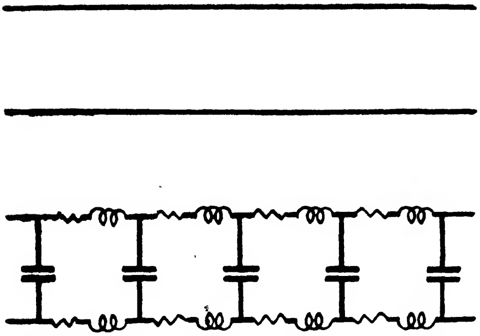


Fig. 9/5. A pair of parallel wires are equivalent to a chain of LCR circuits for radio-frequency currents.

Untuned feeders. Two parallel wires, represented in fig. 9/5 are equivalent for alternating current to a chain of circuits, consisting of the resistance and inductance of the wires and the capacitance between them. In general if energy is transferred from one point to another by means of this arrangement, radiation occurs from each wire. It has been shown above that if the wires have certain length the total radiation can be almost zero. The same result can be obtained in a rather different way with wires of any length. A system of two parallel conductors has a characteristic impedance Z_0 defined as

$$Z_0 = \sqrt{\frac{L}{C}} \qquad \dots\dots\dots\dots\dots\dots 9.1$$

where L is the inductance of each wire per unit length and C is the capacitance between the wires per unit length. It can be shown that if a load impedance of Z_0 is connected across one end of the pair, energy supplied at the other end travels along the line with practically no loss by radiation and is all absorbed in the load. When an aerial is resonant, standing waves of current and voltage, fig. 9/6, are associated with it. These may be compared to the standing sound waves set up in organ pipes and stretched strings. The current flowing in an element of the wire and the potential difference between that element and the generator both alternate at constant amplitude and it is therefore possible to calculate the impedance at each point of the aerial. This impedance is smallest at the centre where the current wave has an antinode and largest at the ends. In the case of a half-wave aerial remote from the earth the impedance at the centre is approximately 73 ohms, whatever the wavelength. At the ends the impedance may be as much as 1000 ohms. Referring now to the discussion of untuned feeders above, it is apparent that if a transmission line of characteristic impedance 73 ohms is connected to the centre of a half-wave aerial no radiation occurs from the feeder and all the energy supplied by the generator

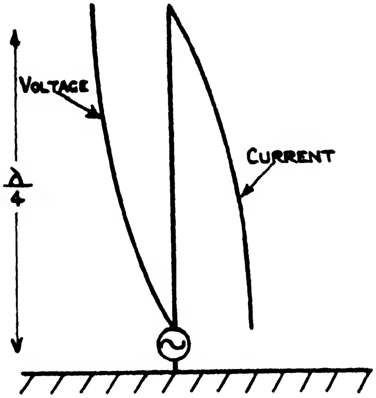


Fig. 9/6. Standing waves of current and voltage in a quarter-wave aerial. The current is a maximum near the source and decreases towards the free end of the aerial, because of leakage through the capacitance between the wire and the ground.

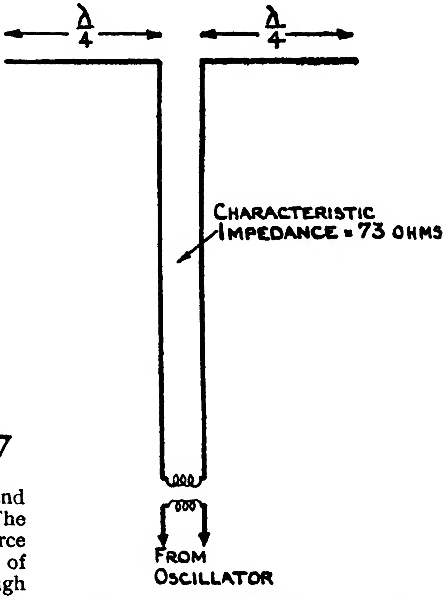


Fig. 9/7. Feeding a half-wave aerial with an untuned transmission line.

is delivered to the aerial, fig. 9/7. The feeder can have any length and the question of tuning it does not arise. The untuned feeder may be made in several forms. Flexible twin cable used for lighting circuits has a characteristic impedance of about 130 ohms. Loss of energy by absorption in the insulation restricts its use to short lengths—less than λ. Air insulated cable has smaller losses and is commonly used in two forms; (1) two parallel wires or tubes have characteristic impedance

$$Z_0 = 276 \log_{10} \frac{2D}{d} \qquad \text{.........................9.2}$$

where D = distance between centres of conductors
and d = diameter of each conductor.

Two quarter-inch tubes separated by 1½ ins. have a characteristic impedance of 300 ohms which is a common value for this type of cable; (2) coaxial or concentric cable is particularly useful at high frequencies where radiation losses are appreciable with other types of feeder. One conductor is a tube and the other is a wire running through it. The two conductors are spaced by washers of insulating material placed at intervals. In some cases the whole space between the conductors is filled with insulating material. The characteristic impedance of coaxial cable is generally between 50 and 100 ohms. It may be calculated from the equation

$$Z_0 = 138 \log_{10} \frac{b}{a} \qquad \text{............................9.3}$$

where b = inner diameter of outer conductor
and a = outer diameter of inner conductor.

It is possible to use a single wire as a feeder, the earth forming the second conductor. In this arrangement Z_0 is usually about 500 ohms. If the transmission line has a characteristic impedance in the neighbourhood of 73 ohms, it may be connected to the centre of a half-wave aerial or to the centre of a $\lambda/4$ aerial without serious loss of energy. If, as is often the case, the characteristic impedance of the feeder is considerably greater than this value, there are two methods by which it may be used to excite the aerial without serious radiation losses; (1) the feeder may be joined to the aerial at a point where the impedance is equal to Z_0 (p. 116 above). This method is commonly employed with single wire transmission lines; (2) a quarter-wave matching section may be used. A line $\lambda/4$ long with characteristic impedance Z_0 terminated by an impedance Z_1 offers at the other end an impedance

$$Z = \frac{Z_0^2}{Z_1} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 9.4$$

If one end of a quarter-wave line is joined to a half-wave aerial at a point where the impedance is Z_a the impedance at the opposite end is

$$\frac{Z_0^2}{Z_a}$$

and, by suitably adjusting the size and separation of the conductors forming the line, this may be made equal to the characteristic impedance of the feeder by which the aerial is to be excited. Thus to supply energy to a half-wave aerial from a transmission line with $Z_0 = 500$ ohms requires a quarter-wave line with characteristic impedance Z_0 given by

$$73 = \frac{Z_0^2}{500}$$

$$\therefore \quad Z_0^2 = 73.500$$

$$\text{that is, } Z_0 = \sqrt{73.500}$$

$$= 191 \text{ ohms.}$$

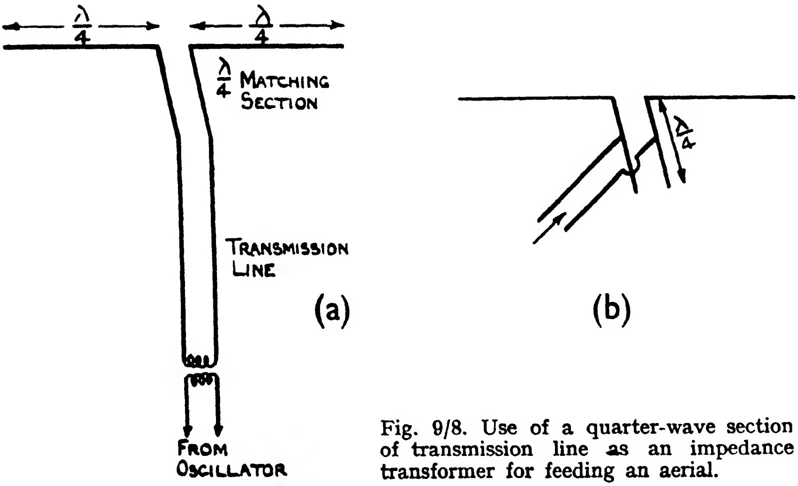


Fig. 9/8. Use of a quarter-wave section of transmission line as an impedance transformer for feeding an aerial.

The quarter-wave line used in this way is known as a matching section, (fig. 9/8 (*a*)), and it may be compared to the transformer used in impedance matching at audio frequency (see p. 92). Even if the characteristic impedance of the matching section is not correct the feeder may be connected to the aerial without serious loss if it is joined to the quarter-wave line at some distance from the open end, (fig. 9/8 (*b*)). The exact position of the points to which the feeder is to be attached may be found from tables or more usually by trial. Quarter-wave conductors used in this fashion are referred to as *Q bars*.

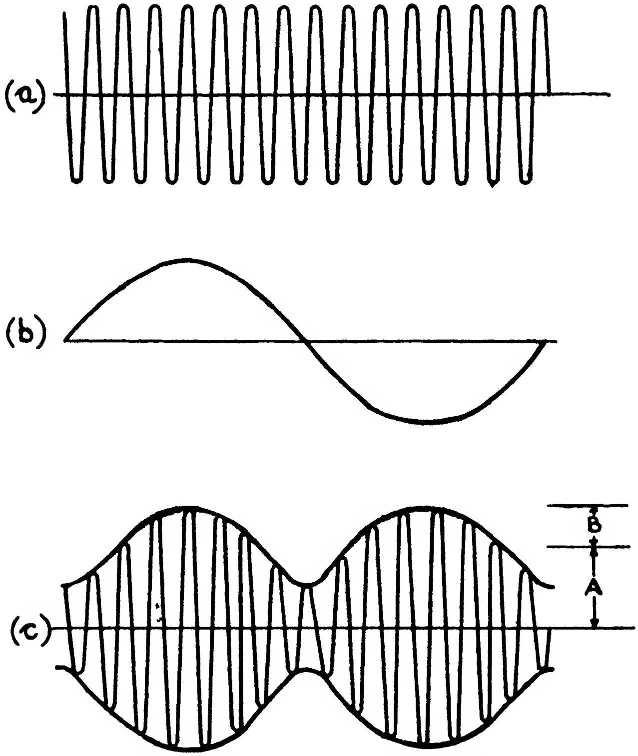


Fig. 9/9. Generation of a modulated wave.

Modulation. In the transmitting devices already mentioned the aerial current may be interrupted by means of a key at a suitable point in the circuit and signals in morse code may thus be radiated. The transmission of speech or music presents a more serious difficulty as these sounds are of low frequencies, seldom greater than 10,000 c/s. At such frequencies direct radiation is not practicable (p. 112), and it is necessary to modify the high frequency current flowing in the aerial in such a way that audio-frequency signals may be carried with the outgoing wave. This process is known as *modulation* in electromagnetic wave technology. Modulation is accomplished usually by superimposing variations on the amplitude of the high frequency *carrier* wave, these variations occurring at the frequency of the sound to be transmitted and having the same waveform. Modulation by varying the frequency or phase of the carrier is also practicable, but will not be discussed here. The radio frequency signal is separated from the high frequency carrier at the receiving station. Before discussing the methods by which amplitude modulation may be obtained we shall consider its effect on the waveform and other characteristics of a sinusoidal oscillation.

Fig. 9/9 (*a*) represents an unmodulated sine wave of amplitude A and frequency F. Fig. 9/9 (*b*), represents the audio frequency signal, with frequency f, which is to be superimposed on it. After modulation the high frequency wave has the shape shown in fig. 9/9 (*c*). The peak amplitude alternates at the frequency of the modulating signal and the extent of this variation depends on the amplitude of the radio frequency wave. If the mean amplitude of the modulated wave is A and if the maximum excursion above or below this value is kA then the equation of the envelope is

$$y = A + kA \sin 2\pi ft \qquad \ldots\ldots\ldots\ldots\ldots 9.5$$

and the modulated wave may therefore be represented by the expression

$$\begin{aligned} I &= (A + kA \sin 2\pi ft) \sin 2\pi Ft \\ &= A \sin 2\pi ft + \tfrac{1}{2}kA \cos 2\pi(F-f)t \\ &\qquad - \tfrac{1}{2}kA \cos 2\pi(F+f)t \quad \ldots\ldots\ldots 9.6 \end{aligned}$$

The modulated wave thus consists of three components:—

(1) the carrier, which is the original high frequency wave and is present whether or not modulation is applied.

(2) two components at frequencies $F+f$ and $F-f$, which are known as the upper and lower side frequencies. It is by means of these two waves that the audio frequency signal may be transmitted. The constant k is the *modulation index* and is usually expressed as a percentage. In fig. 9/9 (*c*) k is approximately 50 per cent. For the most effective radiation of the modulating

signal, k should be unity, (Fig. 9/10 (*a*)). If k is greater than one, (fig. 9/10 (*b*)), overmodulation results and the received signal is distorted. In the transmission of speech and music the intensity of the modulating signal varies and k will on the average be considerably less than 100 per cent. since at the loudest passages it must not exceed this value.

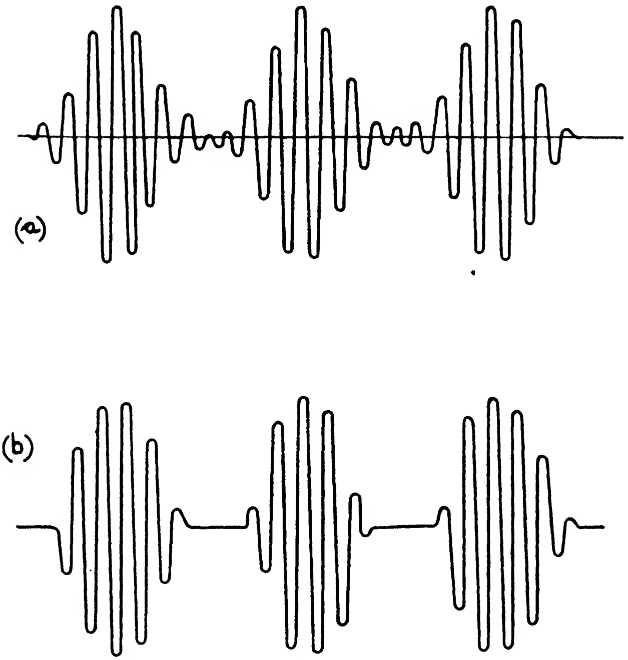


Fig. 9/10. (*a*) 100% modulation.
(*b*) overmodulation.

The energy radiated is proportional to the square of the aerial current. Thus the energy radiated in the carrier is $\propto A^2$, and as the amplitude of each side frequency component is $\propto kA/2$, the energy radiated here is proportional to $k^2A^2/4$. The total energy in the radiated wave at the side frequencies is therefore $\propto k^2A^2/2$. This can never be greater than one-half of the carrier energy or one-third of the total radiation.

In the transmission of speech and music, which are made up of a large number of frequencies, a correspondingly large number of side frequencies are produced. The groups of frequencies are referred to as *side bands*. Thus in the transmission of music with a frequency range from 50 c/s to 5,000 c/sec., the side frequencies for a 1,000 kc./sec. carrier extend from 1000.05 kc/sec. to 1,005 kc./sec. on the upper side, and from 995 to 999.95 kc./sec. on the lower side.

Methods of amplitude modulation. To produce amplitude modulation it is necessary to vary the aerial current in accordance with the modulating signal. Modulation may be applied either to the oscillator stage or to one of the amplifiers following it.

The current flowing through a valve may be altered by changing the relative potentials of the electrodes and modulation takes place by impressing an alternating potential of the appropriate value on the anode, grid or cathode.

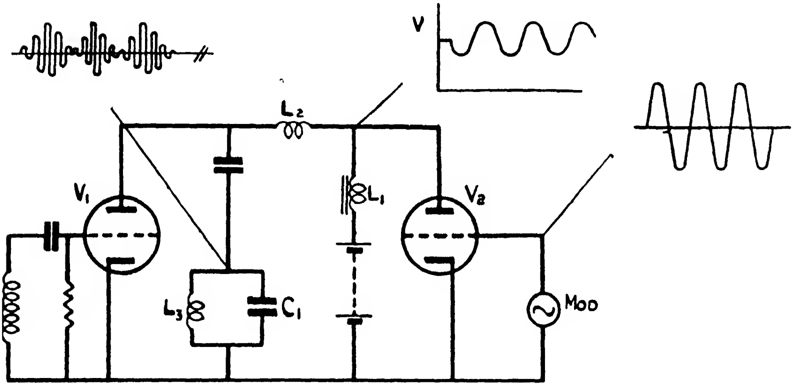


Fig. 9/11. Anode modulation. The oscillator valve, V_1 is supplied with high tension already modulated by the audio-frequency input to V_2. The oscillations generated by V_1 show a corresponding variation of amplitude.

Anode modulation. V_1 is the oscillator valve (fig. 9/11) with the tuned circuit L_3C_1 as its load. V_2 amplifies the modulating voltage applied between its grid and cathode. The anode load of this valve is the audio-frequency choke L_1 which has a large inductance. The anodes of the two valves are connected through the radio-frequency choke L_2. The anode current of V_2 has an alternating component of the same waveform as the modulating signal. Thus the voltage of the anode in this valve, and therefore the high-tension voltage applied to V_1 varies in accordance with the audio-frequency signal applied between the grid and cathode of V_2. As the amplitude of the oscillations is approximately proportional to the anode voltage, the potential difference developed across L_3C_1 shows amplitude modulation (cf. waveforms on fig. 9/11).

Grid modulation. Here the carrier and modulating voltages are applied in series between the grid and the cathode of an amplifying valve (fig. 9/12). The operation of the system is as follows. In this system the valve is supplied with a large fixed grid bias, bringing it well beyond the cut-off point. When the carrier potential is applied to the grid, the valve just reached the cut-off point at the peaks of the oscillations. If now the modulating signal is applied to the grid in series with the carrier, the valve is lifted above the cut-off point during the positive half-cycles of the modulating signal and the anode current alternates at the carrier frequency; the amplitude of these oscillations rises to a maximum as the modulating signal reaches its peak and then falls to zero. In the negative half-cycles of the modulating signal, the valve is driven still further beyond cut-off and no current flows. The anode load is a tuned circuit resonating at the carrier frequency and the periodic impulses delivered to it during the half-cycles in which current flows are sufficient to maintain the electrical oscillations.

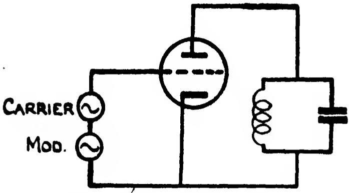


Fig. 9/12. Grid modulation. The valve is biased beyond cut-off and the carrier voltage supplied to the grid produces anode current only during the positive half cycles of the modulating signal.

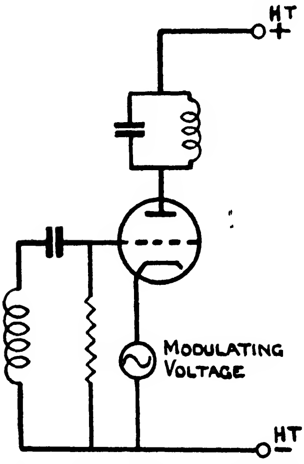


Fig. 9/13. Cathode modulation

Cathode modulation. In this arrangement, the modulating voltage is applied between the negative high tension terminal and the cathode of the oscillator or of a succeeding amplifier stage, (fig. 9/13). The method combines some of the features of the two systems described above. Thus if the cathode becomes positive with respect to the negative terminal of the supply, two effects follow; (1) the potential difference between anode and cathode decreases and the anode current therefore falls; (2) if the grid has a negative bias the potential difference between grid and cathode increases; if the grid bias is positive this potential difference decreases. In either case the anode current becomes smaller. Similarly, if the cathode is made negative with respect to the negative high tension line, by a variation in the modulating voltage, the anode current increases. The anode current thus undergoes fluctuations in amplitude corresponding in waveform to the modulating voltage.

Anode modulation leads to less distortion than either of the other methods and is commonly employed in large installations. Cathode modulation, on account of its simplicity, is useful in small transmitting sets.

Amplification of a modulated wave. In a modulated wave, the side frequencies are normally close to the carrier. For example, a 1,000 kc./sec. wave modulated at frequencies up to 5 kc./sec. has sidebands extending between 995 and 1,005 kc./sec. An amplifier tuned to the carrier frequency thus amplifies the side frequencies satisfactorily if the Q of its anode load circuit is not too high. For a carrier frequency of 100 kc./sec. the anode load of a tuned amplifier might have $Q = 50$ and the falling off in gain at the side frequencies is not serious. For lower frequencies the anode load must be flatly tuned, or aperiodic (untuned).

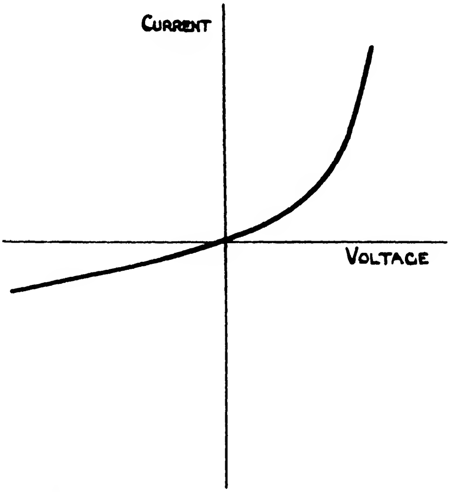


Fig. 9/14 (a). Characteristic curve of a crystal detector.

Demodulation. Demodulation or detection is the process whereby the audio-frequency intelligence is recovered from the modulated signal at the receiving station.

The crystal detector. The junction between two crystals of galena, or between one crystal and a metal point, conducts more readily in one direction than in the other. The relation between current and applied potential difference for such a device is shown in fig. 9/14. This arrangement does not obey

Ohm's law as the current through it is not in a constant ratio to the potential difference across it. The crystal detector is the simplest form of non-linear device. Application of a modulated voltage to the crystal produces a current of the form shown in fig. 9/14. This current may be resolved by Fourier analysis into a number of sinusoidal components, one of which has the same frequency as the modulating signal applied to the carrier at the transmitting station. (It will be remembered (p. 121) that a modulated wave consists of three components, none of them at an audio-frequency.) If the rectified current is passed through headphones, sound is produced corresponding to the modulating signal. The rectification may be rendered more effective if a condenser of small capacitance is connected as shown in fig. 9/15.

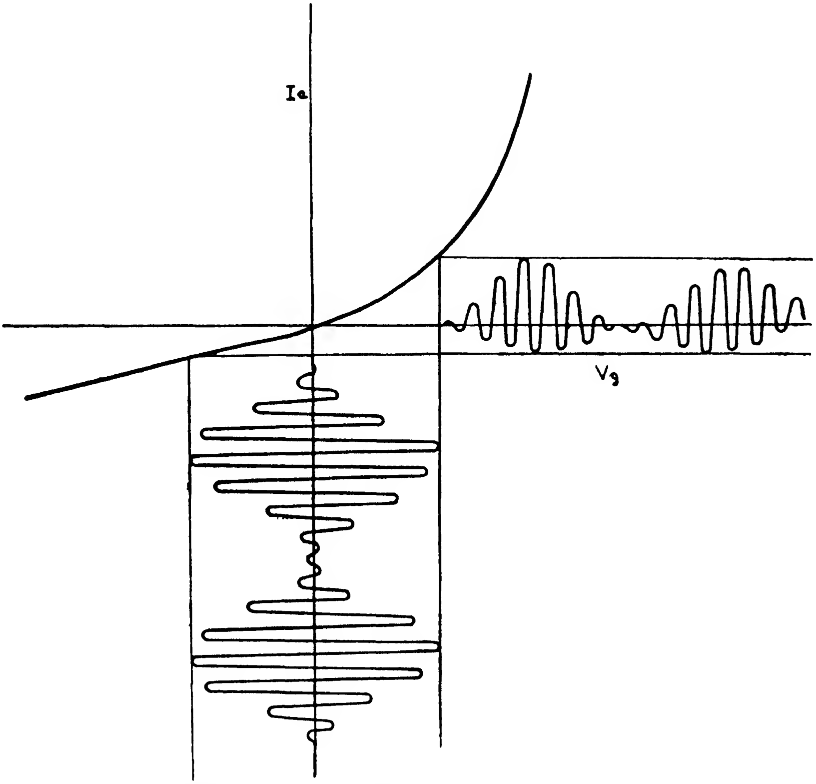


Fig. 9/14 (*b*). Application of a modulated wave to such a device gives a waveform of the kind shown, having a fairly large component at the modulation-frequency.

This condenser offers a small reactance to radio frequency currents, which are thus diverted from the headphones.

The diode valve as a detector. Another non-linear device is the diode valve (p. 21) and a circuit suitable for demodulation is sketched in fig. 9/16. The operation of the circuit is as follows. During the first quarter-cycle of the modulated input wave (*AB*) the condenser *C* is charged almost to the peak voltage of *B* (see p. 64). As the signal voltage decreases the condenser discharges through *R* until the potential difference across it is overtaken by the rising signal voltage during the second cycle. The condenser then charges (*CD*) and discharges during the interval between *D* and *E*. When the input amplitude is falling (*FG*) a similar process occurs though the interval between successive chargings of *C* is somewhat greater. The potential difference across the load

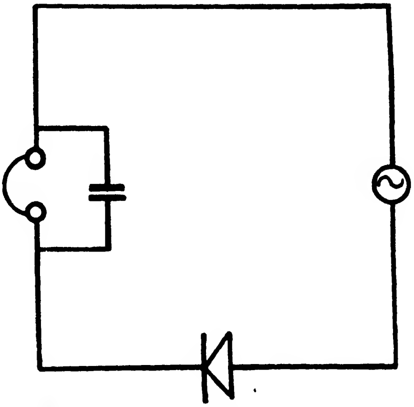


Fig. 9/15. Circuit of simple crystal receiver.

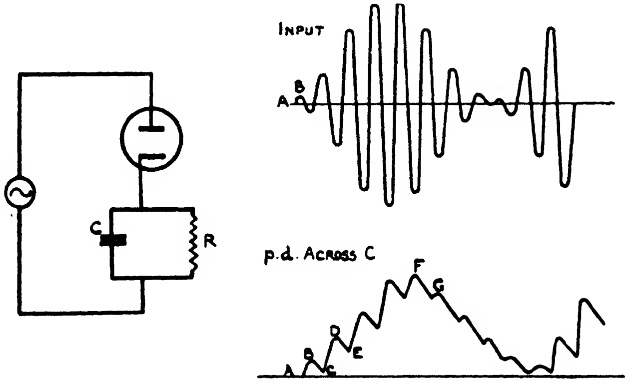


Fig. 9/16 (*a*). Operation of diode detector.

circuit *CR* is of the form shown in fig. 9/16, and has a considerable component at the modulation frequency with only a small radio-frequency residue. The values of *C* and *R* must be carefully chosen. If *R* is too small a considerable part of the output appears as a potential difference across the diode and is therefore not useful. If the time constant *CR* is too large the condenser discharges slowly through *R* and when the amplitude of the signal is decreasing, (fig. 9/16 (*b*)), the voltage across *C* decays along a line which does not follow the modulation waveform. In this case the output finally obtained is distorted. The optimum values of *C* and *R* depend on the modulation index and on the frequency of the modulating voltage as well as on the impedance presented by the diode. As an approximate rule it may be stated that the ratio *Z*/*R* should not be less than *k*, where *Z* is the impedance of the load (*C* and *R* in parallel) at the highest modulating frequency in the signal, and *k* is the modulation index of the signal. If this condition is satisfied, the output from the detector stage is not seriously distorted.

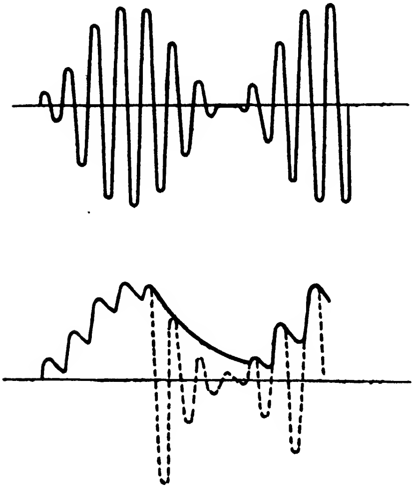


Fig. 9/16 (*b*). If the time constant (CR) of the load circuit is too small, distortion is produced.

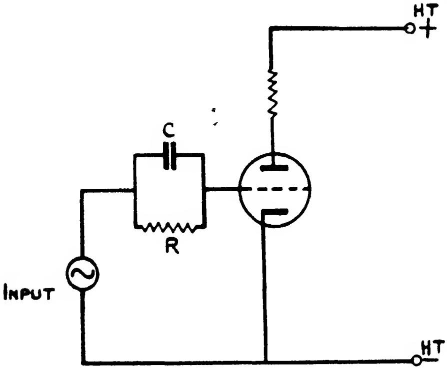


Fig. 9/17. A triode may be used for detection in this circuit, the grid and cathode forming a diode rectifier.

Grid detection. The grid and cathode of a triode (or pentode) may be connected as a diode rectifier, (fig. 9/17). The audio frequency output is amplified and appears as an increased voltage in the anode circuit.

Anode bend detection. It is possible to use a triode or pentode as a non-linear device, for the grid characteristic shows considerable curvature at each end. If a modulated voltage is applied between the grid and cathode of a valve, biased so that the

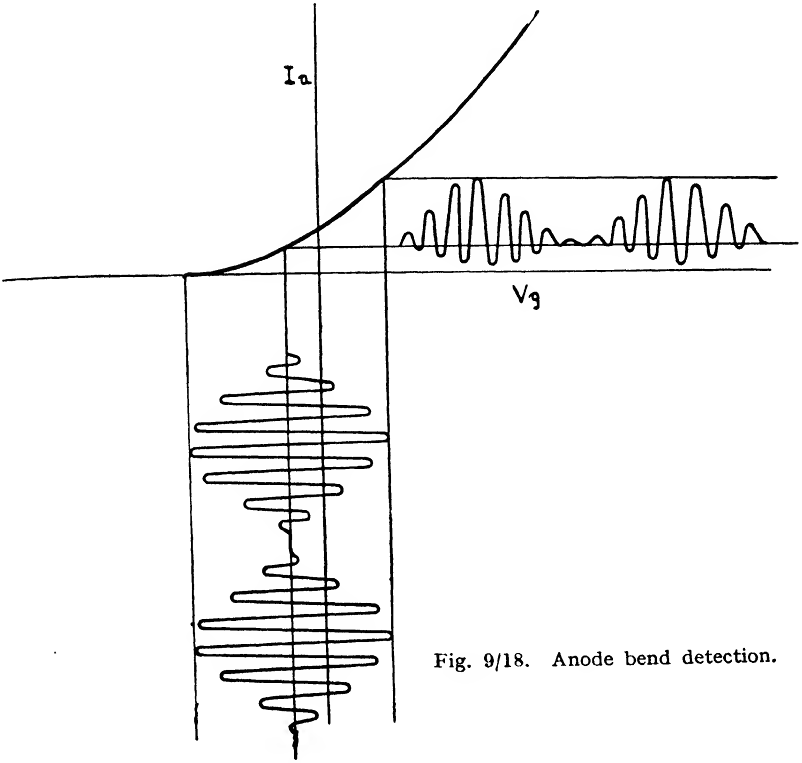
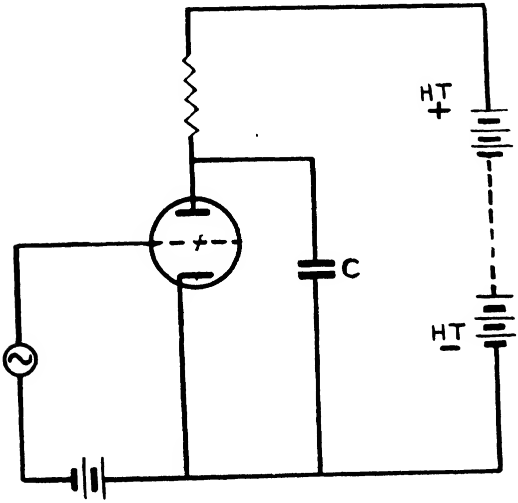


Fig. 9/18. Anode bend detection.



operating point is on the curved portion of the characteristic, (fig. 9/18), the voltage developed across the anode load is a distorted wave of the form shown and has a component at the audio-frequency. *C* is a by-pass condenser inserted to provide an alternative path for the radio-frequency component of anode current so that only an audio-frequency potential appears across the load. The anode bend detector is suitable for the reception of weak signals, on account of its amplifying property, but leads to more distortion than does the diode arrangement.

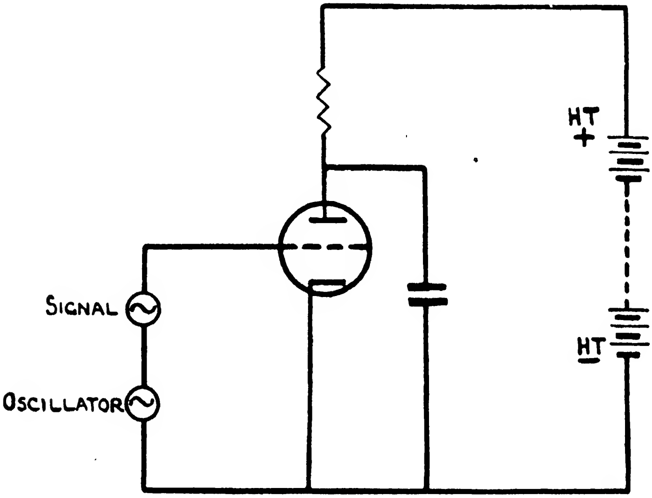


Fig. 9/19. Heterodyne detection. The modulated signal is applied to the grid of a valve, in series with the output of a local oscillator.

Heterodyne detection. This method is suitable for the reception of unmodulated carrier signals which are often used in the transmission of intelligence by morse code. In this case the output wave is interrupted by a key to give a succession of dots and dashes according to the letters of the message. The aerial signal is applied to the grid of a suitably biased amplifying valve in series with the output from an oscillator circuit in the receiver (fig. 9/19). The anode current of the valve in general contains components at the frequencies of these two signals and at their sum and difference. Thus if the carrier frequency to be received is 1,000 kc./sec. and the local oscillator works at 999 kc./sec., the anode current contains a component at 1 kc./sec., together with several radio-frequency components. Headphones inserted in the anode circuit may be used to pick out the audio-frequency. The process of mixing the signal with another oscillation is known as *heterodyning* and the principal advantage of this method of reception is its ability to resolve two signals with closely adjacent frequencies. Thus if two stations are operating on 1,000 kc./sec. and 1,001 kc./s.—a difference of 0.1 per cent.—the result of heterodyning signals from them with local oscillator at 999 kc./s. is to give audio-frequency components at 1 kc./s. and 2 kc./s.—a difference of 100 per cent., which is easily resolved by the ear.

Transmitters. The fundamental component circuits necessary for transmission and reception by radio have now been described and it remains to discuss the way in which they may be combined to form complete units.

The essential requirements for a transmitter are an oscillator stage in which the carrier wave is generated and a modulator stage in which modulation is applied. In the simplest apparatus both stages may be built around a single valve but in general they are separated by amplifying stages.

The oscillatory circuit. To generate a carrier of stable frequency it is necessary that the oscillator shall deliver only a small amount of energy—usually not more than five or ten watts, even in the largest transmitters. This output may be amplified by the methods already described either before or after modulation is applied; details are discussed below (page 132).

Modulation. Modulation may take place at either of two points in the transmitter. Low level modulation is applied at or near the oscillator and therefore requires the expenditure of small power only. In high level modulation the carrier is amplified almost to its final value and then modulated. In this case a greater proportion of the energy radiated has to be provided by the modulation amplifiers. Modulation in the oscillator stage itself is used only when simplicity of design is important, because the variation in high tension voltage which is involved leads to undesirable changes in frequency as well as to modulation of amplitude. If modulation is applied to an amplifier stage following the oscillator this difficulty does not arise, since the high tension voltage of the oscillator valve remains constant. Anode modulation in a class C amplifier stage is the most efficient method available and is widely used. Grid modulation is less satisfactory as the efficiency is low—about 35 per cent. By driving the grid positive this figure may be increased but greater distortion results. Grid modulation is therefore not suitable for broadcast transmission where high quality is required. Cathode modulation is the most suitable method for small transmitters since it does not require an extra valve. Its efficiency may be 50—60 per cent. and the distortion is less than that produced in grid modulation.

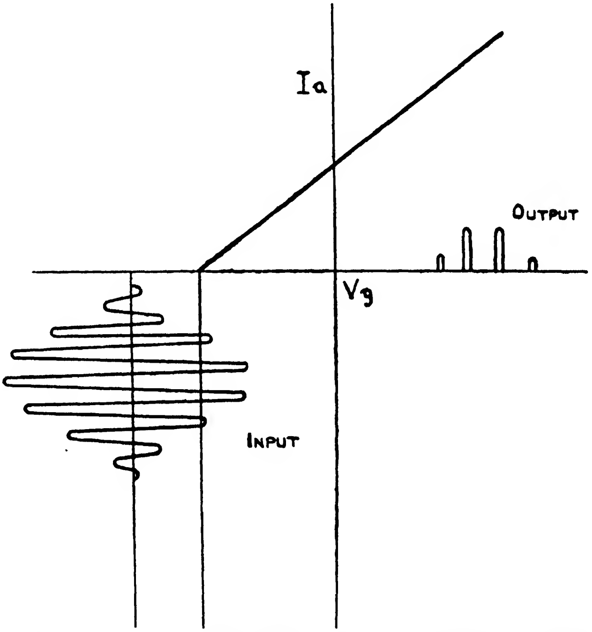


Fig. 9/20. A modulated wave cannot be amplified in Class C without distortion.

Amplification. The carrier wave may be amplified most efficiently by Class C stages using valves in parallel or in push-pull when large power is involved. After the modulator stage, amplification in Class C leads to distortion, as fig. 9/20 illustrates, and Class B is therefore preferable.

Frequency multiplication. It is easier to construct a stable oscillator at low frequency than for a high frequency, and for operation at several mc./sec. it is customary to generate the carrier at a low frequency, which is multiplied to the required final value by the arrangement now to be described. The anode load of a Class C amplifier consists of a tuned circuit, resonating at the frequency to be amplified. The anode current contains numerous harmonics of the input frequency and in normal operation the fundamental is amplified. If, however, the anode load is tuned to a harmonic of this frequency the potential difference developed across it is nearly all at the harmonic frequency, since a rejector circuit offers only a small impedance to frequencies other than that to which it is tuned. The frequency may be doubled or trebled satisfactorily in this way. Conversion to higher harmonics than the third is not practicable because the anode current contains only small components at these frequencies and the gain is therefore small. It is not possible to use frequency multiplying stages after the modulator as this process multiplies the side frequencies and thus leads to distortion.

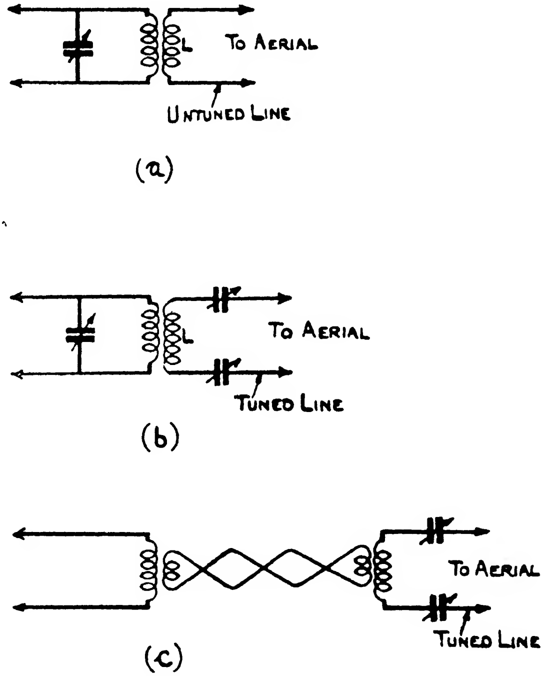


Fig. 9/21. Methods of coupling between oscillator and aerial.

Coupling between oscillator and transmission line. Various methods of transferring energy from the oscillatory circuit to the transmission line are shown in fig. 9/21.

This simple method (fig. 9/21 (*a*)) has the disadvantage that the coupling inductance L is added to that of the aerial. If L is kept small, close coupling is necessary to ensure the transfer of sufficient energy. To reduce the impedance reflected into the oscillatory circuit on this account the inductance L may be tuned as in fig. 9/21 (*b*), by the addition of condensers; L, C_1 and C_2, then form a circuit resonating at the frequency of the oscillator. The transfer of energy is thus facilitated and the impedance added to the transmission line is purely resistive. The line may be further isolated from the oscillator by the use of a link coupling (fig. 9/21 (*c*)) which may be several feet long.

Radio receivers. Methods of demodulation and audio frequency amplification have already been described and it remains to consider how the electromagnetic wave—or a voltage derived from it—may be applied to the detector circuit. The first problem is that of selecting the desired signal from among the many which reach the aerial. Aerials for broadcast receivers are usually of simple construction and it is only in special conditions that much care need be taken in their design and erection. The aerial wire forms part of a resonant circuit (fig. 9/22). The

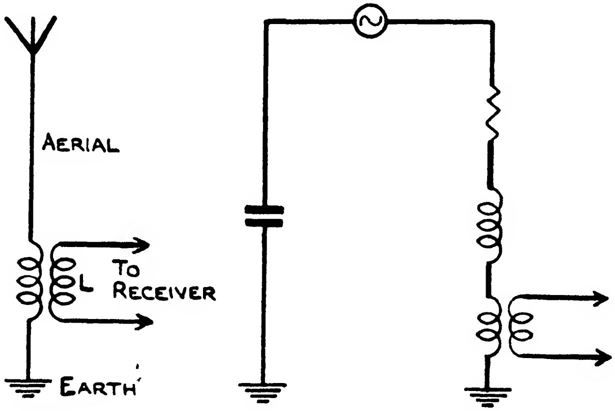


Fig. 9/22. The aerial circuit (left) is equivalent, for radio-frequency currents, to that shown on the right, which includes the e.m.f. induced in the aerial wire by electromagnetic waves passing it.

radiations from various transmitting stations within range each cause an e.m.f. to be induced in the aerial. If the aerial circuit is tuned to one of these frequencies, a magnified voltage is developed across L and may be passed to succeeding stages by magnetic coupling. The aerial circuit may be tuned by a variable capacitance but as the capacitance and therefore the resonant frequency of the circuit is affected by such factors as swaying of the wire due to wind, it is advisable to use an untuned or aperiodic aerial, coupled to a resonant circuit as in fig. 9/23 (*a*). Energy is then transferred most readily from the aerial at the frequency to which L_2C_2 is tuned. This is an acceptor circuit and the effect of the radiation reaching the aerial is to induce an e.m.f. in it. The potential difference across C_2 is greatest when the circuit resonates at the frequency of one of the signals reaching the aerial and by varying C_2 it is possible to select any of these signals. If the Q of the circuit is high a negligibly small potential difference appears across C_2 at frequencies other than that to which L_2 and C_2 are tuned. Other arrangements are shown in fig. 9/23, (*b*) and (*c*).

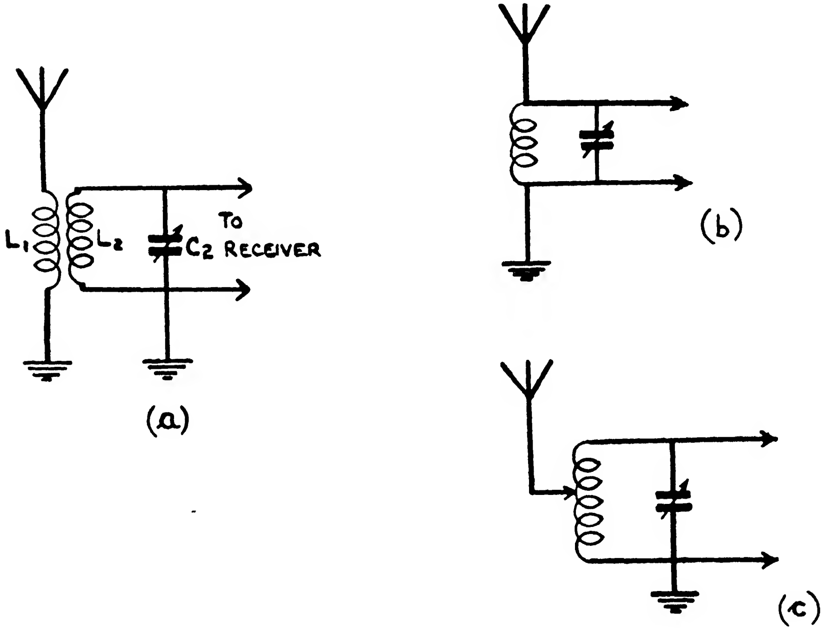


Fig. 9/23. Other forms of aerial circuit for a receiver.

Coupling between the aerial and the receiver must be loose if the circuits involved are to remain sharply tuned and it is therefore not possible to obtain a very large amplification of the signal voltage between the aerial and the grid of the first valve—generally no more than 10. Further amplification is needed before a demodulator can be operated satisfactorily.

The radio-frequency amplifier stage (fig. 9/24) usually has

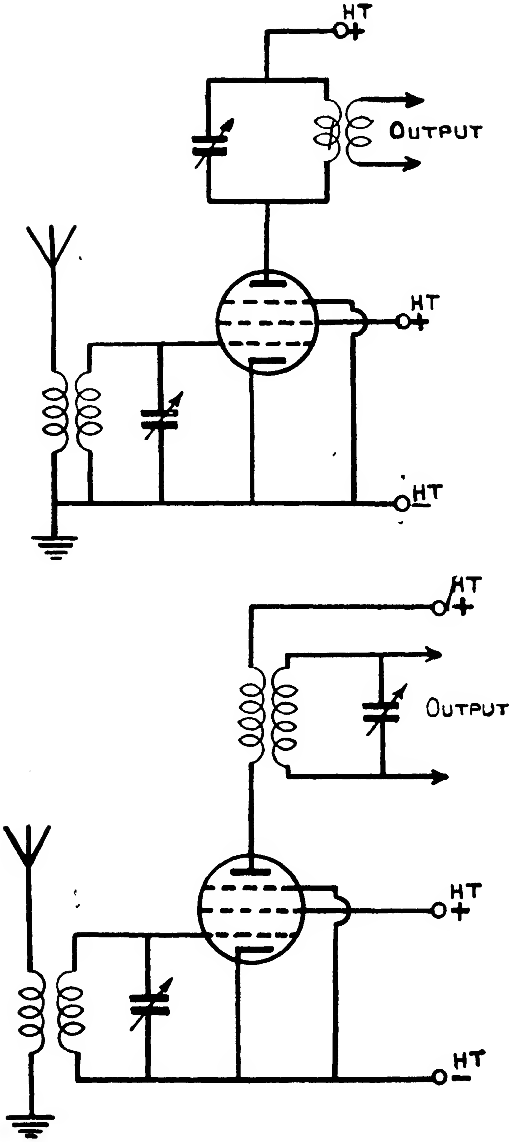


Fig. 9/24. Radio frequency amplifiers. Above: the anode load is tuned to the frequency of the transmission which it is required to receive. Below : alternatively, the tuning condenser may be connected as shown here, where it is effectively in the grid circuit of the following stage.

an anode load tuned to the frequency of the input by another variable condenser, and coupled by mutual inductance to the grid circuit of the succeeding stage. Alternatively, the anode load may be untuned, in which case it is coupled to a tuned grid circuit. Several radio-frequency stages may be used in cascade, but it is not usually practicable to employ more than two on account of the mechanical and electrical difficulties of tuning a number of circuits to the same frequency by separate variable condensers. Apart from this, the inter-electrode and other stray capacitances allow some degeneration by negative feedback and thus impose a limit on the gain which it is practicable to seek, especially at high radio frequencies.

The principal characteristics of a receiver are selectivity and sensitivity. *Selectivity* is the property of resolving two signals with a small frequency difference, and is usually expressed in kc/sec. Broadcasting stations are separated by 9 kc/sec. and selectivity of this order is therefore required. As the sidebands fill the whole of the 9 kc/sec. allotted to each station they will not be reproduced faithfully if the selectivity is increased beyond this figure. It may be observed here that as the limited range of frequencies available allows only a 9 kc/sec. band to each station, modulation frequencies above 4500 c/sec. cannot be transmitted

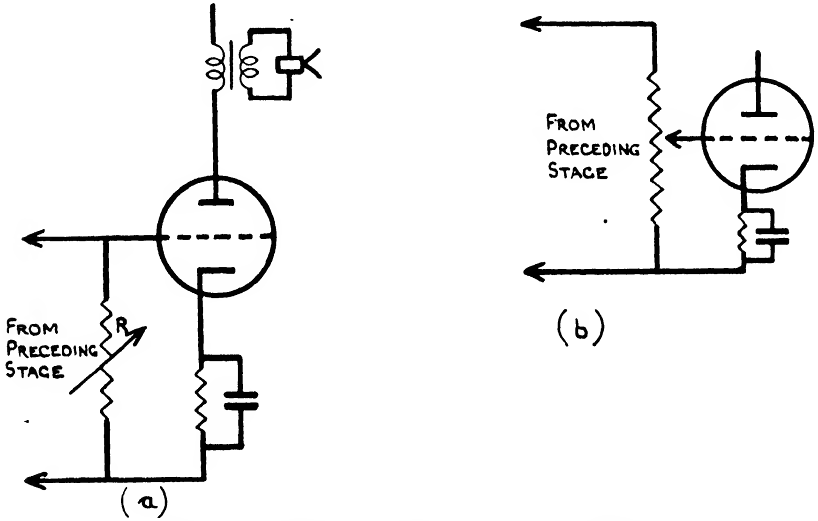


Fig. 9/25. Methods of volume control in a receiver.

without overlapping the sidebands of adjacent transmitters. The selectivity of a receiver depends on the number of tuned circuits which it contains and on the Q of each. As the Q of a circuit depends on the frequency, the selectivity is also a function of frequency.

Sensitivity is the property of responding to weak signals and depends on the amplification produced in the radio-frequency stages. The limit to the number of these stages which it is practicable to include in a receiver, mentioned on page 137, is important in this connection. Sensitivity is expressed as the voltage which must be induced in the aerial circuit to produce a stated output from the loud speaker.

Volume control. The output of a receiver may be controlled in the audio-frequency stages by using potentiometer arrangements such as those shown in fig. 9/25, but this form of control is not sufficient. Application of a large signal to the detector stage causes distortion in the resulting audio-frequency output and some control of the gain in the radio-frequency stages is necessary in all but the simplest receivers. It is not possible to use the methods shown in fig. 9/25, for resistances in series or in parallel with a tuned circuit would impair its resonant properties, and the difficulty is overcome by the use of a ***variable-mu valve.*** The negative grid voltage needed to produce cut-off

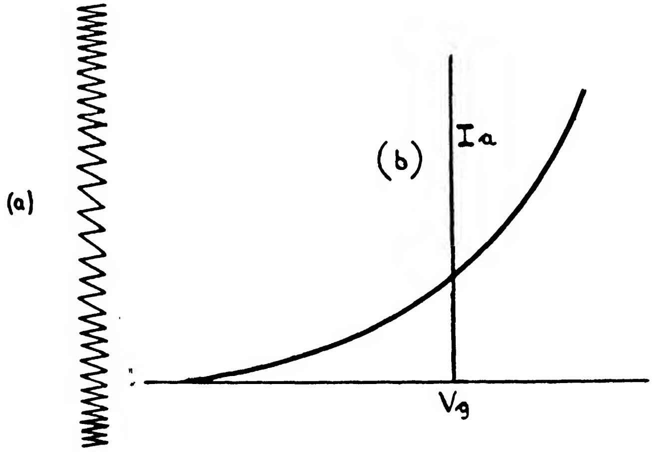


Fig. 9/26. Grid construction and characteristic curve for variable-mu valve.

of anode current in a valve depends on the spacing of the grid wires. If neighbouring turns are close together, a small bias suffices and if a greater separation is used, a correspondingly greater grid voltage is necessary to stop the flow of current through the valve. In the variable-mu valve, which is usually a pentode, the spacing of the grid is close at the ends and becomes more open towards the centre (fig. 9/26). At a small negative bias, the portion of the cathode opposite the closely-wound turns of the grid spiral cease to contribute to the anode current which therefore falls a little. As greater negative voltages are applied to the grid, less and less of the cathode emission is available for the anode current, which thus falls gradually, reaching zero at the cut-off voltage corresponding to the spacing in the centre of the grid. The characteristic curve (fig. 9/26) has a long gentle slope at its lower end, in contrast to the sharp bend obtained with a uniformly-wound grid. Over a range of a volt or two, the slope of the curve is fairly constant and amplification is therefore possible without much distortion. Since the amplification produced by the valve depends on the slope of the grid characteristic, it may be altered by changing the grid bias voltage, using the circuit shown in fig. 9/27, in which the variable cathode resistor is the gain control.

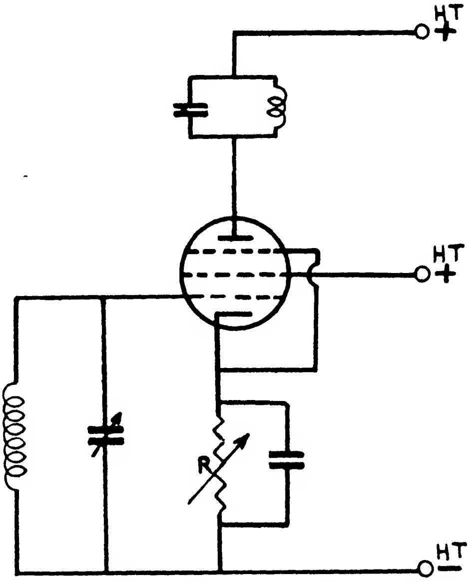


Fig. 9/27. Tuned grid radio frequency amplifier with variable-mu valve and volume control.

In the foregoing discussion, the term *variable-mu* is an abbreviation of *variable mutual conductance.* It is this quantity which depends on the slope of the grid characteristic curve, and which is adjusted by varying the grid bias voltage.

The superheterodyne receiver. It has already been explained that if more than three radio-frequency stages are used, mechanical and electrical complications ensue, and to make a really sensitive and selective receiver it is necessary to employ the supersonic heterodyne (or *superheterodyne*) principle (fig. 9/28). The signal frequency is amplified and converted to some lower fixed frequency (the intermediate frequency) by heterodyning it with an oscillation of variable frequency generated at the receiver. The stage in which this process takes place is known as the mixer or frequency changer. If the aerial signal is a sine wave, the output of the mixer is also a sine wave, but if a modulated signal is heterodyned, the result is a carrier at the difference frequency, with the original modulation still superimposed on it. Thus the effect of the mixer stage is simply to alter the carrier frequency to some lower pre-arranged value. Amplification may be performed more satisfactorily at the intermediate frequency because (1) it is lower than the signal frequency—465 kc sec. is a common value—and losses due to inter-electrode capacitance and other such effects are not so great as at higher frequencies. (2) As the frequency at which the intermediate frequency amplifiers operate is fixed, they contain no variable condensers and need no adjustment. It is thus possible to use several stages of amplification in this part of the receiver without encountering the difficulties mentioned above. The output from the final intermediate-frequency stage is demodulated by a diode detector and an audio-frequency amplifier drives the loud speaker. Fig. 9/28 shows the essential component stages of a superheterodyne receiver.

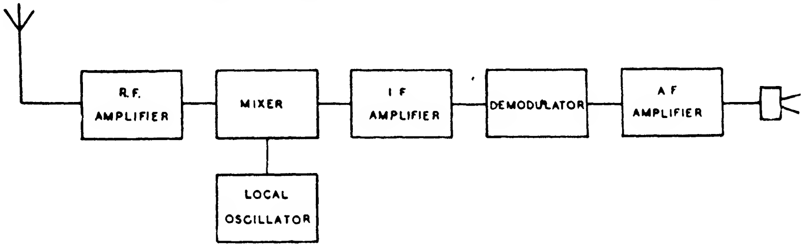


Fig. 9/28. Arrangement of the superheterodyne receiver.

In commercial receivers the signal is often applied directly to the frequency changer valve without amplification. Greater sensitivity and selectivity may be obtained if one or two radio-frequency stages are included. The design of these amplifiers has already been discussed (page 135).

Frequency changer. The heterodyne detector (page 130) is a frequency changer but is used only at low radio frequencies. At higher frequencies where the percentage difference between the signal frequency and the local oscillator frequency is less, the simple heterodyne arrangement does not operate satisfactorily on account of the interaction between the tuned circuits in the radio-frequency stage and those in the oscillator. Mixing is performed indirectly by the use of a multi-grid valve, applying the signal voltage to one grid and the output of the local oscillator to another. The anode current then contains components at several frequencies including the required intermediate

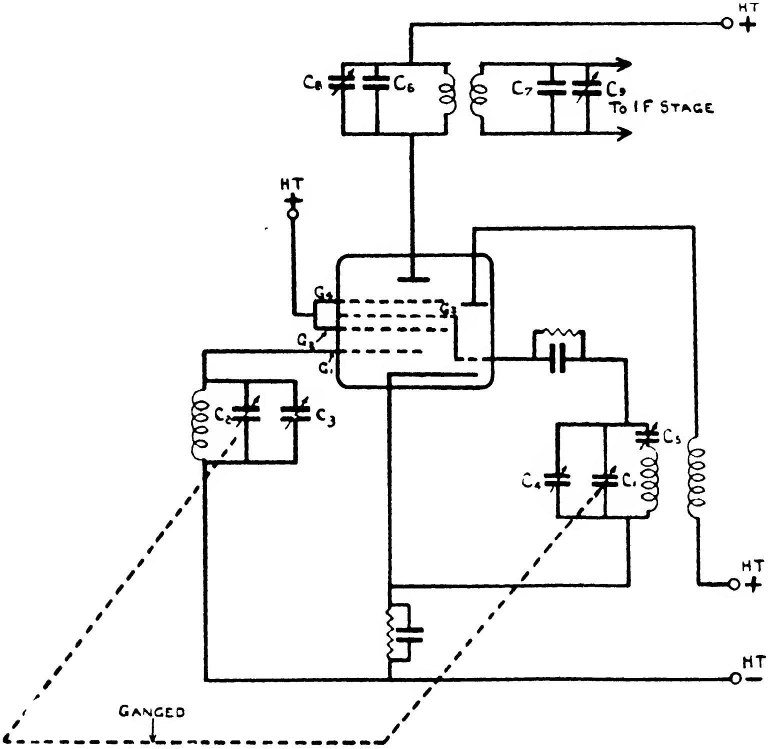


Fig. 9/29. Frequency changer circuit using triode-hexode valve.

frequency, which is the difference between the signal and oscillator frequencies. The triode-hexode valve is widely used for mixing, and a suitable circuit is shown in fig. 9/29. The triode portion is built into an oscillator circuit—in this case, of the tuned grid type—and is connected internally to the third grid of the hexode portion. The signal from the aerial or R.F. stage is applied to the first (control) grid and the anode load consists of a parallel resonant circuit tuned to the intermediate frequency. G_2 and G_4 form a screen around G_3, thus eliminating any electrostatic coupling between the signal and oscillator circuits. The intermediate frequency output is taken to the next stage by coupling with another tuned circuit, identical with that forming the anode load of the mixer hexode. The frequency of the local oscillator is adjusted by means of the variable condenser C_1 which is "ganged" (mounted on the same spindle) with C_2, tuning the input to the hexode control grid and, if necessary, with the tuning condenser of a preceding R.F. stage. As the capacitances C_1 and C_2 must vary in such a way that the resonant frequencies of the corresponding circuit always differ by a fixed amount (the intermediate frequency) small "padding" condensers are added to the circuit. By suitable adjustment of C_3, C_4 and C_5 it is possible to keep the difference frequency substantially constant over the whole tuning range of C_1 and C_2. The condensers C_6 and C_7 are fixed in value, since the circuits in which they occur are intended to resonate at a fixed frequency.

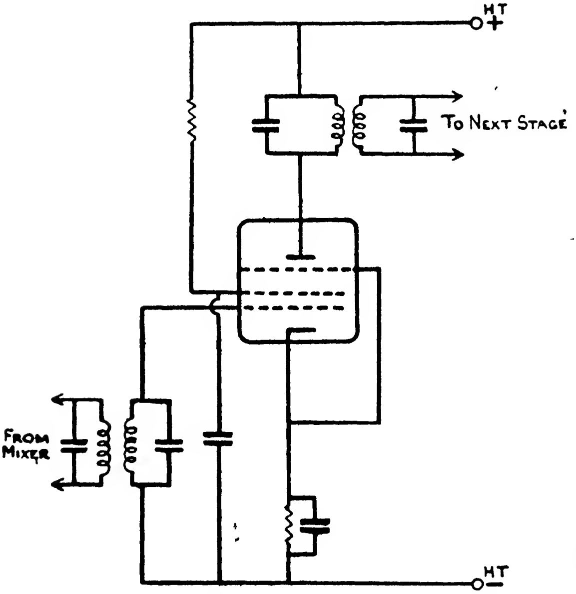


Fig. 9/30. Intermediate frequency amplifier circuit.

Small adjustments are made by means of trimming condensers C_8 and C_9.

Intermediate-frequency amplifier. These are of simple design as shown in fig. 9/30. Pentodes are normally used and the anode load consists of a parallel *LC* circuit tuned to the intermediate frequency and coupled inductively to the succeeding stage. In small receivers, one intermediate-frequency stage is commonly used but as many as four or five are practicable.

Demodulation. After these stages comes the demodulator which is in most cases a diode (page 127) and is followed by audio-frequency stages.

It will be observed that considerable amplification may be performed in the intermediate frequency stages which, operating at fixed frequency, are quite simple in design and construction. The superheterodyne method has other advantages of which the most important is increased selectivity. Suppose that it is desired to receive signals from a transmitter operating at 1000 kc/sec. but to reject signals from another transmitter at 1010 kc/sec. The frequency difference is 10 kc/sec. or 1%. If a superheterodyne receiver is tuned to the required station, the local oscillator frequency is 1465 kc/sec. (taking an intermediate frequency of 465 kc/sec., which is usual). The difference frequency produced by the interfering transmitter is $1465-1010 = 455$ kc/sec. and the frequency difference in the intermediate frequency is therefore $10/465 = 2{\cdot}2\%$. The two signals are therefore more easily separated than they could have been in a straight receiver.

CHAPTER X

APPLICATIONS OF PHOTO-ELECTRIC CELLS

THE principal characteristics of the various types of photo-electric cells under normal working conditions are summarised in Table 10/1. The applications of photo-electric cells are so numerous that no comprehensive account of the subject can be given here. The principles involved are, however, of an elementary nature, and discussion of a few photo-electric cell devices will serve to illustrate the methods employed. Photo-cells have really only one purpose, the conversion of luminous energy into electrical energy, and the uses which are made of this property fall into three main categories, according to the way in which the output from the cell is used.

TABLE 10/1

CHARACTERISTICS OF PHOTO-ELECTRIC CELLS.

Type	Optimum load res'tance	Current sensi'ty	Time lag	Battery needed	Colour response close to that of human eye	Linear relation between output and intensity	Usual connection
Vacuum photo-emissive	> 1 megohm	10–50 μA/lumen	No	Yes	No	Yes	thyratron or valve amplifier
Gas-fi'd photo-emissive	> 1 megohm	75–150 μA/lumen	Yes	Yes	No	No	thyratron or valve amplifier
Photo-voltaic	100–1000 ohms	100 μA/lumen	Yes	No	Yes	No	galvano-meter or relay
Photo-cond'ive	0·1–25 megohms	about 100 μA/lumen	Yes	Yes	Yes	No	relay or valve amplifier

1. **Conversion of luminous alternations into sound.** This is required in the reproduction of films. The processes by which

a sound track is impressed on a film and later extracted from it may be understood by reference to fig. 10/1 (*a*). The microphone current is amplified and used to modulate the intensity of a discharge lamp. A system of lenses throws an image of a narrow slit illuminated by this lamp on to the moving film.

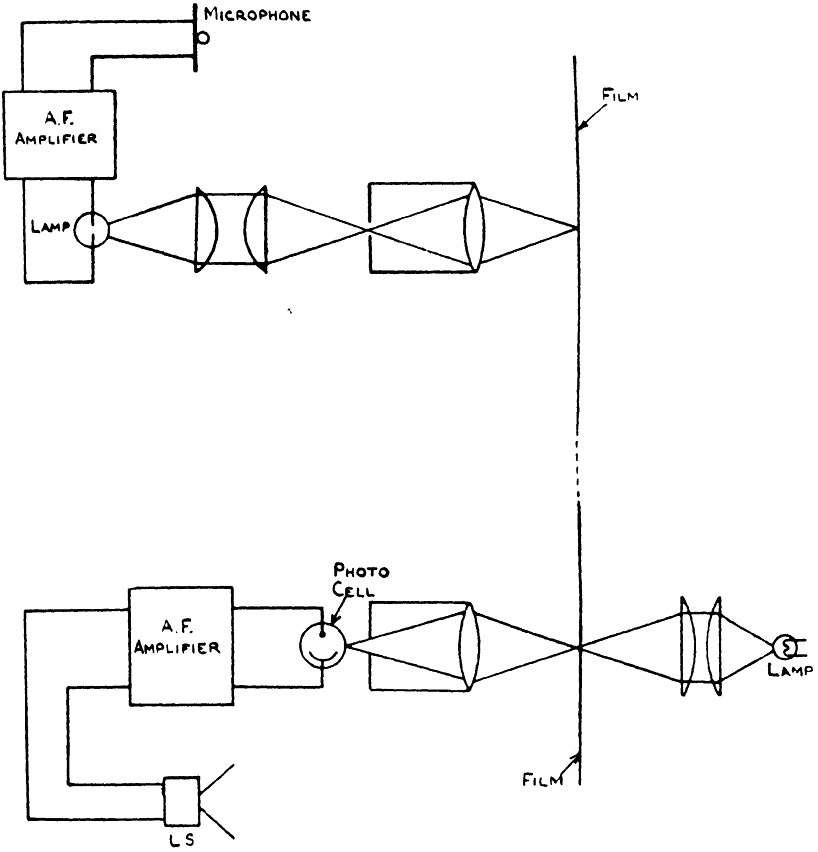


Fig. 10/1 (*a*). Preparation and reproduction of sound film.

The density of the image formed in this way varies as the intensity of the lamp changes and thus corresponds to the fluctuations of microphone current. The appearance of a variable area sound track is shown in fig. 10/1 (*b*). It is also possible to record by using constant illumination and varying the width of the slit as the microphone current alters (fig. 10/1 (*c*)). The process of reproduction is the same in either case and is briefly as follows. A narrow strip of the sound track is illuminated by a lamp and an image of it is formed by lenses on the sensitive surface of a photo-emissive cell. The amount of light reaching the cell is determined by the density or width of the track and the output from the cell therefore consists of an alternating current having the same waveform as the original sound.

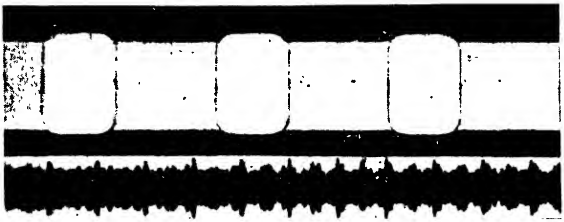


Fig. 10/1 (*b*). Variable area sound track.

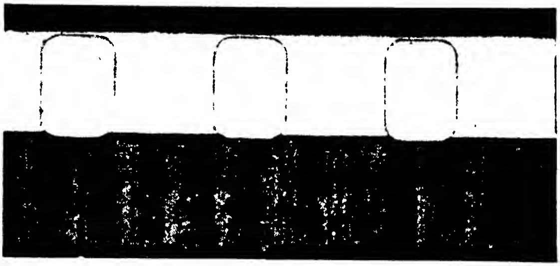


Fig. 10/1 (*c*). Variable density sound track.
(From E. G. Richardson: " Physical Science in Modern Life.")

An essentially similar process is performed by the iconoscope (page 195) of a television transmitter

2. **Operation of Relays.** The photo-emissive cell is widely used as a switch and the basic circuit arrangements are shown in figs. 10/2 to 10/4. In fig. 10/2 the anode current of the valve decreases when the photo cell is illuminated and the relay therefore operates if a beam of light directed towards the cell is interrupted. When light falls on the cell current falls in the

direction *ABCD*, making the grid end of R_1 negative with respect to the cathode, and the anode current of this valve diminishes. By adjusting R_2 the grid bias may be altered so that the available change of illumination will cause the relay to operate. If the relay is required to operate when the cell is illuminated, the connections of fig. 10/3 are necessary. Here a single high tension source may be used to operate both the cell and the valve. When the photo-sensitive surface is illuminated, current flows along the path *ABCD*, making the grid of the valve positive with respect to the cathode and thereby increasing the anode current. A modification of the circuit of fig. 10/3 for alternating current operation is shown in fig. 10/4. The valve can only pass current when its anode is positive and since under normal conditions a negative grid bias potential is needed, the grid and anode voltages are derived from opposite ends of a transformer winding, so that they are in opposite phases. C_1 is necessary to balance out the phase change between grid and anode voltages caused by the interelectrode capacitances of the valve and the photo-cell. The purpose of C_2 is to by-pass the alternating component of the anode current, leaving the steady component to operate the relay. Three illustrations of the way in which a photo-emissive cell may be used as a switch will suffice.

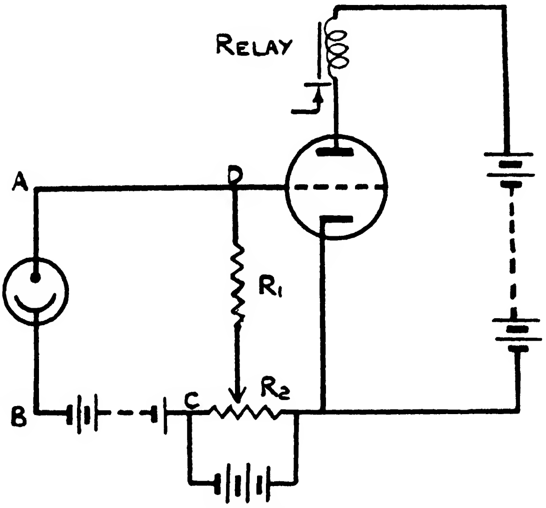


Fig. 10/2. In this circuit the relay closes when the illumination of the photo cell falls below a certain level. Such a system may be used to switch on street lamps or shop-window lighting when darkness falls.

(*a*) *Counting*. Objects which it is desired to count rapidly may be passed on a moving belt so as to interrupt a beam of light passing from a lamp at one side to a photo-cell opposite. If the circuit of fig. 10/3 is used, the relay closes every time the beam is intercepted and if an electromagnetic counter is connected to the relay contacts the number of objects passing may be recorded.

(*b*) *Sorting.* In making electric lamps it is generally necessary to reject those which give illumination below a certain level. The lamps are conveyed on a belt and connected to the electric supply shortly before passing a slit behind which is a photo-electric cell, incorporated in a circuit such as that of fig. 10/2. Mechanical devices may be arranged to remove lamps which do not produce sufficient illumination to operate the relay.

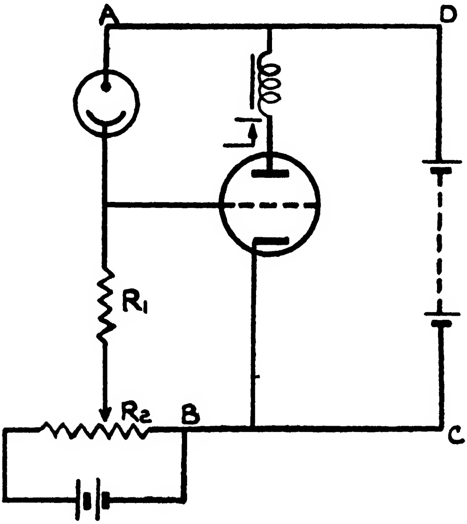


Fig. 10/3. Here the relay closes when the cell is illuminated.

(*c*) *Matching.* Small resistors and other electrical components made by mass-production methods are by no means uniform in value and it is necessary to reject those lying outside certain specified limits. In testing resistors for this purpose a Wheatstone bridge circuit is used, with a mirror galvanometer to indicate balance. If the unknown is not equal to the standard against which it is compared the beam of light from the galvanometer is deflected by an amount proportional to the difference between the two. A pair of photo-electric cells are arranged so that one or other of them is illuminated if the galvanometer beam reaches a deflection corresponding to the maximum permissible deviation above or below the standard value, and an arrangement of relays and mechanical devices may be used to remove the unsatisfactory components.

3. **Measurement of Light Intensity.** Photo-electric cells are widely used in photometry for the measurement or comparison of luminous intensities. To obtain an indication of the intensity of a light source it is necessary only to observe the photo-electric current which it produces in a suitable cell. The photovoltaic

type, connected to a galvanometer, is appropriate for this purpose. The response of the cell must be calibrated by the use of standard sources and care is necessary to ensure that variations in output arising from the use of sources with different spectral characteristics do not invalidate the results. A photovoltaic cell connected to a microammeter may be calibrated to read foot-candles directly and is then useful for checking the adequacy of illumination in work places. By the use of filters, an arrangement of the same kind may be employed in the analysis and matching of colours.

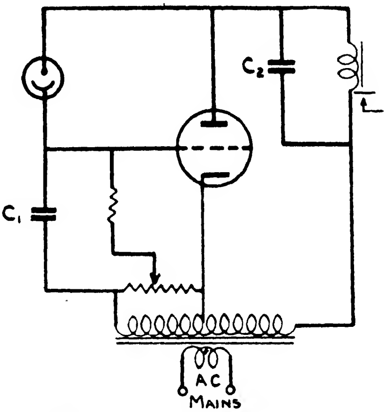


Fig. 10/4. The circuit of fig. 10/3 is here shown with modifications for operation from an a.c. supply.

It is often advisable to measure the density of smoke emerging from a factory chimney to ensure that efficient combustion is taking place and to avoid excessive atmospheric pollution. In this case, a photovoltaic or photo-emissive cell is mounted at a suitable point inside the chimney and illuminated by a lamp on the opposite wall. The photo-cell output (after amplification if the emission type is used) may be employed to give a continuous record of the smoke density, or to actuate an alarm when necessary.

A rather similar application is the measurement of the turbidity of a liquid containing solid matter in suspension. A common arrangement involves a comparison of the light passing through a vessel containing a clear liquid with one containing the turbid suspension; the ratio of the two light intensities measured on a photo-electric cell being a measure of the turbidity. (Such instruments are also known as nephelometers.) If the powder in suspension is allowed to sink under gravity and the light passing through at a certain level is continuously recorded, the distribution of size of the particles of the powder may be calculated.

CHAPTER XI

DISCHARGE LAMPS

Voltage stabilisers. It has already been explained that the potential difference across a glow discharge tube is substantially constant for wide variations of the current flowing through the tube. This property is used in voltage stabilising. It is sometimes necessary to maintain a constant voltage across a valve or between two points in a circuit. If the power is supplied from batteries no difficulty arises, but it is usually more convenient to operate from alternating current mains and the normal type of rectifying arrangement (page 63) delivers an output voltage which varies considerably with the current drawn by the load circuit. The voltage output may be rendered independent of load current by the use of glow discharge lamps, as in the circuit of fig. 11/1.

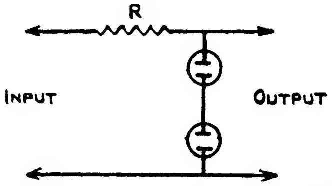


Fig. 11/1. Glow discharge lamps may be used to stabilise the potential difference supplied by a power pack or other d.c. source.

Neon lamps are usually used for this purpose. The operating voltage, which depends on the pressure of the gas and on the separation of the electrodes, is generally in the region of 100 volts, so that it may be necessary to employ several tubes in series. In fig. 11/1 the unregulated supply is connected in series with a resistance R and the stabilising tubes. The load is joined across the discharge tubes. As the load current varies, the current flowing through the neon tubes varies, but the potential difference between the output terminals remains constant, the voltage across the resistance R therefore fluctuating in accordance with the current.

Illumination. Discharge tubes are widely used as light sources, and for this purpose may be divided into the following classes :

1. Glow tubes, in which a glow discharge is set up between metal electrodes in a suitable gas or vapour. The luminosity

originates in the positive column and its colour depends on the nature of the substances in the tube. Glow tubes may be operated from direct current supply, but alternating current operation is more usual, since it is often necessary to use voltages greater than that of the mains. If alternating current is used, each electrode is alternately positive and negative, and the discharge reverses after each half cycle of the supply. The eye does not distinguish this rapid fluctuation and the tube appears to be filled with a single luminous column. The electrodes may be simply pieces of metal, as in the cold cathode tube. Alternatively, the discharge current may be reinforced by using electrodes from which thermionic emission takes place, as in the hot cathode tube. In this case, the cathodes may be heated by an external current source, though in most cases, an adequate rise in temperature is obtained through bombardment of the electrodes by the various particles released in the discharge.

2. Arc tubes, in which an arc is maintained between metal electrodes, usually in an atmosphere of sodium or mercury vapour. Cold cathode tubes usually contain neon, argon or helium or a mixture of the three in suitable proportions, the total gas pressure being of the order of 1 mm. of mercury. The colour of the discharge, which is red (neon), blue (argon) or pink (helium), makes this kind of lamp unsuitable for interior illumination. In a typical neon tube, with a diameter of 0·5 cm. and a current of 5 milliamps the voltage drop in the cathode dark space may be as much as 400 volts. In the positive column the potential gradient is about 15 volts per cm. and to obtain a reasonable efficiency it is therefore necessary to use a long tube so that the voltage drop in the positive column—which emits light—shall be large compared with that in the cathode dark-space which is useless as regards illumination. Apart from this electrical loss, the luminous efficiency of a neon lamp is still below that of a hot cathode or arc tube.

Hot cathode tubes generally contain sodium or mercury vapour.

Sodium lamps. Tubes containing sodium vapour give an intense yellow illumination. The vapour pressure of the sodium is usually about 1 cm. of mercury and a little neon is included to facilitate starting. The electrodes consist of tungsten spirals coated with barium compounds, for thermionic emission, each surrounded by a metal tube which acts as the anode during half of each cycle. To start the discharge, a current is passed through the cathode spirals in series (fig. 11/2). After about 30 seconds—longer in the case of large tubes—this circuit is disconnected and the supply voltage is connected across the two electrodes. A discharge then starts in the neon, and after a short time, the vapour pressure of the sodium, which has already been raised by the heating of the cathodes, reaches a value at which a discharge can be supported. A discharge in neon is easily initiated at a pressure of a few mm., but, with a moderate tube voltage, cannot be established in an atmosphere of sodium until the vapour pressure has risen to a value higher than that corresponding to atmospheric temperature. The necessary rise in the temperature is brought about by the various collision processes which take place in the neon discharge. The tube in which the discharge occurs should be maintained at a temperature of about 280°C., and is therefore insulated from the atmosphere by a double walled glass vessel.

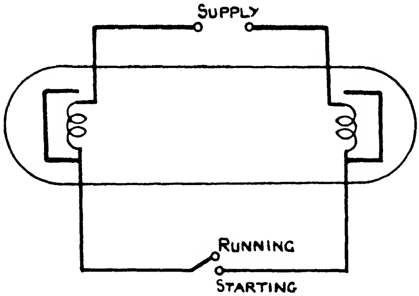


Fig. 11/2. Hot cathode sodium vapour lamp. With the switch in the starting position, current is passed through the cathode coils, heating them and vapourising the sodium in the tube. After a time, the switch is put in the running position, disconnecting the cathode coils and connecting the whole supply p.pd p.d. across the discharge.

The luminous efficiency of a typical sodium lamp is about 50 lumens per watt.

Mercury lamps. These may be considered under two headings.

1. Low pressure discharge tubes. Here the pressure is about 1 atmosphere.

2. High pressure tubes in which the pressure may be several hundred atmospheres. Tubes of this kind have a greater luminous efficiency and give a spectrum approximating to that of sunlight.

Low pressure tubes. The construction of a typical 400 watt tube is shown in fig. 11/3.

A little argon or helium is included in the tube to facilitate starting of the discharge. The electrodes consist of tungsten spirals coated with barium compounds to induce copious emission. The filaments are usually heated sufficiently by the discharge to ensure an adequate supply of electrons. An auxiliary electrode

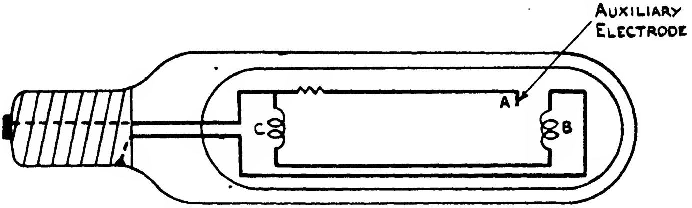


Fig. 11/3. Typical mercury vapour discharge lamp.

A near one filament B and connected through a resistance to the other electrode C facilitates the start of the discharge by maintaining an electric field B and reducing the effective length of the tube to the separation between B and A. The 400 watt tube draws a current of 5 amps. at 20 volts initially and after about ten minutes, when the discharge has become stable, requires about 2·7 amps. at 150 volts. Good heat insulation is essential since, if the temperature is allowed to fall, the vapour pressure of the mercury decreases to a value at which the discharge cannot be maintained. The luminous efficiency of this type of lamp is about 40 lumens per watt.

Since the lamp is usually operated from 240 volt supply mains, a choke is connected in series with it to provide the voltage drop required. The choke also serves to limit the surge of current which may occur when the lamp is switched on. If a lamp of this kind is extinguished, the discharge cannot be restarted until the vapour pressure has fallen below a certain value and the time needed for the vapour to cool for this purpose may be a few minutes.

High pressure lamps. A typical example of this type of tube is the A-H6,* rated at 1 kilowatt. The light source here consists of a tube 2 mm. in diameter and 25 mm. long, enclosed in a quartz bulb surrounded by a water jacket, through which the cooling liquid is forcibly circulated by a pump. The vapour pressure of the mercury in the tube is about 75 atmospheres, though in other models it may be as high as 300 atmospheres. As the vapour

---

* General Electric Company.

pressure is increased the line spectrum characteristic of mercury vapour becomes less distinct and merges into a continuous background radiation. At high pressures, the luminous output approximates to that from a white light source.

Comparison of sodium and mercury discharge lamps. Sodium lamps are operated at low pressures and low current. In mercury lamps on the other hand, both vapour pressure and current density must be high for satisfactory operation. The reason for these differences in operating conditions is revealed by an examination of the fundamental processes involved in the discharge.

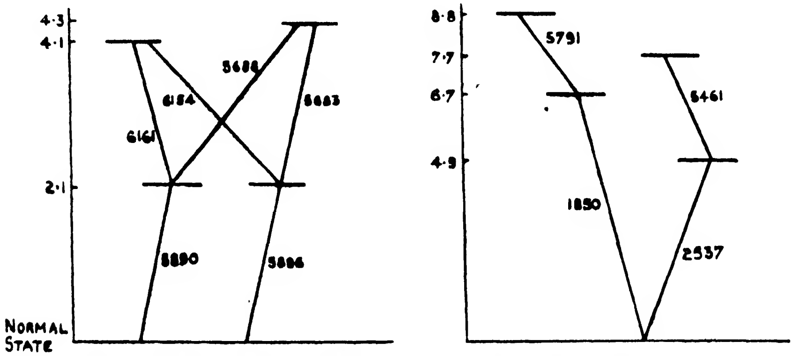


Fig. 11/4. Energy levels in electron volts for sodium (left) and mercury (right) atoms. The differences between various levels are expressed as wavelengths, in Ångstrom Units corresponding to the radiation emitted when an electron moves from a high level to a lower level.

In the case of sodium vapour, the most intense lines are produced by resonance radiation (see page 18). Now these wavelengths are absorbed by the vapour (some energy being re-radiated at the original frequency) to an extent which increases as the pressure is made greater. The necessity of a low vapour pressure is now apparent.

Secondary processes, leading to transitions between higher energy levels and consequently to emission of undesired radiation, occur to an increasing extent as the current becomes greater and may therefore be minimised by using a low current density. The current must however be large enough to maintain the optimum temperature for the discharge, which is about 280°C. In other words, some of the energy of the electrons must be turned into heat rather than into visible radiation.

Conditions in a mercury discharge lamp are somewhat different.

The energy level diagram of fig. 11/4 shows that the resonance radiation is in the ultra-violet region and that the useful wavelengths for illumination (5461 and 5791 Å.U.) are produced by transitions between higher levels. Thus it is necessary to have a high vapour pressure (to promote the absorption of the resonance radiation) and a high current density (to promote the secondary processes which lead to emission of visible radiation).

Arcs. Radiation from a mercury arc contains a good deal of ultraviolet light, which produces biological and bactericidal effects useful in medicine. Ultraviolet radiation is also used to catalyse certain chemical processes. Mercury arcs of large dimensions are used as rectifiers.

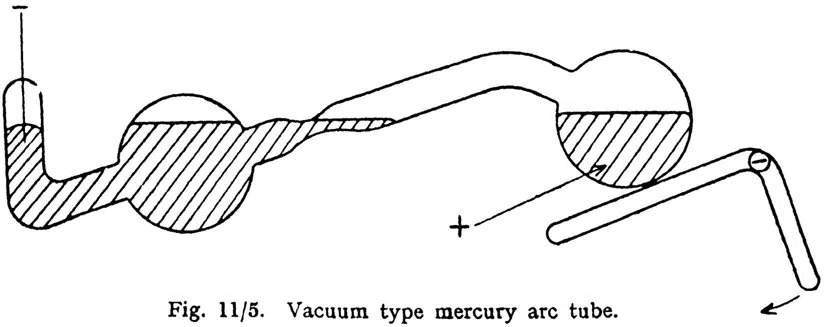


Fig. 11/5. Vacuum type mercury arc tube.

Low power arcs. (a) Vacuum type. This arrangement consists of a quartz tube—glass absorbs ultraviolet radiation—containing mercury in two reservoirs (fig. 11/5). These mercury electrodes are connected to a direct current supply. To start the arc, the tube is tilted so that the mercury forms a continuous thread and is then relevelled. When the thread breaks, a large potential gradient is set up at the point of separation and electrons are drawn from the cathode across to the anode. After a short time the cumulative ionisation produced by collisions between these electrons and atoms of mercury vapour which are always present in the tube leads to the production of an arc discharge. Positive ions formed by the discharge bombard the cathode and one small region—the cathode spot—becomes heated to a high temperature and supplies electrons by thermionic emission for the maintenance of the arc.

(*b*) Atmospheric arcs. Lamps of this kind are made as shown in fig. 11/6. The arc chamber is at atmospheric pressure and is normally filled with mercury. A heating coil placed outside the tube and connected in series with the direct current supply to the electrodes vaporises the mercury at one point and thus establishes an arc, the mercury then receding into the anode chamber until the discharge occupies the whole of the straight portion of the tube. The atmospheric lamp has an advantage over the vacuum type in that the danger of accidental breakage, through movement of the mercury column inside the tube, is reduced by the presence of the air which slows down any such motion. The output of ultraviolet radiation from vacuum type arcs falls off considerably after a time owing to an allotropic change in the quartz of the bulb ; this deterioration occurs to a much smaller extent in atmospheric lamps.

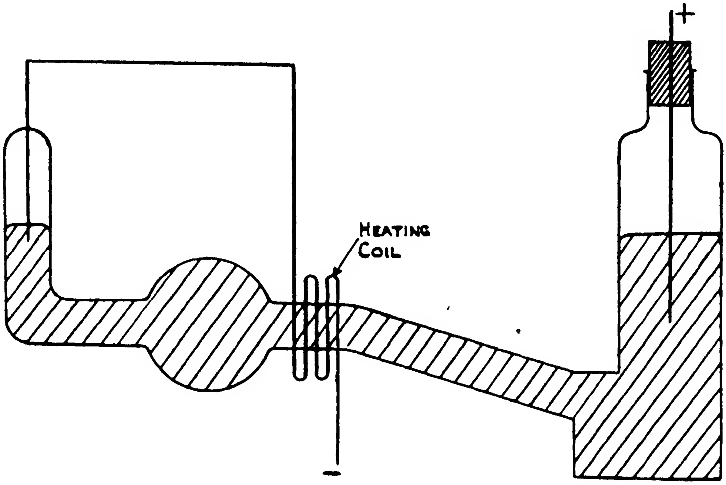


Fig. 11/6. Atmospheric mercury arc.

High power arcs. The use of an arc as a rectifier, with a mercury pool as cathode and a carbon or metal anode, was suggested in 1903 by Cooper-Hewitt, and rectifiers of this kind are now extensively used for supplying direct current to electric railways and tramways as well as for the operation of cinema projectors and in many industrial processes. A simple type of arc rectifier is shown in fig. 11/7. An arc is started between the two mercury pools at the bottom by tilting the tube, and after a short time enough ionisation occurs to permit the transference of the arc to one of the anodes, which are connected to opposite ends of an alternating current supply. When the arc has begun, bombardment of the cathode by positive mercury ions produces a small region (the cathode spot) at high temperature, from which electrons are released by thermionic emission to maintain the discharge. These electrons, together with those formed by collision processes in the discharge, travel to whichever of the anodes is positive at the time and the tube acts as a full-wave rectifier. The electron emission which can be obtained from a mercury pool is much greater than that provided by any solid cathode and the mercury arc rectifier is therefore suitable for applications involving heavy currents.

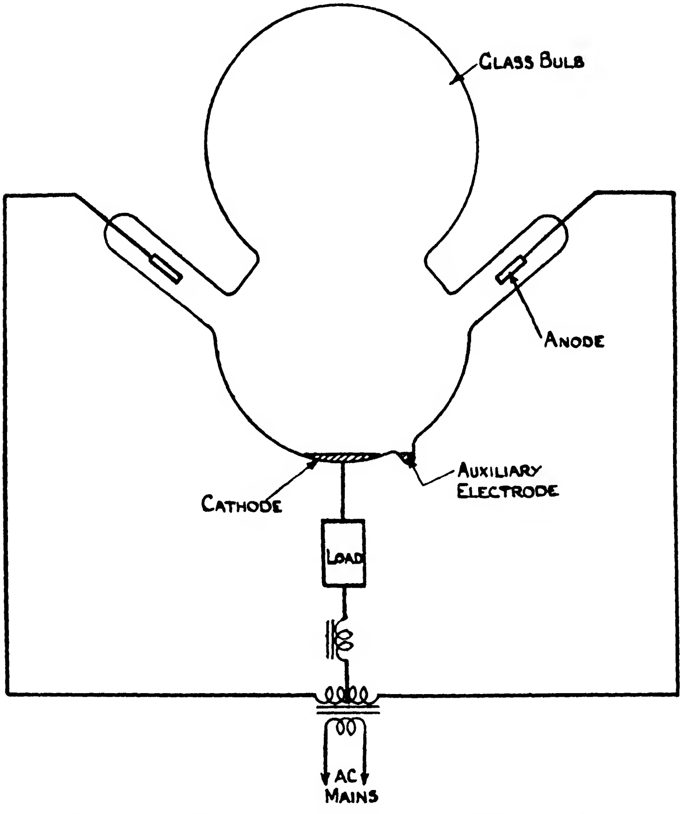


Fig. 11/7. High power mercury arc in glass bulb form.

To maintain the arc when the alternating current supply passes through zero, a choke may be connected in the cathode lead as in fig. 11/7. The back e.m.f. generated by self induction in this coil serves to maintain a potential difference between the cathode and the anode until the supply voltage has risen again. An alternative method is to use an auxiliary anode, maintained permanently at a higher potential than the cathode and situated near to that electrode so that an arc is always present.

Conduction in the reverse direction, known as arc-back, may occur in a mercury rectifier if thermionic emission takes place from the anode. To avoid this possibility the anodes are made of substances such as iron or carbon, which do not emit profusely even at high temperatures and are not wetted by mercury which might otherwise condense on them. The anodes are usually placed in side tubes so that there is no direct path for mercury vapour thrown off from the cathode and they are sometimes protected by metal shields.

Rectifier units for outputs up to 250 kilowatts are generally made in the form of glass bulbs. Before being put into service, the container is evacuated to about 0·0005 mm. of mercury and then sealed off. Since the quantity of air in the bulb is negligible, the mercury vapour is always saturated and its vapour pressure therefore depends on the temperature to which the bulb rises during operation. For satisfactory rectifying action, the vapour pressure of the mercury should be between 0·001 mm. and 0·01 mm. ; the temperatures corresponding to these pressures are about 20°C. and 80° C. Glass bulbs are not used for units with outputs much in excess of 250 kilowatts, because the heat generated by the arc is then so great that the tube cannot be kept below 80°C. unless it be made inconveniently large, when it becomes more liable to damage. For very high outputs, a metal tank is used to house the arc and electrodes. The tank cannot be effectively sealed so as to maintain a high vacuum and is therefore continuously evacuated by pumps. Cooling is provided by forced circulation of water around the outside of the container. Metal tank units have been built for outputs as high as 3000 kilowatts. Rectifiers of this size operate with polyphase alternating supply and have three, six or twelve anodes, the arc travelling to each in turn.

High power-arc rectifiers are often fitted with grids, consisting of metal gratings, placed close to the anodes. If a negative potential of 10—20 volts (with respect to the cathode) is applied to the grids, no current can pass to the anode and the establishment of an arc is thus prevented. If the grids are momentarily made positive, the arc is formed and the usual rectifying action takes place, each anode drawing current until its voltage falls to zero when, the grid being now again negative, current ceases until the grid receives another positive impulse. In this way back-firing and cross-firing (arcing between two adjacent anodes) are completely eliminated. By adjustment of the phase of an alternating voltage applied to the grid, it is also possible to control the intervals during which each anode draws current and thus to control the output of the system.

Fluorescent lighting. Mercury vapour lamps emit a considerable amount of their energy at a wavelength of 2537 Å.U. which is in the ultraviolet region. If the inside of the tube is coated with suitable fluorescent material, this radiation is absorbed and re-emitted without much loss at wavelengths in the visible region. Fluorescent lamps now available contain mercury at a vapour pressure of about 0·0003 mm. which is low enough to ensure that most of the emission is at 2537 Å.U. (see page 154). The construction of a fluorescent lamp is similar to that of the sodium lamp described previously (page 151). Thermionic cathodes are fitted and are heated initially by being connected in series with the supply. When sufficient electronic emission has occurred to start the discharge, the supply voltage is applied between the electrodes, which are maintained at an adequate temperature by ion bombardment. The inner surface of the tube is coated with a layer of fluorescent material. Among the substances commonly used are calcium tungstate, zinc silicate and cadmium borate; to obtain white light it is necessary to use a mixture of several substances and the fluorescent activity is stimulated by the presence of magnesium as an activator.

The luminous efficiency of a fluorescent lamp in the visible region may be 30-40 lumens per watt. Lamps operated from alternating current supply show a flicker which sometimes leads to dangerous stroboscopic effects in the illumination of machinery. The flicker can be overcome by using sources containing two lamps, with currents having a phase difference, produced by resistance and capacitance in the supply circuit, so that a more nearly uniform illumination is obtained.

CHAPTER XII

ELECTRONIC HEATING

ELECTRONIC devices may be used to generate heat in two main ways.

(1) Induction heating. In this method the object to be heated is placed near a conductor carrying an alternating current, usually at radio frequency. Induced e.m.f's set up in various parts of the load give rise to eddy currents by which the heating is accomplished.

(2) Dielectric heating. Here the object to be heated is placed between the plates of a condenser connected to a source of radio-frequency current. The load must be an insulating substance. The alternating electric stress imposed on the plates of the condenser gives rise to an alternating displacement of electrons in the medium and heating is thereby produced.

Induction heating. Eddy currents at frequencies up to 100 kilocycles/sec. are used in several metallurgical processes and in this case a rotary generator, working on the principle of the dynamo, is a suitable source of energy. For heating at higher radio-frequencies, which is becoming more important, mechanical devices are inefficient and valve oscillators are used. Since induction heaters are generally operated without skilled attention, the design and construction must be simple and the circuit should be such as to simplify repairs or adjustments. For these reasons a single stage self-excited oscillator is commonly employed. The generator must also be capable of working with no loss of efficiency into a variable load. During the heat treatment of metals the resistance may be doubled or trebled as the temperature rises and in the case of steel a further complication is introduced by the fact that at the Curie point, a temperature of about 800°C., the magnetic properties disappear altogether, thus eliminating the inductive component of the load impedance. These variations make it impossible to keep the load circuit tuned to a fixed frequency while heating is in progress and the use of tuned circuits by way of coupling between the oscillator and the load is ruled out. The heater coil must, in fact, form part of the oscillator tank circuit. Then, as the load impedance alters, the frequency of operation adjusts itself so that the oscillator gives a reasonably constant output. The tank circuit, part of which is led away to surround the load, is generally in the form of a copper pipe, water-cooled in large installations where currents of 100 amperes or more are not uncommon. The tuned anode and the Colpitts circuit (page 106) are most often used for induction heating ; the output may be adjusted by means of a variable grid-bias resistor.

Certain limitations to the frequency of operation are apparent. It is not economical to build a radio frequency oscillator for operation below 100 kilocycles/sec. as the coils and condensers then become inconveniently large. On the other hand, operation at very high frequencies is usually characterised by reduced efficiency, though the permissible range is being extended by the development of new valves. At frequencies of several megacycles/sec. it is necessary to arrange the wiring of the oscillator and the position of the load carefully so that energy is not lost through radiation from leads and leakage in insulators.

The heating effect produced by eddy currents is proportional to the square root of the frequency and therefore rises only slowly as the frequency is increased. The optimum frequency is therefore determined mainly by the nature of the process to be undertaken and by the other considerations mentioned above. Frequencies between 200 and 2000 kilocycles/sec. are commonly used. The depth to which eddy currents penetrate is an important consideration. A high frequency current in a conductor tends to travel by way of the outer layers and is most concentrated near the surface. This skin effect, which is more pronounced at high frequencies, is a nuisance in many branches of radio work, but is a convenience here since it enables the specimen to be heated to whatever depth is appropriate merely by adjustment of the operating frequency. About 90% of the heat developed in the material is confined to a layer of thickness t, given by

$$t = 5300 \sqrt{\frac{\rho}{\mu f}} \quad \text{cms.}$$

where ρ is the specific resistance and μ the permeability of the specimen and f is the frequency. From this formula it is apparent that as the frequency is increased the depth of penetration becomes less and at frequencies of more than a few megacycles/sec. the heating is confined to a layer which is too thin to be useful for most purposes.

Induction heating is most successful with iron and steel for the following reasons. (1) The load should have some inductance, so that energy may be transferred to it from the heater coil, and magnetic materials are therefore more easily heated. (2) The impedance of the load circuit is always low and matching to the oscillator is difficult. It is however more nearly accomplished with metals which have a fairly high specific resistance such as iron and steel than with good conductors such as copper. In the case of iron and steel, the magnetic properties disappear at the Curie temperature, which may be between 700° and 800°C. At temperatures above this the load impedance drops considerably, since the permeability falls to 1, and the depth of heating increases.

Applications of eddy current heating.

1. *Case hardening.* It is a common practice to treat tools and engine shafts by case hardening, which involves the production of a thin hard layer on the surface. The process may be accomplished by heating the object concerned and then cooling the surface rapidly by immersion in oil. If eddy current heating is applied at a suitable frequency, the heating effect is confined to a surface layer and since the inside of the metal retains its original temperature, the surface is cooled by conduction as soon as the inducing field is removed.

2. *Tinning plates.* Tin plate is normally made by dipping sheets of steel into molten tin. It is more economical to use a thinner layer of tin, deposited electrolytically, but if this is done the surface is dull and uneven, so that heating is necessary to produce the customary polished finish. The tinned strip is therefore passed through a flat coil carrying a heavy current at a frequency of a few hundred kilocycles/sec. from an oscillator. Speeds of as much as 1,000 ft./min. can be obtained.

3. *Metal fusion.* To avoid the possibility of contamination by furnace gases, metals may be melted by eddy current heating.

To secure uniform heating throughout the material, a low frequency—often in the audio-frequency range—is necessary, and in this case mechanical generators are more suitable. Radio-frequency heating is suitable where the size of the object to be melted is not too great.

4. *Evacuation of valves.* However vigorously the air in a valve is pumped out, some gas remains in the walls and on the electrodes and can only be removed by heating. The envelope of the valve is raised in an oven to the highest temperature which can be reached without softening—about 450° for glass and 1,000°C for silica. In this way the gas is driven out of the walls and it is then necessary to apply the same treatment to the electrodes. The valve, still connected to the vacuum pump, is placed inside a coil carrying radio-frequency current. The eddy currents produce little effect on the envelope, but the electrodes may be raised to 800° or 900°C. After this the valve is sealed off and the final stage in the evacuation process—known as "gettering"—may be undertaken. At an early stage in the assembly of the valve, a little magnesium or barium, covered by gauze, is attached to the electrode assembly, usually near the base of the valve so that it is not affected by the heating of the electrodes. Eddy current heating is then applied to vaporise the "getter," which settles on the walls in a finely-divided condition forming the mirror-like surface commonly seen inside a valve. Any trace of gas still remaining is absorbed by this surface.

5. *Soldering.* Condensers, transformers and other radio components are often sealed in metal cans to avoid the effects of exposure to tropical conditions. The soldering operation necessary to seal the container may be accomplished rapidly by passing the tins on a moving belt past a coil carrying radio frequency current. The solder and flux are placed in position when the tin is assembled and may be raised to the required temperature in less than a second. Apart from this very great speeding-up, induction heating has the additional advantage that it can be applied for so short a time that components inside the container are not appreciably heated by the soldering operation.

Some further applications of eddy current heating are discussed in Chapter xiii.

Dielectric heating. This method is suitable for raising the temperature of non-conducting materials. It is difficult to heat such substances effectively by the normal methods, which involve conduction or radiation of heat from the source and further conduction inside the material. In the technique to be described, the heat is generated within the material and is thus uniformly distributed without any steep temperature gradients.

A body placed between the plates of a condenser connected to a radio frequency generator may become heated in the following four ways:—

1. *Electronic displacement.* In an insulating substance there are no conduction electrons and a movement of charge within the body from one electrode to the other is therefore not possible. When the electric field is applied, however, the atoms become distorted by a displacement of electrons towards the positive plate of the condenser. In an alternating field the rapid oscillation of electrons which results from these displacements absorbs energy from the field and gives it up as heat within the material. Thus whenever the condenser is charged or discharged some of the energy involved is dissipated as heat instead of being taken from, or returned to, the generator.

2. *Atomic displacement.* Whole atoms may be displaced from their usual positions in the same way and another source of heating is therefore apparent. Electronic and atomic displacement do not produce noticeable heating at radio frequencies.

3. *Dipole rotation.* Many materials are known to have unsymmetrical molecules, with a concentration of electrons at one end and a deficiency at the other end. These are known as polar molecules or dipoles and react to an electric field by setting themselves along the lines of force. Thus in an alternating field they rotate first in one direction and then in the other, releasing energy in the form of heat. Many organic substances show molecular polarisation.

4. *Ionic displacement.* If the material is an electrolyte, or is dissolved in one, movement of ions occurs when the electric field is applied. This displacement constitutes a current within the material, though the rapid reversal of its direction prevents any charge from escaping into the external circuit.

In each of these four processes, some of the energy supplied to the condenser in charging it is not stored in the dielectric until the next discharge, but is dissipated as heat. The ratio of the heat lost in charging the dielectric to the total energy stored is called the power factor of the material. This quantity is independent of the size and shape of the object but depends to some extent on the frequency.

The four possible types of displacement occur most readily at different frequencies. If an electric field is applied the movement of the electron, atom, molecule or ion is opposed by electrical forces analogous to friction and viscosity, and the displacement does not begin for a short time—known as the relaxation time. A similar delay in response occurs when the field is removed. If the period of the field is small compared with the relaxation time, no appreciable movement of charge takes place and little heat is generated. On the other hand, if the field has a period much greater than the relaxation time, the movements of charge can follow the variations in field without difficulty, and again no losses are occasioned. The most favourable conditions for heat generation arise when the period of the field is of the same order as the relaxation time; the hysteresis effect between the applied force and the resulting displacement is then greatest. There is thus a resonance frequency for each of the types of displacement; as might be expected this frequency is highest for electronic resonance. Substances which show electronic or atomic resonance within the optical range usually absorb energy heavily at these frequencies and are therefore opaque to light. In transparent materials the resonances are outside the visible range.

Materials suitable for dielectric heating. The heat generated in a material exposed to an electric field is proportional to

$$fX^2K\phi$$

where f = frequency
X = potential gradient across specimen
K = dielectric constant
ϕ = power factor

The product of power factor and dielectric constant is known as the loss factor and indicates in a general way the suitability of a material for dielectric heating. Approximate values for some common substances are set out on page 166.

APPROXIMATE LOSS FACTORS AT 1 MEGACYCLE/SEC.

Mica	0.001
Rubber	0.002
Polystyrene	0.001
Methacrylic resins (e.g., perspex)	0.06
Urea-formaldehyde resins	0.2
Spruce wood, dry	0.1
Phenolic glue	5
Urea glue	20

Frequency of operation. Since the heat generated is proportional to the frequency, it is advisable to use a wavelength as short as possible for dielectric heating. There are, however, other considerations. The loss factor of most materials changes as the temperature rises and the load circuit becomes detuned —an effect which is most pronounced at high frequencies. Difficulties also arise when the object to be heated is comparable in size with the wavelength of the oscillator. In this case standing waves of voltage appear on the electrodes—as on a transmitting aerial—and modifications are necessary to ensure uniform heating. Frequencies between 5 and 50 megacycles/sec. cover most of the applications of dielectric heating. It is possible to work up to 300 megacycles/sec., although in this region it is not possible by methods commercially available to generate more than a few hundred watts. High power output is not always essential, for the materials subjected to dielectric heating are invariably insulators and have a high impedance, so that matching of the load to the oscillator is accomplished more effectively than in eddy current heating. To secure efficient heating it is necessary to establish a high potential across the load, which is usually connected across the tank circuit of the oscillator, either directly or by a transmission line. A variable inductance is sometimes placed in the load circuit so as to facilitate the adjustment of its impedance to the optimum value.

Applications of dielectric heating.

1. *Dehydration of foodstuffs.* Most vegetables contain a high proportion of water and considerable economies in transport and storage may be obtained by dehydration. Experiments which have been performed with cabbages, onions, carrots and beetroot indicate that the moisture content can be reduced to

less than one per cent. without any sign of cooking, by the application of dielectric heating in a vacuum. The temperature is raised to about 60°C. and the oscillator is switched off. To keep the temperature constant after this it is only necessary to apply the heating for about ten per cent. of the total time. The pressure in the heating chamber is reduced to about one inch of mercury to facilitate evaporation of the water and the drying process may last about two hours. This is only a tenth of the time which would be necessary to carry out dehydration to the same extent by heating in an oven.

2. *Cooking.* Cooking by dielectric heating is an obvious application and experiments have been carried out on a number of foodstuffs. The results are satisfactory in some cases and poor in others, but the cost of the equipment is at present too great to permit of domestic use. Further development might be worth while if valves delivering a kilowatt or so of power at an anode voltage of 200—300 were available ; at present the high-tension network accounts for most of the cost of the oscillator.

3. *Wood glueing.* The manufacture of plywood, in which several thin layers of wood are glued together, is a process suitable for the application of dielectric heating. The glues generally used are urea-formaldehyde or phenolic resins, which, to give a satisfactory joint, must be cured at about 300°C while under pressure. It is wasteful and very slow to heat the glue by means of a hot conductor applied to the wood above it, though this has been the normal method for some time. It will be observed from the table above that the loss factor of the glue is many times greater than that of the wood which surrounds it. If, therefore, the pile of sheets is assembled and subjected to a radio-frequency field, the heat concentrates in the glue layer, which may be heated to the curing temperature while the wood remains quite cool. In this way the heat is delivered directly to the point at which it is required. A pressure of about 150 lb. per square inch is necessary during the curing process and the metal plates between which the wood is pressed are made the electrodes of a condenser connected to the oscillator tank circuit. Frequencies between 10 and 50 megacycles/sec. are commonly used in this work and the curing time is no more than a few minutes. The finished bond between adjacent sheets is stronger than the wood

itself. In the manufacture of aircraft it is often necessary to produce curved sheets of plywood and the same technique may be employed. In this case curved electrodes are used but the same principles apply.

4. *Pre-heating of plastics.* The plastic materials now so widely used are in most cases not plastic at ordinary temperatures and must be raised to 150—200°C before undergoing moulding or other operations. When the appropriate temperature has been reached the material may be shaped by pressure, and on cooling it retains whatever shape has been imparted in this way. It is possible to combine the two processes by heating the plates of the press, but it is difficult to produce mouldings more than a fraction of an inch thick, since the materials used are not good conductors of heat and the inside of the specimen may be quite cool while the outside is at the curing temperature. To obviate this difficulty it is becoming common to cure the material, usually in the form of a powder, by dielectric heating before moulding is begun. In this way the whole of the sample may be brought to the proper temperature at the same time and the uncertainties of the moulding process thereby eliminated. It would be convenient if the heating took place in the press itself, but this is precluded by insulation difficulties—the press is usually made entirely of metal.

CHAPTER XIII

ELECTROTHERAPY

THE electrical properties of the human body and the effects produced by the passage of current through it may be summarised as follows.

1. **Direct current.** The skin offers a high resistance, which is reduced by the presence of moisture, and the resistivity of the tissues and organs is in most cases quite low. The fluids which permeate the tissues contain mineral salts in solution and current is carried by ions of these substances. Migration of ions under the influence of a current is sometimes of therapeutic value and it is possible to secure the entry of ions into the body by placing a pad soaked in a suitable solution between the skin and one of the electrodes used to pass the current. Direct currents have the property of stimulating muscular tissue by involuntary contraction; if an excessive current is passed the contractions may become sufficiently violent to interrupt the action of the heart with fatal results.

Effects generally similar to these are obtained with interrupted direct current obtained from a mechanical make-and-break device. It is not necessary to discuss this subject further, since we are interested principally in the applications of alternating currents.

2. **Low frequency alternating current.** Alternating current of low frequency, such as that obtained from the mains, produces much the same effect as direct current. There is no migration of ions in the tissue fluid, since the direction of the current is reversed so rapidly.

3. **High frequency alternating current.** At higher frequencies, the impedance offered by the body diminishes and capacitance rather than resistance predominates. At frequencies of several megacycles per second the body behaves almost entirely as a capacitance. When the frequency is increased above that of the mains supply, the current necessary to produce stimulation of

muscle increases also. Thus a few milliamperes of direct current or low frequency alternating current causes appreciable contraction, but to produce the same effect at 500 kilocycles/sec. requires a current of several amperes. At radio frequencies it is possible to pass large currents through the body without causing any contraction of the muscles and a considerable heating effect may be produced in this way. The use of electricity to produce internal heating is known as *diathermy.* The three possible techniques, all of which are in common use, are as follows.

1. **Conduction heating.** A direct current or alternating current with frequency below one million cycles per sec. is passed through the body by means of two electrodes attached at suitable positions. This method is sometimes referred to as *classical diathermy.*

2. **Dielectric heating ;** (*short-wave diathermy*). If a material is made the dielectric of a condenser connected to a source of alternating electromotive force, some of its electrons are displaced, first in one direction and then in the other, on account of the alternations in electric stress. These displacements of charge constitute a current and produce the usual heating effects. To heat the tissues by this method, electrodes connected to a radio-frequency oscillator are placed near the skin, separated from it by air or glass, and internal heating occurs although there is no direct contact with the electrodes. Frequencies between 10 and 100 megacycles/sec. are commonly used.

3. **Eddy current heating.** If a conductor is placed in an alternating magnetic field, currents are induced in it. The process known as *inductotherapy* involves placing a coil carrying a large radio-frequency current near the part of the body which it is desired to heat.

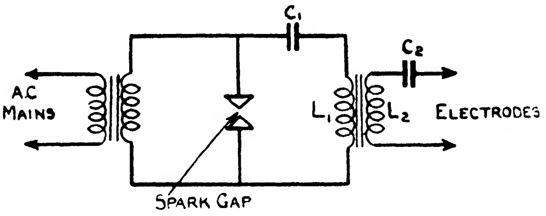


Fig. 13/1. Spark-gap generator used in diathermy.

Generators. High frequency oscillations for conduction heating are generated by methods which in other branches of electronics would be considered antiquated. The commonest arrangement is the spark-gap generator, (fig. 13/1). A condenser C_1 and an inductance L_1 are connected across the secondary winding of a step-up transformer, the primary being connected to alternating current mains. A spark gap is joined across the secondary winding. As the e.m.f. rises from zero C_1 charges, and when the potential difference across it becomes enough to break down the insulation of the air in the spark gap, the condenser discharges through the coil. If the resistance in the circuit is small enough to make

$$\frac{R^2}{4L_1^2} < \frac{1}{L_1C_1}$$

the discharge is oscillatory and a damped train of oscillations with frequency

$$f = \frac{1}{2\pi\sqrt{L_1C_1}}$$

is produced (cf. p. 50). The condenser charges again and the process continues. The gap is usually adjusted so as to break down near the peak of the secondary potential difference, so that one train of waves is generated in each half-cycle. If the breakdown voltage is reduced, by placing the contacts of the gap closer together, more than one wave train is obtained in each half cycle. In this case, the amplitude of the oscillation is smaller but the current flowing is more nearly constant. The current set up in the coil is transferred by magnetic coupling to the patient's circuit, consisting of another coil L_2, tuned by a condenser C to the resonant frequency of the transformer secondary circuit, and connected to two electrodes which are placed on the patient's skin. At the frequency employed (usually about 500 kilocycles/sec.), the impedance of the body is largely resistive and the capacitance between the electrodes need not be considered in calculating the resonant frequency of the patient's circuit. The coils in the patient's circuit and the generator circuit are usually wound on "pancake" formers; the coupling between them, and therefore the output amplitude, may be altered by moving one coil relative to the other. Valve oscillators of simple design are sometimes used in conduction heating. The anode supply is not rectified, and oscillations occurs in alternate half-cycles.

For dielectric heating valve generators are used, giving power outputs of a few hundred watts in most cases. Here the capacitance between the electrodes is comparable with that in the oscillator tank circuit and a variable condenser is connected in the patient's circuit so that it may be adjusted to resonance. A thermocouple in the patient's circuit is connected to a galvanometer which indicates the current flowing in the circuit and shows when resonance has been obtained. The electrodes, which are attached by movable arms to the coil coupling the patient's circuit to the oscillator, are usually circular metal plates, faced with glass or some other insulating material.

Heating effects. In conduction heating, the greatest rise of temperature takes place at the skin, which has a higher resistivity than the underlying tissues and the extent to which internal heating may be accomplished is limited by the danger of burning at the surface. In dielectric heating, the resistance of the skin plays no part and more intense heating is thus possible. The distribution of energy in conduction heating depends on the resistivity of the material through which the current passes ; in dielectric heating the deciding factor is the dielectric constant of the material. Since the various tissues and organs in the body have a high fluid content, they all have approximately the same dielectric constant as water and uniform heating is more easily obtained. This technique is particularly suitable for application to the brain and spinal cord, which are both surrounded by poorly-conducting material and are therefore inaccessible to conduction currents.

Inductotherapy. In this method a well-insulated cable connected to a short-wave oscillator, is placed near, or wound round, the part of the body under treatment. Heat is produced more rapidly than in either of the two methods discussed above, for the same expenditure of electrical energy. The eddy currents—and therefore the heating effect—are greatest in regions of least resistivity and again there is no danger of burning the skin.

Electrosurgery. If current from a diathermy generator is passed through a portion of the body, using one small and one large electrode, the current density and therefore the heating effect is much greater under the smaller electrode than elsewhere. In this region the rise in temperature may be sufficiently great to coagulate the tissue by evaporation of the fluids in it.

CHAPTER XIV

CATHODE RAY OSCILLOGRAPHY

An oscillograph is a device for delineating the waveform of an alternating quantity. Electromechanical oscillographs are used for some purposes and are satisfactory at low frequencies. This type of oscillograph consists, in principle, of a mirror galvanometer, in which alternations of the current or voltage under examination are reproduced by the movement of a coil of wire in a magnetic field and magnified by an optical lever system. The moving element which traces out the waveform has inertia and its response is not sufficiently rapid for use at frequencies above about 3000 c/s. At high audio-frequencies or at radio-frequencies, mechanical devices are unsuitable and it is necessary to find a recording arrangement with extremely small inertia. A beam of electrons, having negligible mass under normal conditions, fulfils this requirement and forms the moving element of the cathode ray oscillograph. In this instrument a narrow pencil of electrons is deflected by electric or magnetic fields associated with the voltage or current to be examined and strikes a screen coated with fluorescent material. The screen is perpendicular to the path of the electron beam which therefore produces a luminous spot moving in accordance with the alternations under investigation. By the application of a suitable time scale the trace may be spread out into a graph of the alternating quantity. The essential requirements for a cathode ray oscillograph are therefore :

1. A source of electrons.
2. Arrangements to focus them into a narrow pencil.
3. Arrangements to deflect the electron beam.
4. A time scale.
5. A fluorescent screen.

Details of the construction are as follows.

1. **Source of electrons.** A cathode, directly or indirectly heated is used. To obtain a narrow beam of electrons, the area

of the emitting surface must be small and the hairpin filament and tubular cathode common in thermionic valves are therefore not suitable. The cathode usually takes the form of a tungsten loop with a tiny bead of active material at the tip.

2. **Focussing.** The term focussing is used here to describe the process of concentrating the electrons into a narrow beam, and may be accomplished to a certain extent by the electrodes of the tube.

(*a*) A cylindrical metal shield, or *grid*, surrounds the cathode and is maintained at a negative potential with respect to that electrode. Any electrons which approach it are repelled and the emission is thus concentrated into a narrow beam along the axis of the tube. The grid bias voltage is usually between −20 volts and −100 volts in small tubes.

(*b*) The electrons are accelerated by a number of anodes which may take the form of circular discs, each with a small opening at the centre, through which the beam is constrained to pass. The potential applied to the anodes depends on the size of the tubes and on the intensity with which it is desired to illuminate the screen. In small tubes a few hundred volts may suffice and in larger tubes the anode potential may be several thousand volts. Information on this point is set out in table 14/1.

TABLE 14/1

OPERATING CHARACTERISTICS OF SOME CATHODE RAY TUBES.

Maker's number	Screen diameter inches	Heater	Grid bias volts	Anode voltages			Maximum sensitivity mm/volt
				1st	2nd	3rd	
E46-12	12	4 volts 1 amp.	−60	250	1100–1400	4000–5000	0·17
E42-G6	6	,,	−35	200–400	1000–2000	—	0·27
A41-G4	4	,,	−40	400–500	1000–1200	—	0·39
E40-G3	3	,,	−25	140–200	500–800	—	0·19
914	1	2·5 volts 2·1 amps.	−65	100	500	—	0·7

The focussing in these two ways is not sufficient to give a really sharp trace on the screen and additional methods which may be employed are as follows.

(*c*) An electron projected into a magnetic field executes a circular path if the lines of force are perpendicular to its initial direction of motion, and a helical path if the lines are parallel to the direction in which it is moving (page 12). Current passed through a coil wound over the neck of a cathode ray tube produces a magnetic field parallel to the axis of the tube and it is possible to arrange that the helical paths, traced out by the electrons with different initial speeds and directions, all reach the screen at the same point.

(*d*) In the system known as gas-focussing a small amount of an inert gas such as argon or helium is introduced into the bulb after the air has been removed. The electron stream ionises this gas by collision (page 17), producing electrons which are added to the main current and positive ions which, being much heavier, remain in the path of the beam. Thus a core of positive ions is formed along the beam and any electrons which deviate from the normal path are attracted to it. If gas focussing is employed, it is usually necessary to have only one anode, and satisfactory operation may be obtained at audio-frequencies. At high deflecting frequencies, however, the movement of the beam is sluggish on account of the inertia of the positive ions and blurring of the trace results.

(*e*) Electrostatic focussing is the method commonly used. A charged particle in an electric field tends to move at right angles to the lines of force. In a uniform field the path of an electron is therefore a straight line but in a non-uniform field a curved path is executed and it is possible to construct an electrostatic lens—a system of conductors which deviates a beam of electrons in the same way as an optical lens refracts a beam of light. (Electron optics is discussed more fully in chapter XVI). By the use of suitably-shaped electric fields, focussing analogous to that of an optical instrument may be obtained. To produce the required fields the anodes are made in the form of cylindrical tubes. A typical assembly is shown in fig. 14/1. The focal lengths of electron lenses depend on the potentials of the anodes and sharpness of focus may thus be altered by changing the potential of one anode.

3. **Deflection of the beam.** The electron pointer may be deflected by passage through an electric or a magnetic field.

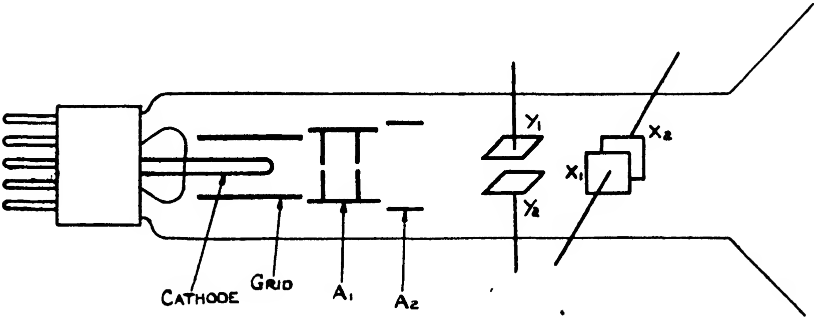


Fig. 14/1. Construction of typical cathode ray oscillograph tube.

In the former case an alternating potential difference is established between two metal plates, one above and one below the path of the beam. For magnetic deflection a current having the wave-form to be examined is passed through coils wound round the tube thus producing a magnetic field in which the electrons are deflected.

To record the variation with time of an electrical quantity it is necessary to deflect the spot in two directions. It is convenient to have these at right angles and if electrostatic deflection is used the deflector plates are arranged in two pairs (fig. 14/1). A potential difference set up between the first pair Y_1, Y_2, which are horizontal, causes a vertical displacement of the spot on the screen. The second pair X_1, X_2 are vertical and are used to give a horizontal deflection which usually forms the time scale.

4. **Time bases.** Suppose that it is required to delineate the waveform of an alternating voltage with amplitude 100 volts and frequency 1000 cycles/sec. A potential difference of this size connected directly across the Y plates causes the spot to execute a simple harmonic motion and the trace of this is a straight line. To examine the waveform it is necessary to spread the motion across the screen, a process which may be accomplished by applying a suitable potential difference to the X plates. If the potential difference between X_1 and X_2 is increased uniformly starting from zero, the spot moves across the screen at constant

speed and the displacement caused by the Y plate voltage is spread out in the horizontal direction. If the X voltage can be made to sweep the spot across the screen in 1/1000 second, a single cycle of the Y potential appears on the screen. This trace disappears in a short time, but by returning the spot to its original position and repeating the sweep, the same trace may be produced again and again, giving the impression of a stationary pattern. If the spot moves steadily across the screen in 1/500 second and quickly back, two cycles of the Y potential are seen. In general, a stationary pattern containing a whole number of cycles is exhibited when the time taken by the spot to sweep across the screen and back is a multiple of the period of the alternation which is being examined. The potential difference applied between X_1 and X_2 to produce these results is called the *sweep* or *time base* voltage and its waveform should be of the form shown in fig. 14/2. Along AB the spot traverses the screen in one direction and along BC it returns quickly in the opposite direction. Circuits suitable for the generation of such a saw-toothed voltage wave will now be described.

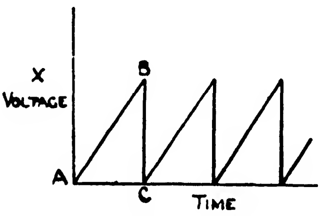


Fig. 14/2. Saw-toothed waveform of deflecting voltage for a linear time base. The spot is swept slowly across the screen as the deflecting potential rises along AB and moves back quickly to its starting point along BC.

1. *Neon lamp time base.* A tube containing neon between two electrodes at low pressure becomes conducting when the potential difference across it reaches a certain value—about 170 volts for the types commonly used. The discharge, which allows current to pass, persists if the potential difference is reduced below the starting value and ceases at some lower value—perhaps about 130 volts. The time base now to be described depends on this difference between the starting and stopping potentials of the discharge. A condenser C is charged from a direct current source through a resistance. The neon tube is connected in parallel with C (fig. 14/3 (*a*)). The potential difference across the condenser, and therefore across the tube, rises exponentially (fig. 14/3 (*b*)). When it reaches the value required to make the tube conducting, C discharges until the potential difference across it falls below the stopping potential

for the discharge, at which point the neon tube ceases to conduct, the condenser recharges and the cycle is repeated. The potential difference across the lamp is of the waveform shown in fig. 14/3 (*b*), and is a fair approximation to the required saw-toothed pattern, but is not sufficiently close for many purposes. The output voltage—about 40 volts—is not enough to sweep the spot across the screen of a normal cathode ray tube and the curvature of the portion *AB* indicates that the time scale is not linear. The waveform under examination is therefore distorted. Although the neon time base is of little practical value, it is informative to examine its properties further. The frequency of the sweep voltage may be altered by adjusting the time constant *CR* which determines the charging rate of the condenser (page 49). Thus *R* may be a rheostat and *C* may be selected by a switch from among several condensers of different values. As has been mentioned, the amplitude of the sweep voltage is fixed at the difference between the starting and stopping voltages for the discharge. It is usually necessary to have a greater voltage than this and to have some means of controlling its amplitude.

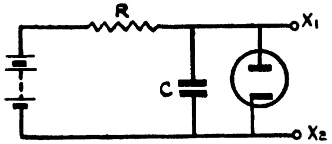


Fig. 14/3 (*a*).

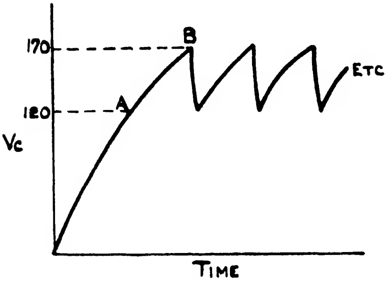


Fig. 14/3 (*b*).
Neon lamp time base.

2. *Thyratron time base.* Some of the shortcomings of the neon lamp time base may be eliminated by the use of a three electrode tube in which the starting potential is controlled by the bias voltage applied to an auxiliary electrode—the grid. The gas discharge triode or thyratron is similar in construction to the ordinary thermionic valve and has an indirectly heated cathode. It contains a small amount of mercury vapour or of an inert gas such as argon or helium. Its operation in a time base depends on the same principles as the neon tube arrangement. The anode voltage required to initiate the discharge is, however, controlled by the potential applied to the grid, and the relation between these two quantities is shown in fig. 14/4. In some

cases the operating part of the characteristic is in the region of positive grid voltage. The grid control ratio is

$$\frac{\text{anode voltage required to start discharge}}{\text{grid bias potential}}$$

This ratio is analogous to the amplification factor of a normal triode and, except for low grid voltages, is constant ; in the valve to which fig. 14/4 refers, it is about 20.

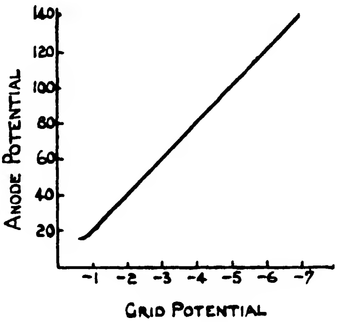


Fig. 14/4. Characteristic curve for a typical gas-filled triode or thyratron.

Once the discharge has begun it is not influenced by grid potential* and continues until the anode voltage is reduced to a value less than the ionising potential for the gas or vapour in the tube—about 15 volts for mercury. Fig. 14/5 shows the basic circuit of a thyratron time base. The condenser C_1 is charged through R_1 from a high tension supply. When the potential difference across C_1 reaches a value equal to the grid bias voltage times the grid control ratio, the valve becomes conducting and the condenser discharges until the anode voltage has fallen to the extinction value. The condenser then begins to charge again and the process is repeated. The waveform of the potential difference across the valve is of the form shown in fig. 14/5. The frequency of operation may be adjusted by varying R_1 or C_1 and the amplitude of the output may be controlled by adjusting the grid potential. The resistance R_2 is necessary to prevent damage to the cathode of the valve by the passage of excessive current during the discharge.

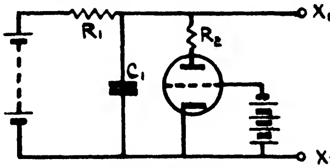


Fig. 14/5 (a).

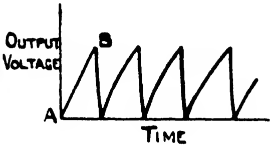


Fig. 14/5 (b).
Simple thyratron time base.

The thyratron time base, though more satisfactory than the neon

---

* A thin layer of positive ions is formed around the grid and acts as an electrostatic screen in that it isolates the grid from the discharge.

lamp arrangement, has a number of disadvantages in the form illustrated by fig. 14/5 (*a*) :—

1. If mercury vapour is used, the operating characteristics of the tube depend on the vapour pressure of the mercury and therefore on the temperature. This difficulty does not arise in the case of gas-filled valves.

2. A thyratron does not generate the required triangular oscillations (fig. 14/2) at high frequencies—the limit is about 50 kilocycles/sec.—and is thus unsuitable for oscillography at radio frequencies. If C_1 and R_1 in fig. 14/5 (*a*) have values calculated for operation at radio frequency, the discharge begins in the usual way when the condenser is first charged to a sufficiently high potential. The valve then remains in the conducting state and the potential difference across it becomes constant—at about 15 volts for mercury. This behaviour is commonly explained by the statement that the mercury ions formed during the discharge require a certain time to recombine. At high sweep frequencies the condenser voltage is below the stopping potential of the discharge for only a short time, during which the discharge may be maintained by residual ions which have not recombined. Recent research shows that the deionisation time is not the limiting factor. For operation at high frequencies C_1 must have a low value and it is found that in this case the anode voltage falls to zero while the condenser discharges, but quickly rises again to a value at which conduction through the valve can be maintained. This effect can be prevented only by reducing the charging current of the condenser. To do this it is necessary to increase R_1 and then the frequency of the time base is lowered. The thyratron circuit is thus not suitable for use at radio frequencies and sweep circuits employing high vacuum or hard valves are employed. A detailed discussion of these devices is outside the scope of the present volume.

3. The output waveform is not linear. The charging stroke (*AB* in fig. 14/5 (*b*)) is part of an exponential curve and the time scale is therefore not uniform. If the condenser is charged at a constant current, the potential difference across it rises at a uniform rate and the curvature is eliminated. The simplest device for obtaining a constant current is a saturated diode (page 24). If R_1 in fig. 14/5 (*a*) is replaced by a diode (V_2 in

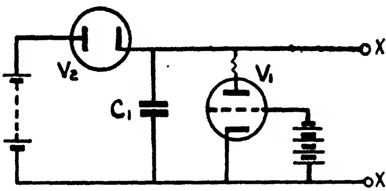


Fig. 14/6. Improved linearity is obtained by substituting a diode valve for the resistance R_1 in fig. 14/5 (*a*).

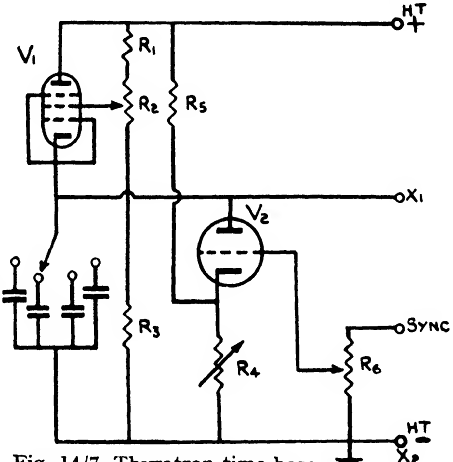


Fig. 14/7. Thyratron time base with pentode valve as the charging resistance.

fig. 14/6) the current flowing in the condenser circuit is almost independent of the potential difference across V_2 and therefore remains constant as C_1 becomes charged. Suppose for example that a high tension supply at 300 volts is used and that C_1 may be charged to a potential of 200 volts before the thyratron conducts. At the beginning of the charge, the potential difference across C_1 is zero and the whole high tension voltage is developed across V_2. When C_1 is charged to 200 volts the potential difference across the diode is 100 volts. The anode voltage of the diode therefore varies between 100 and 300 volts and if a valve is used in which voltage saturation exists over this range, the current flowing is almost constant. With this arrangement the charging current cannot be altered (unless the cathode temperature is changed) and the frequency of the time base may be controlled only by adjusting the value of C_1. It is usually necessary to have a more precise control over the sweep frequency, and a pentodevalve is commonly used instead of a diode. Over the working part of the anode characteristic the current flowing through the valve is almost independent of anode voltage (page 81) and it thus remains constant while the condenser is charging. The anode current is, however, sensitive to changes in the potential of the screen grid and this electrode may therefore be used to control the charging rate of the condenser. With a large value for the screen potential, the pentode behaves as a small resistance and a large current passes, resulting in rapid charging of C_1 and a high sweep frequency. If this potential is reduced the charging current and the time base frequency are both reduced. A typical circuit including these modifications is shown in fig. 14/7. In this circuit the potential applied to the screen grid of the pentode valve V_1 controls the current which charges the time base condenser, connected between anode and cathode of the thyratron V_2. The resistances R_1, R_2 and R_3 form a potential divider for the screen of V_1 and fine adjustment of the sweep frequency is made with R_2, which is variable. The output amplitude is controlled by R_4 which, with R_5, applies a portion of the high-tension supply as a grid bias voltage to V_2. A synchronising signal is applied to the terminals of the potentiometer R_6 and part of it goes to the grid of the thyratron.

It has been explained that, to produce a stationary pattern of a whole number of cycles on the screen, the time base should operate at the frequency of the deflecting voltage, or a submultiple thereof. It is usually not possible to satisfy this requirement for more than a very short time unless a synchronising arrangement is incorporated in the sweep circuit. The method normally adopted is to apply a small part of the deflecting voltage in series with the grid bias of the gas discharge tube. The operation of this arrangement may be explained by reference to fig. 14/8 in which the time base frequency is supposed to be a little below that of the signal under investigation. The grid bias voltage now has an alternation at the signal frequency superimposed on its steady value. The condenser charges along OA and would in the normal way discharge along AB after slightly more than one cycle of the deflecting voltage. With a synchronising signal, the grid bias becomes sufficiently small at X to make the valve conduct and the condenser thus discharges along CD. In the next cycle the discharge which without the synchronising impulse would occur along EF, is initiated at Y. The sweep

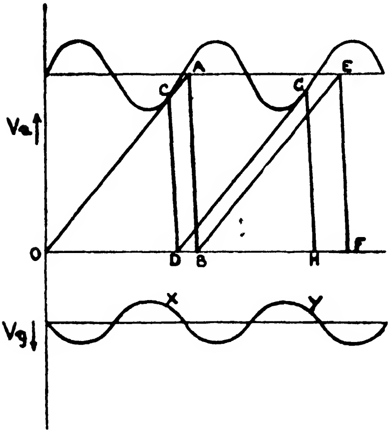


Fig. 14/8. To illustrate synchronisation of a time base.

circuit thus operates at the same frequency as the deflecting signal and minor variations in either do not affect the appearance of the trace. If the time base operates at a frequency higher than that of the signal, synchronisation by the method described is not effective. It has been assumed in this discussion that one cycle of the deflecting waveform is required on the screen ; the time base may equally well be synchronised to a sub-multiple of the signal frequency, in which case the trace contains a number of cycles.

5. **Fluorescent screens.** Characteristics of some commonly-used fluorescent materials are shown in table 14/2.

TABLE 14/2

PROPERTIES OF SOME FLUORESCENT MATERIALS.

Substance	Colour of fluorescence	Persistence	Uses
Zinc orthosilicate (Willemite)	green	moderate	visual observation
Calcium tungstate	blue	very short	photographic observation high-voltage tubes
Zinc phosphate	red	long	visual observation at low frequencies
Zinc sulphide-zinc cadmium sulphide	white	short	television tubes

Double beam oscillograph. It is often convenient to observe simultaneously the variations of two alternating quantities and for this purpose the double-beam tube is employed. The electron stream is divided by a horizontal plate midway between the *Y* plates. The two beams so produced are deflected independently by potentials applied to the upper and lower *Y* plates, and the same time-base is used for both. Alternatively, two complete sets of electrodes, each with its own cathode, may be enclosed in one glass envelope and inclined at a small angle to one another, so that both electron beams reach the single screen. This arrangement has the advantage that time-bases of different speeds may be used for the two beams, though its construction is rather complicated and expensive.

Power supply. The power supply network for a cathode ray oscillograph consists essentially of a transformer and a half wave rectifier valve with simple smoothing arrangements supplying a large potential difference to a high resistance potential

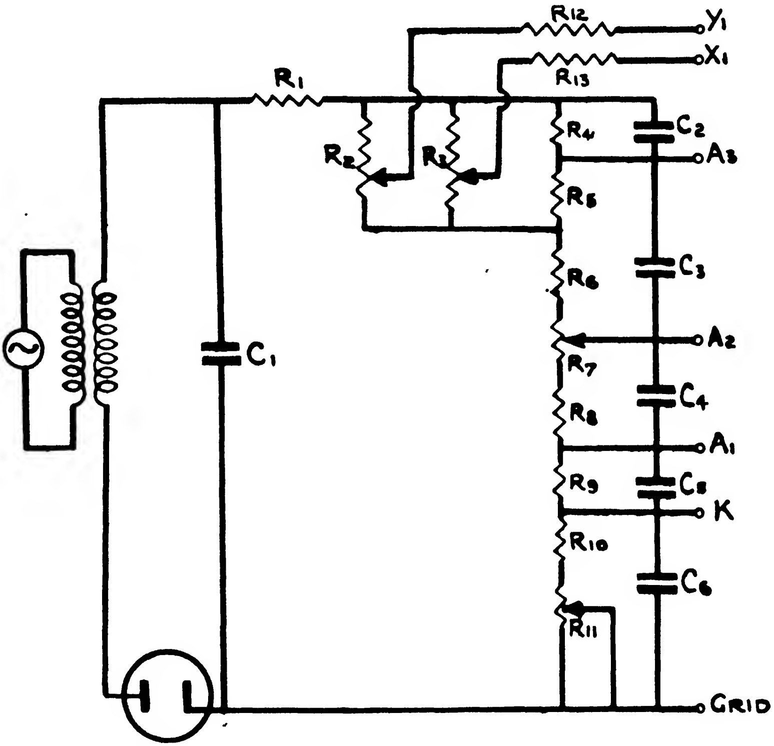


Fig. 14/9. Power pack for a cathode ray oscillograph.
C_1 and R_1 form a simple smoothing network for the rectified output supplied by the diode valve, and the electrodes are connected to appropriate points in the potential divider formed by R_4-R_{11}. R_2 and R_3 are the shift controls, enabling the spot to be centred on the screen. Typical values for the components are :

R_1, R_2, R_3 R_4, R_5, R_7	0.5 megohm
R_8, R_9	0.25 megohm
R_{11}	0.1 megohm
R_{10}	0.01 megohm
R_{12}, R_{13}	4 megohms

C_1—0.1 μ F
C_2, C_3, C_4, C_5—1 μ F
C_6—4 μ F

With a transformer output of 5000 volts R.M.S., the electrode potentials referred to the cathode, are approximately as follows :

Grid : —15 to —150 volts
1st anode : 500 volts
2nd anode : 900-1700 volts
3rd anode : 5500 volts

divider to which the various electrodes are connected at suitable points. In fig. 14/9 the smoothing components are C_1 (0·1 μF) and R_1 (0·5 megohm) and the resistances R_4 to R_{11} provide the voltage divider. The final anode is connected near the positive end of the supply. One of each pair of deflector plates is connected through a 5 megohm resistance to A_3 and the other plates are connected through high resistances R_{12}, R_{13} to the X and Y shift potentiometers which enable the spot to be centred on the screen or, if necessary, to be put in some other position by small displacements in the two directions. A_2 and A_1 are connected to points at lower potentials and the cathode is joined near the low potential end of the divider. The bias voltage applied to the shield is adjusted by means of R_{11} and further smoothing for each electrode is provided by the condensers C_2-C_6.

High voltage oscillographs. In recording phenomena of short duration, such as transients in electric power systems and lightning flashes the electron beam travels at high speed and to obtain an adequately illuminated trace for visual or photographic observation the beam current must be large. If high anode voltages are used the tube must be long—partly to secure adequate insulation between the various connecting leads and partly to secure adequate deflection on the screen. (The deflecting sensitivity is inversely proportional to the accelerating voltage and if a bright trace is required a higher deflecting voltage is therefore necessary.) A tube several feet long is made more conveniently in metal than in glass and this construction is used for high voltage oscillographs. The metal tube has the additional advantage that it can be continuously evacuated, thus allowing

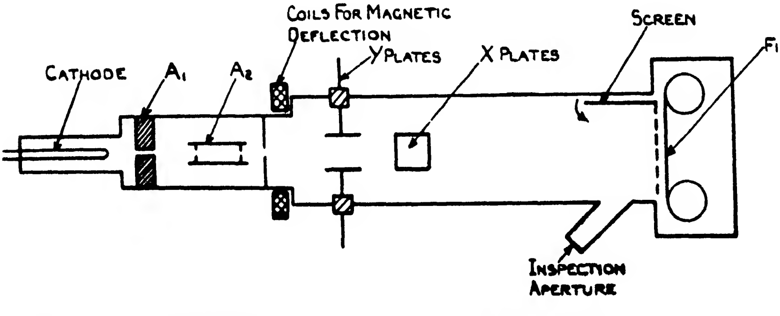


Fig. 14/10. Arrangement of electrodes in a high-voltage oscillograph tube.

replacement of the electrodes of the screen whenever necessary. The life of a metal tube is unlimited whereas the sealed-off type ceases to be useful after about a thousand hours, owing to the falling off in cathode emission.

The construction of a metal tube oscillograph is illustrated in fig. 14/10. A thermionic cathode may be employed though longer life is obtained, with high beam current, by using a metal block or ball as the cathode, extracting electrons in the cold state by applying a high voltage to the first anode. Accelerating arrangements are similar to those described above for the sealed-off tube and if a fluorescent screen is employed it falls into one of the categories enumerated in table 14/2. It is possible to record by allowing the electron beam to fall directly on to a photographic plate or film placed inside the tube. If film is used, it may be wound on a drum which is rotated to provide the time scale. The portion of the tube in which electrons are drawn from the cathode is separated from the main tube by a narrow channel. For cold cathode emission the optimum pressure is about 0·1 mm. mercury, but in the main tube a pressure greater than 0·001 mm. is undesirable because of the possibility of electrons in the beam losing energy by collision with gas molecules in their path. A needle valve is therefore used to allow a slow leak of air into the portion of the tube containing the cathode.

It is customary to use one anode—or two at the same potential—and magnetic focussing is accomplished by means of a coil wound round the neck of the tube near the anode. Though large voltages are needed for accelerating the electron beam, the current drawn by the tube is no more than two or three milliamperes and a satisfactory power supply may be derived from a transformer and rectifier with simple smoothing.

Applications of the cathode ray oscillograph. Since the cathode ray oscillograph may be used for the examination of any alternating or varying quantity it has been put to a great number of uses in physics and engineering. It is not possible here to describe these applications in detail, and all that can be done is to give an account of a few which illustrate the general principles of oscillographic testing and measurement.

1. *Amplifier and transformer tests.* A phase difference between the input and output—phase distortion—of an amplifier

or transformer may be detected by the arrangement shown in fig. 14/11. The input voltage, after amplification if necessary, is applied to the X plates and the output provides the *Y* deflection. A Lissajous figure is formed on the screen. If the output and input voltages are identical in phase and waveform the figure is a straight line and 45° to the horizontal. Amplitude distortion, caused by overloading, is shown by curvature of the trace near its ends and if any phase difference is present the trace opens out into an ellipse. A more precise investigation of these effects is possible with the double beam tube.

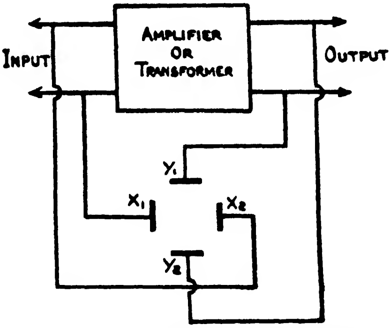


Fig. 14/11. Arrangement for amplifier or transformer tests with the cathode ray oscillograph.

2. *Valve characteristics.* A circuit suitable for plotting $I_a - V_g$ characteristics is shown in fig. 14/12 (*a*). An alternating voltage of 50–100 volts provides the sweep and a portion of it is applied to the grid of the valve. The anode current is indicated by the potential difference across a load resistance *R* connected to the *Y* deflecting plate. For $I_a - V_a$ characteristics the circuit of fig. 14/12 (*b*) may be used.

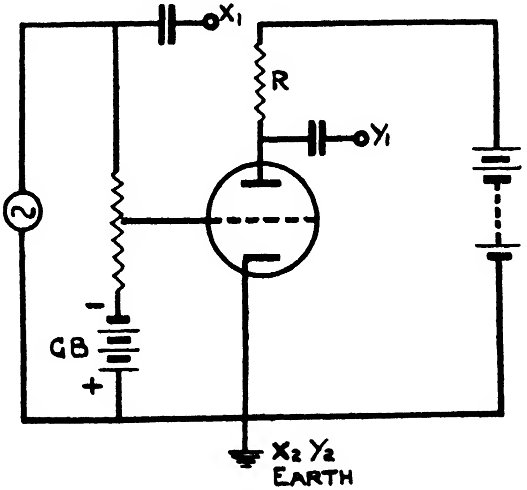


Fig. 14/12 (*a*). Circuit for plotting grid characteristics of a valve on the screen of an oscillograph.

3. *Modulation measurements.* Application of a modulated deflecting voltage gives a trace of the form shown in fig. 14/13, from which the modulation index may be calculated. The arrangement of fig. 14/14 is widely used and gives traces of the kind shown.

4. *Echo sounding.* Most materials reflect electromagnetic radiation and the position of an object may therefore be estimated if it is possible to find the time interval between the mission of a radio signal and the reception of its echo. The cathode ray oscillograph enables this time to be measured accurately and has been used, and in researches on the ionosphere—a diffuse region of ionised air in the upper atmosphere which affects the propagation of radio waves, chiefly by reflecting them back towards the earth. To find the height of the layer, a radio signal in the form of a pulse is sent out from a transmitter and recorded on the screen of an oscillograph. The reflected signal is amplified and applied to the same oscillograph. The time base circuit is synchronised by the transmitted pulses and the time delay of the reflected signal may be calculated from the position of the return pulses on the screen. The same principle is used in the detection of aircraft and other objects by radar. In this case, by using directional aerials for transmitting and receiving, the direction and position of the object may be estimated, the apparatus constituting a searchlight with radio waves instead of light waves.

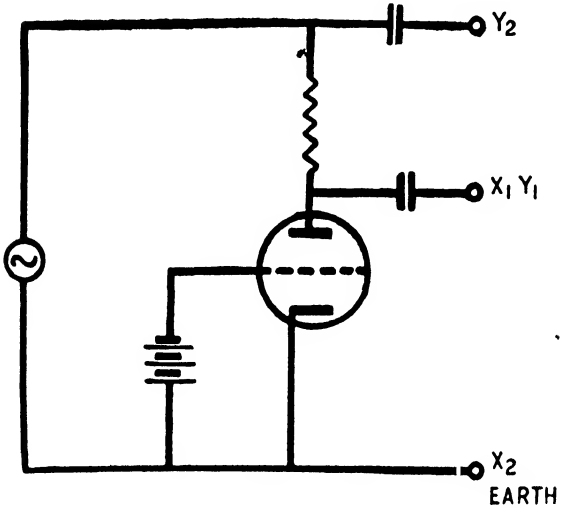


Fig. 14/12 (*b*). This arrangement is used for plotting anode characteristics.

5. *Electrocardiography.* Movement of muscles is accompanied by electrical changes which, in the case of a large mass of tissue such as the heart, are of measurable amplitude. Due to the contraction and expansion of the heart, an alternating potential difference of about a millivolt may be detected between the right and left arms or between one arm and one leg. The frequency of the alternations is that of the heart-beats and the waveform

is used for the diagnosis of many pathological conditions (fig. 14/15). Electrocardiograms may be recorded photographically, using a string galvanometer to move a beam of light over a revolving photographic paper, or with a cathode ray oscillograph. If the oscillograph is used a very slow time base is necessary and a zinc phosphate screen is usually employed, on account of its long afterglow.

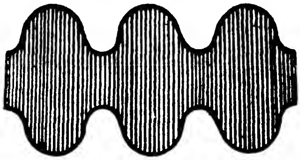
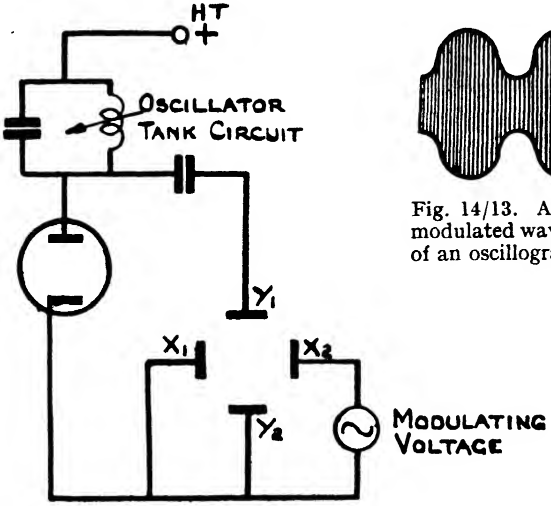


Fig. 14/13. Appearance of a modulated wave on the screen of an oscillograph tube.



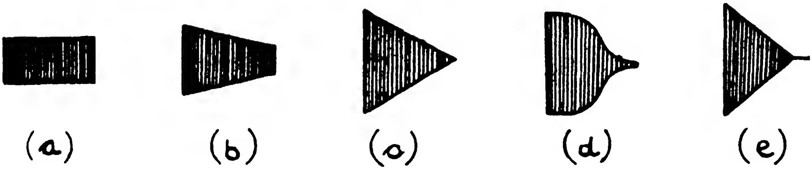


Fig. 14/14. This circuit is commonly used for modulation measurements and gives patterns as follows—(*a*) no modulation, (*b*) normal modulated wave, (*c*) 100% modulation, (*d*) non-linear modulation, due to distortion in the modulator or in an earlier stage, (*e*) overmodulation.

Leads from the patient may be connected directly to the string galvanometer if this instrument is used for recording, but if a cathode ray oscillograph is employed, an amplifier must be included in the apparatus. A gain of 50,000 is usually adequate and the frequency response should be uniform from 1 c s. to 50 c s.

6. *Electroencephalography.* The brain generates electrical energy in small amounts and an irregular alternating potential difference is developed between two electrodes placed on the scalp. Its amplitude is usually between 10 and 50 microvolts and the waveform is complex, with a number of components at frequencies between 1 c/s and 50 c/s ; the most prominent is at about 10 c/s in normal subjects. A record of the brain potential is called an electroencephalogram and is usually obtained in the form of an ink tracing made on a moving strip of paper by a pen which is attached to the armature of an electromagnet. The current energising the magnet is derived from the output of an amplifier, which has leads from the patient connected to its input terminals. A very large gain—10^6 at least—is required and it is customary to use five or six stages, often with a separate power supply to each.

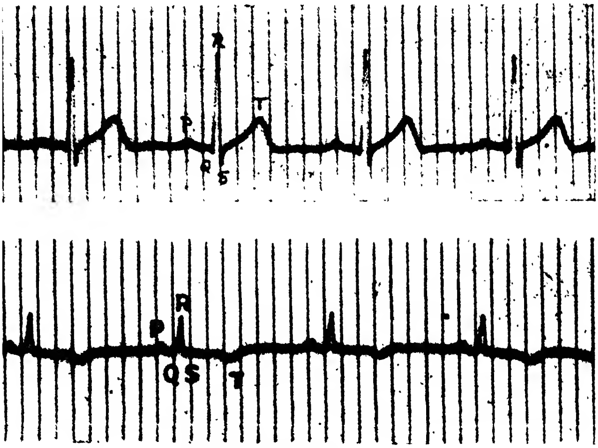


Fig. 14/15. Electrocardiograms. Above—normal.
Below—abnormal.

(By courtesy of Messrs. A. C. Cossor Ltd.)

7. *Pressure recording.* For the recording of mechanical pressures a condenser pick-up is commonly used. This device consists of a condenser with one plate fixed and one which moves in accordance with the pressure to be investigated. The resulting variation in capacitance is changed into an alternation of potential difference across a resistance, (fig. 14/16), and after amplification this voltage is applied to the deflector plate of a cathode ray oscillograph.

The condenser element is usually made in the form of a small cylinder which may, for example, be screwed into an engine cylinder so as to record the gas pressure during the cycle. A similar device may be used to record accelerations in vehicles and aircraft.

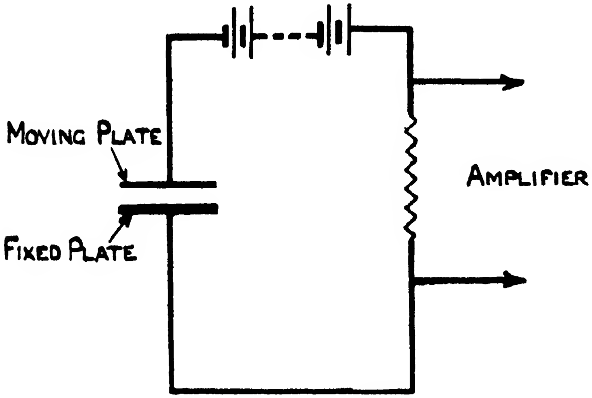


Fig. 14/16. Illustrating the method by which a mechanical displacement is changed into a potential difference suitable for electrical recording. The potential difference supplied by the battery is divided between the resistance and the capacitance. If the capacitance alters, the potential difference across it, and therefore the p.d. across the resistance, is changed.

For investigating higher pressures, such as those in structures, and for experiments in connection with the propagation of explosion waves the piezo-electric pick-up is commonly used. This device consists essentially of a quartz crystal which generates a small electromotive force if subjected to mechanical strain in a certain direction (p. 110), and it takes the place of the condenser used in the low-pressure apparatus. The potential difference across the crystal has the waveform of the pressure which causes it and may be applied to an oscillograph after suitable amplification.

CHAPTER XV

TELEVISION

RADIO communication most often involves transmission of sound by modulating an electromagnetic carrier wave. It is also possible to send intelligence in the form of an optical image from one point to another, again by means of electromagnetic radiation. This process is called television and it is at once apparent that many of the problems involved are quite different from those encountered in sound broadcasting. Thus the transmission of speech or music causes the generation at the receiving station of a series of sounds which, heard consecutively, convey to the ear a replica of the sounds originally delivered to the transmitter microphone. If, however, an optical image is to be transmitted, it is necessary to present the whole of the picture at one instant to the receiver. This is a difficult problem and would in fact be insuperable but for the phenomenon of persistence of vision which is displayed by the eye. The variable quantities concerned in picture transmission are intensity, or amplitude, and colour, corresponding to frequency in sound. A scene to be televised contains in general a wide range of both intensity and colour, but for the moment it is sufficient to consider only the transmission of variations of intensity. With this limitation, an optical image of a scene or object may be defined completely by specifying the brightness at each point of it. The transmission of a complete image at one time is not practicable and the procedure adopted is to divide the scene into small elements of equal size over each of which the illumination is approximately uniform. A series of electrical impulses corresponding in amplitude to the brightness of each successive element is obtained by means of photo-electric cells and transmitted by modulation of a carrier. The image is reassembled at the receiving station after demodulation. The process of dividing the scene into elements is called *scanning* and transmission in this way is possible because of (1) the limited resolving power of the eye which enables a collection of small separate images to be perceived as a single large picture and (2) the persistence of vision which causes the receiving screen to appear completely illuminated, though actually only a small portion of it is receiving light at any instant.

To take full advantage of these properties of the eye it is necessary to scan the object at a rate sufficiently fast to avoid flicker. If about 25 successive complete images are thrown on to the screen per second the eye cannot appreciate that they are separate images and it is thus possible to reproduce scenes containing moving objects. It is also necessary that the elements into which the picture is divided shall be too small for the eye to separate them at the normal viewing distance. If these conditions are fulfilled it is practicable to reproduce with adequate size and detail any sufficiently illuminated scene at the transmitting station.

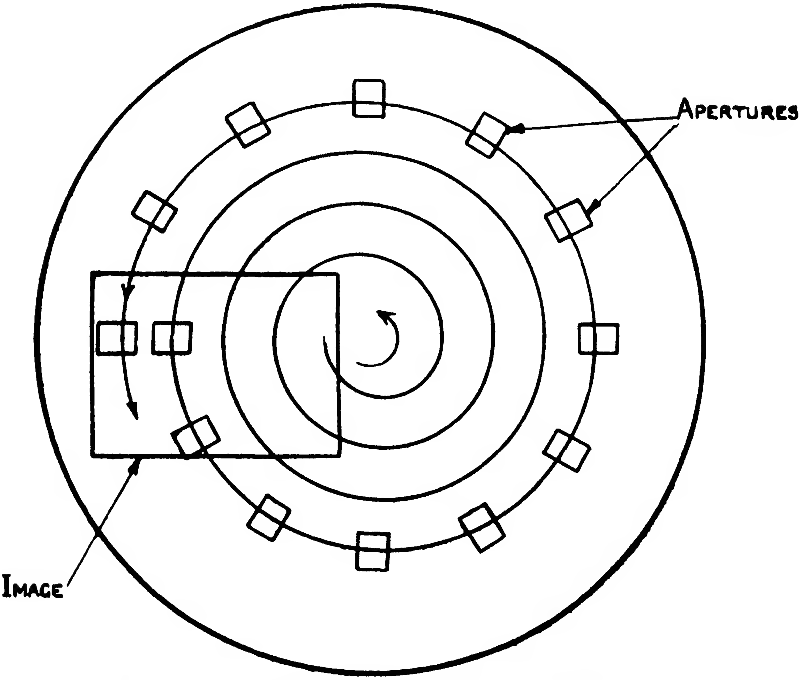


Fig. 15/1. Illustrating the principle of the scanning disc used in early television systems.

Scanning. In early television systems, scanning was performed by a mechanical device. The method is as follows. Light from the scene or object is focused by a lens and forms an image on the scanning disc, which rotates at 25 revolutions per second The disc is perforated with circular or rectangular holes which lie on a spiral, (fig. 15/1.) The size of the image is adjusted so that

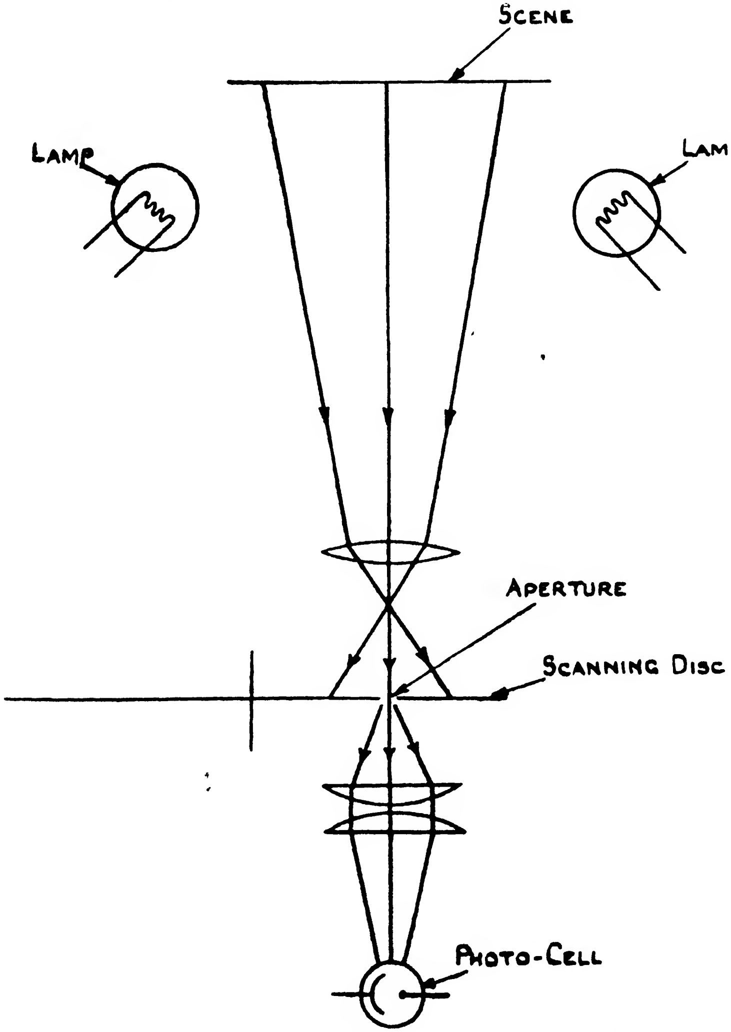


Fig. 15/2. Arrangement of scanning disc and optical system in early television transmission.

the light from each point of it passes through only one hole in the disc during each revolution, (fig. 15/1.) Each aperture thus sweeps across a strip of the image and the light transmitted is brought to a focus on the sensitive surface of a photoelectric cell by another lens, (fig. 15/2.) The current output from the cell at any instant has a value proportional to the brightness of that part of the image in front of the appropriate aperture of the scanning disc at the same instant. The whole image is scanned once in each revolution and the photoelectric cell delivers a fluctuating current corresponding to the variation of intensity over the image. Increased illumination of the received image may be obtained if the photo-electric cell and the lamps of fig. 15/2 are interchanged. Light then reaches the object through one hole in the scanning disc and is reflected from it on to the photo-electric cell. The narrow beam sweeping rapidly over the object can be made much more intense than a broad beam illuminating a greater area continuously.

The output from the photo-electric cell is amplified and used to modulate a carrier wave. At the receiving station the output from a demodulator stage is amplified and used to vary the brightness of a neon lamp, which is viewed through a scanning disc synchronised with that at the transmitter. In this way an image of the original scene is built up in the plane of the scanning disc at the receiver. Other mechanical systems, not essentially different in principle, effect the process of scanning by means of rotating mirrors and such devices. They all suffer from the disadvantage that high definition is unattainable. To reproduce considerable detail in the received image it is necessary that the object shall be divided by the scanning process into small elements. In other words a large number of lines must be employed. In the methods described above this means that the scanning disc has a large number of small holes. The illumination of the photo-electric cell is thus less intense and the brightness of the received image is correspondingly reduced. It is therefore possible to increase the definition at the receiver only by reducing the illumination of the image. For broadcast television, mechanical arrangements in the transmitter and receiver have been replaced completely by electronic devices.

The Iconoscope. Electronic scanning is performed with the iconoscope, fig. 15/3. This consists essentially of a cathode ray tube with a photo-sensitive plate *P* instead of the usual fluorescent screen. This plate consists of a layer of caesium globules on a mica plate with a silver backing. Each globule constitutes the anode of a photo-emissive cell—the cathode of the tube being the other electrode—and with the silver backing forms a small capacitance. An image of the object or scene to be televised is formed by lenses on the sensitive surface of P and the electron beam produced in the usual way is swept across this image by electromagnetic deflection with coils X and Y.

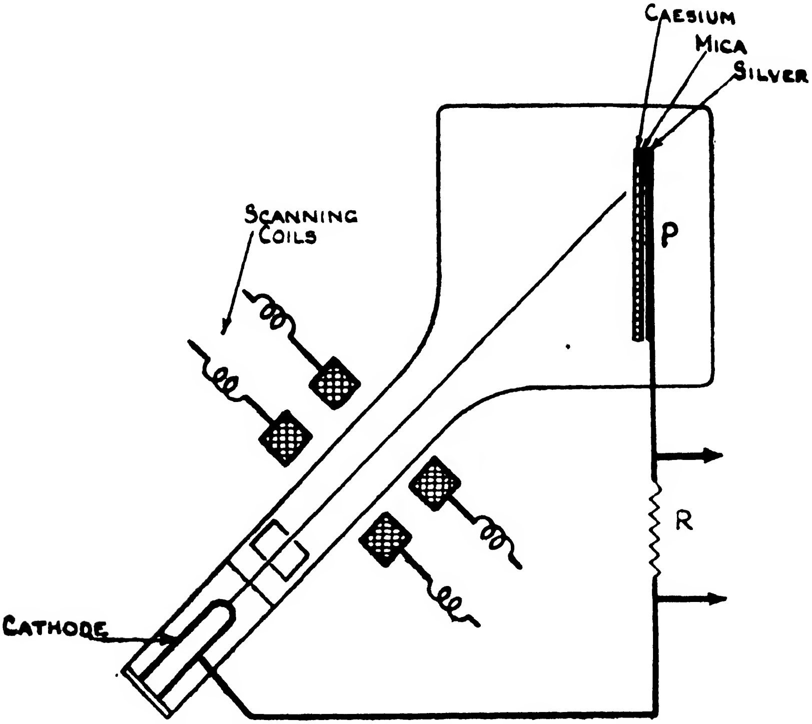


Fig. 15/3. The iconoscope, used for electronic scanning in television.

When the image is formed on P, electrons are ejected from each caesium globule at a constant rate depending on the brightness of the image at the point concerned and the condenser formed by the caesium with the silver backing becomes charged. The electron beam sweeps across every point of the image and in passing restores to each element of the mosaic the electrons which have been lost by it in photo-electric emission. This causes the condenser to discharge suddenly and a burst of current flows in the resistor R connected between the plate and the cathode of the tube. The potential difference developed across *R* fluctuates in a manner corresponding to the variations in brightness as the electron beam sweeps across the image and, after amplification, is used to modulate a carrier. The principal advantage of the iconoscope is its high sensitivity. The current produced when the scanning beam falls on an element of the image is proportional not to the intensity of illumination at that moment, but to the total amount of light which has fallen on the element since the beam last touched. Thus the caesium-silver condensers store the energy falling in the form of light on to the plate and release it periodically as a current which flows back to the cathode through *R* whenever the circuit is completed by the electron beam. The high sensitivity permits rapid scanning, resulting in high definition, and it is possible to televise exterior scenes without artificial illumination. Another and widely used type of iconoscope employs secondary emission to obtain greater sensitivity. In this case, the mosaic is coated with a suitable substance and the scanning beam promotes secondary emission from each element as it passes. The portions of the mosaic which have, since the previous transit of the beam, lost many electrons by photo-electric emission suffer a correspondingly smaller loss through secondary emission, and brightly illuminated parts of the image give current minima in the output which is collected by an extra electrode in the form of a ring kept at a positive potential and placed just in front of the mosaic plate. The image is scanned in horizontal strips, or lines and the whole pattern is referred to as a frame. The rate at which scanning takes place can be indicated by specifying the frame frequency and the number of lines per frame.

Sidebands in television. Any given television waveform may be resolved into a component at the carrier frequency and numerous side frequencies. The extent of these sidebands may be estimated by an approximate calculation. Assume that the object to be scanned consists of alternate black and white squares, arranged in a chess-board pattern, and of side equal to one line-width. Then if the ratio width/height for the complete frame (known as the aspect ratio) is r and the number of lines is n, the number of elements in each frame is rn^2 and with a frame frequency of f the number of elements scanned in a second is frn^2.

The output from the iconoscope has the form shown in fig. **14/2** and the fundamental frequency of the modulating current is therefore

$$\tfrac{1}{2}frn^2.$$

Typical values for the quantities involved are $n = 405$, $f = 25$, and $r = 5/4$, so that the fundamental modulation frequency is approximately 2.6 megacycles/sec.

This calculation is not in any way complete, but it gives a modulation frequency of the correct order. In this case, the sidebands, considering only the fundamental frequency, extend to 2.6 mc/s above and below the carrier. The carrier frequency must therefore be high, so that the tuned circuits in the transmitter and receiver may respond to both carrier and sidebands. The B.B.C. system used a frequency of 45 mc/s.

Interlaced scanning. The extent of the sidebands and therefore the width of the frequency channel occupied by a television transmission is proportional to the frame frequency. To avoid flicker it is desirable to use a higher value than 25 frames per second, but if this change is made, retaining a sufficient number of lines per frame to give adequate definition, the sidebands become inconveniently wide, leading to difficulty in the design of transmitter and receiver amplifiers, as well as to a limitation in the number of stations which can operate in the rather small band of suitable frequencies. For these reasons, a system of interlaced scanning is used in the B.B.C. transmission. On its first journey, the scanning beam of electrons covers alternate lines in the frame; after 202½ lines (in a 405-line frame) the beam is returned to the top of the mosaic, where it completes the unfinished line and then proceeds to fill in those which were omitted previously. In this way the flicker produced is that corresponding to 50 frames per second, which is not objectionable to the eye, and the sideband width is the same as though 25 complete frames instead of 50 half frames were scanned in every second. The principle of interlaced scanning is illustrated in fig. 15/4.

The scansion is performed by the use of time base circuits similar to those described in Chapter XIV, the electron beam being swept across the mosaic slowly and returned quickly for the beginning of the next line, 405 times per second. Superposed on this deflection is a vertical sweep at 50 times per second to return the spot to its original position at the end of each frame. Electromagnetic deflection is commonly used and coils are wound round the neck of the tube for the purpose ; the time bases are designed to give a large current output rather than the voltage output necessary in electrostatic deflection.

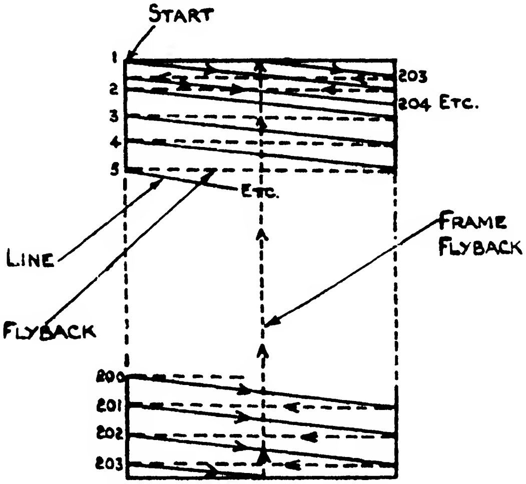


Fig. 15/4. Interlaced scanning.

Reception. At the receiver, an image is built up on the fluorescent screen of a cathode ray tube. The electron beam is swept across the screen in synchronism with the scanning beam at the transmitter. The received signal, after demodulation and amplification, is applied to the shield or grid of the tube and used to vary the brightness of the spot, thus reproducing the variation of intensity over the original scene. To ensure accurate synchronism between the transmitter and the receiver time bases, a number of pulses are sent out from the transmitter, some at the end of each line and others, of different period, after each frame. These pulses are separated from the remainder of the incoming signal by suitable circuits in the receiver and are applied as synchronising voltages to the line and frame time bases which, if adjusted originally to the correct frequency, are then locked to the exact value required. The envelope of the wave form transmitted has the shape shown in fig. 15/5 from which the position and dimensions of the synchronising pulses may be seen. Maximum amplitude of the modulated* wave corresponds to maximum

---

* The term modulation as applied to television has not the same meaning as in radio communication. In the latter case, the mean amplitude of the wave after modulation is equal to the amplitude of the carrier. In television the mean amplitude of the wave has no particular value and the depth of modulation or percentage modulation at any instant is simply the relative amplitude of the wave.

intensity and zero illumination produces an amplitude of 30 per cent. of the maximum value ; the range between zero and 30 per cent. is used for the synchronising pulses.

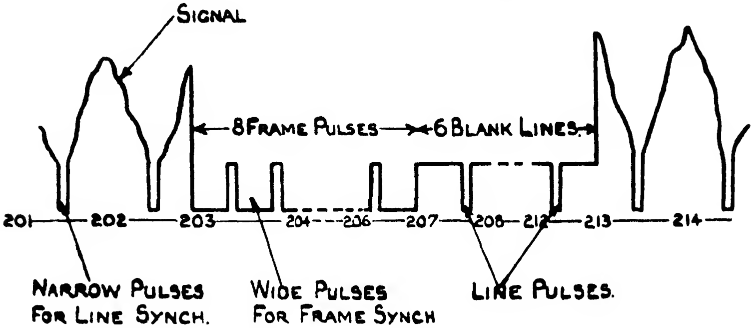


Fig. 15/5. Waveform of the pre-war B.B.C. television signal. Since this illustration was prepared, television transmissions have been resumed and the waveform has been slightly modified in certain details.

Television transmitters. The television transmitter consists essentially of an oscillator generating the carrier frequency and a modulator, to which the iconoscope output is applied. Apart from the problem of obtaining enough energy at the high frequencies involved, there are no points of great interest in the design of the oscillator and its amplifiers. The iconoscope signal, which has components at frequencies ranging from 25 cycles/sec. to at least 3 megacycles/sec., cannot be amplified by conventional methods without some loss at the ends of the frequency scale. A resistance-capacitance coupled amplifier is the most satisfactory, and its diminution in gain at high frequencies is compensated for by the method shown in fig. 15/6. An inductance in series with the anode load resistance has a

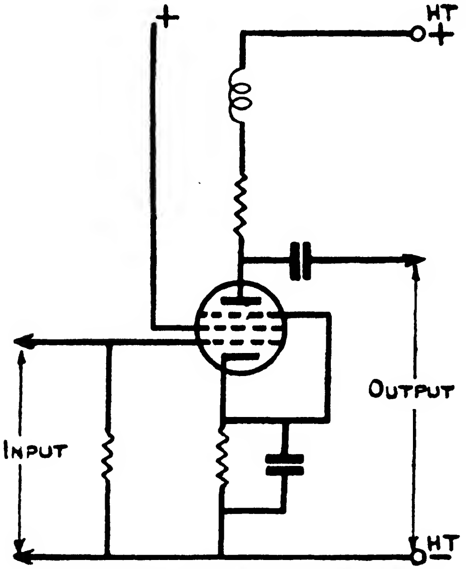


Fig. 15/6. The high frequency response of a resistance-capacitance coupled amplifier may be improved by inserting an inductance in series with the anode load.

value which causes it to resonate with the anode-cathode capacitance of the valve at some frequency well above the upper limit of the range to be amplified. The anode load is then equivalent to a rejector circuit and its impedance rises as the frequency increases, thus compensating for the reduction in gain due to stray capacitances. By suitable design it is possible to keep the gain constant for frequencies up to 5 megacycles/sec.

Television receivers. The television receiver is invariably of the superheterodyne type.

Radio frequency amplification. Amplifier circuits with tuned anode load are used, but on account of the wide frequency range the selectivity must be reduced by shunting the anode load with a resistance.

Frequency changing. A high intermediate frequency—usually about 12 megacycles/sec.—is used and a triode-hexode valve is the most suitable for mixing.

Intermediate frequency amplification. This amplifier usually contains three or four stages coupled by mutual inductance. As the response of the intermediate frequency stages must be fairly flat (again on account of the extensive sidebands), one of the coupled circuits is often left untuned, fig. 15/7.

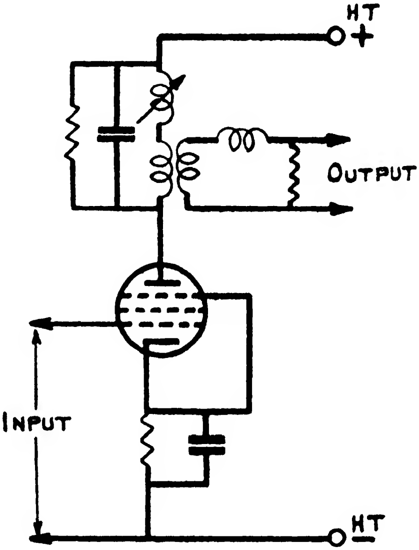


Fig. 15/7. Television I.F. amplifier. Both circuits are shunted by resistances to give a flat frequency response and one of them is untuned.

Detection. Diode detection is used, and the output from the demodulator stage consists of the vision or video-frequency signal. After the detector stage the synchronising pulses are separated from the remainder of the signal by means of valve circuits which it is unnecessary to describe in detail.

Vision-frequency amplification. These stages consist of wide-band amplifiers similar to those described in connection with the transmitter (p. 200). The amplified vision signal is applied between the shield and cathode of the cathode ray tube and varies the brightness of the trace. A range of bias voltage from 0 to −30 is usually sufficient to give adequate contrast in the final image and an output of this order is therefore required from the receiver. As the brightness varies in accordance with the vision signal the electron beam is swept across the screen horizontally and vertically by two electromagnetic time bases similar to those used in the transmitter, and synchronised with them by means of the pulses separated from the vision signal.

CHAPTER XVI

THE ELECTRON MICROSCOPE

THE optical instruments in common use are all based on the phenomenon of refraction, whereby a ray of light changes direction in passing obliquely from one medium to another. The velocity of an electron beam may be altered by subjecting it to an electric or magnetic field and a change of direction results. It is in fact possible to reproduce the optical phenomena of reflection, refraction, interference and diffraction using electrons in place of light rays. The electron lens is an arrangement of conductors designed to produce an electric or magnetic field so shaped as to deflect the electron beam in whatever way may be desired. Both electrostatic and magnetic lenses are used, the former type being common in cathode ray oscillographs and the latter in electron microscopes.

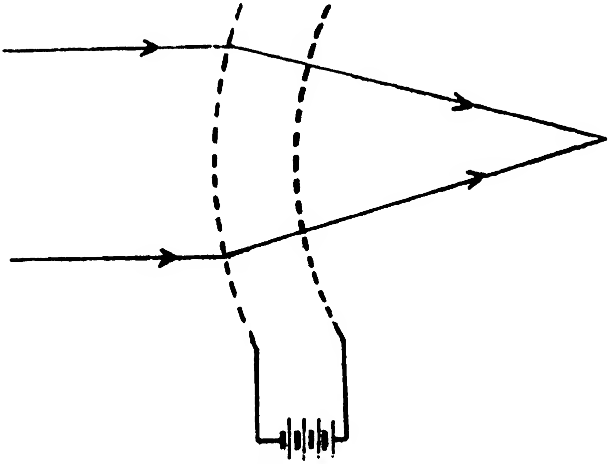


Fig. 16/1. A simple electrostatic electron lens.

Electrostatic lenses. The force exerted on a charged particle in an electric field is in a direction parallel to the lines of force, thus the two gauzes of fig. 16/1 constitute a converging lens for electrons coming from the left. By reversing the potentials a diverging lens may be obtained. Another arrangement is shown in fig. 16/2. Here the electron beam begins to converge at *AA* and would, if not acted on further, come to a focus at *B*. On entering the second cylinder, however, a curvature in the reverse direction is applied. Since the electrons have been accelerated through the potential difference between the two cylinders they are now travelling faster and the divergence produced is not sufficient to cancel the convergence already impressed on the beam which accordingly comes to a focus at a point *C* further along the axis of the lens.

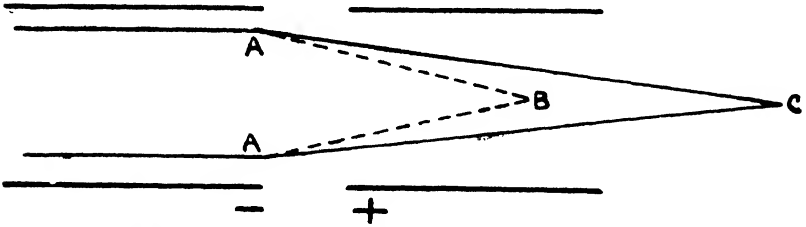


Fig. 16/2. Focussing action of a simple electron lens.

Magnetic lenses. A magnetic field made up of curved lines of force may be produced by sending current through a solenoid and a suitably shaped coil forms the basis of the magnetic lens; two common types are shown in fig. 16/3.

From the foregoing it is apparent that a beam of electrons may be treated in much the same way as a beam of light and the analogy is, in fact, very close. A great amount of theoretical and experimental evidence indicates that the electron and, indeed, any portion of matter, must be regarded sometimes as a particle and sometimes as a wave motion. The wavelength to be associated with an object of mass m and velocity v is h/mv, h being Planck's constant. Rearrangement and substitution of numerical values for the electron leads to the more convenient form

$$\lambda = \sqrt{\frac{150}{V}}$$

where λ is in Ångstrom units and V is the potential difference in volts through which the electron has been accelerated.

Resolving power. The resolving power of an optical instrument, that is, the least separation at which two neighbouring points will give distinct images, is proportional to the wavelength used and inversely proportional to the diameter of the objective lens. If the magnifying power of the instrument is increased beyond a certain value, no further detail becomes apparent unless the resolving power is also improved, and for visible light the greatest magnification which it is practicable to obtain is about 2,000. It is apparent that the use of electrons with equivalent wavelengths of perhaps a hundred thousand times smaller than the wavelength of visible light should make possible a great increase in useful magnification. For reasons which will be explained shortly the increase in magnification is not so great as the reduction in wavelength would suggest, but values up to 100,000—an improvement of fifty times over the light microscope—may be obtained. It is not possible to produce direct visual effects with electrons, but they affect a photographic plate in much the same way as visible radiation and observation on a fluorescent screen is also possible, as in the cathode ray oscillograph.

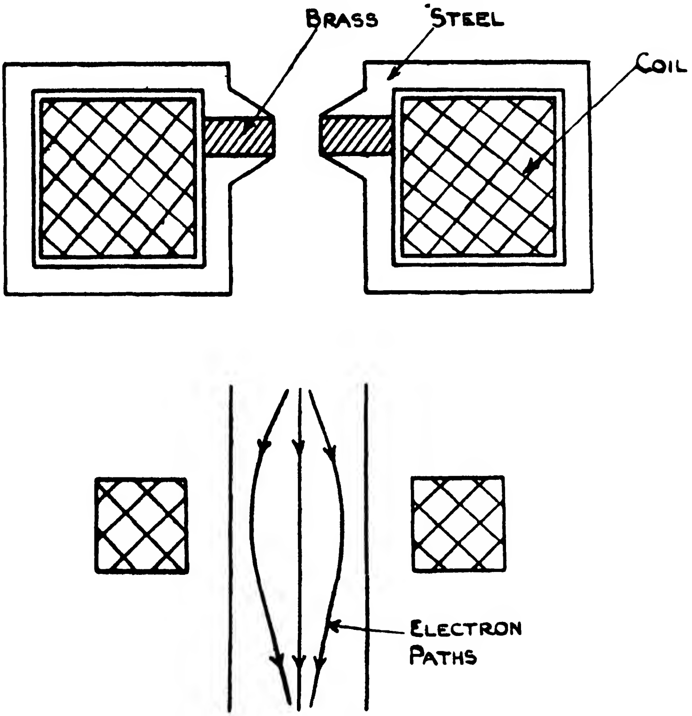


Fig. 16/3. Construction of a magnetic electron lens.

Aberrations. The defects known as spherical and chromatic aberration, characteristic of glass lens systems, are also found in electron optics. Spherical aberration in the visible region arises because, in general, a point source does not give a point image with a lens having spherical surfaces. The difficulty may be overcome in three ways: (a) by reducing the aperture of the lens so that only rays close to the axis are refracted, (b) by the use of two or more separate lenses, and (c) by the use of non-spherical surfaces for the lens. The design of glass optical systems is so far advanced that lenses with apertures of f/1.5 (f = focal length) are common in cameras and microscopes.

Chromatic aberration arises from the fact that the refractive index of glass depends on the wavelength of the light passing through it so that, if white light is used, the rays of different colours do not come to a focus at the same point. This defect may be eliminated almost completely by the use of two or more lenses made from different varieties of glass.

In electron optical systems the design problems are much more difficult because the speed of the beam changes continuously within the lens and the refraction is not localised at two surfaces as in a glass lens. For this reason it has not so far been possible to design an electron lens giving good definition at an aperture greater than about f/1,000. Thus the excellent resolving power of electron optical systems is largely wasted, and will remain unused until lenses with apertures of f/1 or more can be developed. At present an electron microscope gives a magnification of perhaps fifty or a hundred times that of its optical counterpart. This improvement has, however, been of great value in the study of viruses and other micro-organisms and in examining the surface structure of various materials; examples are shown in the frontispiece illustration.

Constructional details. Microscopes using both electrostatic and magnetic lenses have been made on a commercial scale. The RCA model, fig. 16/4, using the latter type, consists essentially of a vertical metal tube divided horizontally into four compartments which are all connected by a pipe to vacuum pumps in the base of the instrument. The top compartment contains a directly heated cathode in the shape of a hairpin with a spot of emitting material at the tip and an anode to which a potential of 60 kilovolts is applied. After acceleration the electrons pass through the condensing lens, emerging as a parallel beam which falls on the object under examination in the second compartment. The specimen must be in a thin film, and various techniques which do not directly concern us have been developed in this connection. The electron beam next traverses the field of the objective lens which has a short focal lěngth and small aperture. The third compartment is separated from the fourth by an eyepiece lens, also of short focal length, which project the image on to a fluorescent screen or a photographic plate.

Magnification. The objective in this instrument gives a magnification of 100 and that produced by the eyepiece may be varied between 20 and 300, giving a range of overall magnification from 2,000 to 30,000 diameters. A further increase of two or three times by photographic enlargement is practicable.

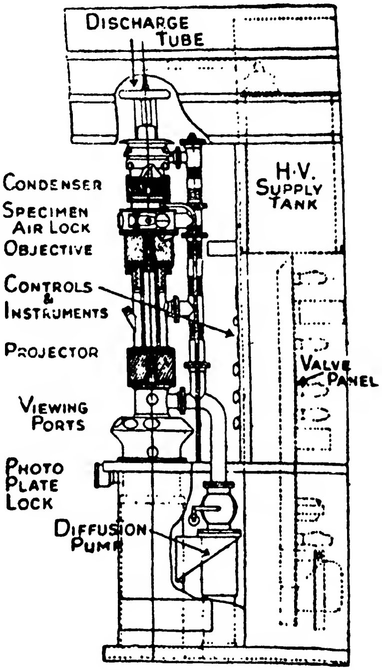


Fig. 16/4. Arrangement of parts in the R.C.A. electron microscope.

Evacuation. The pressure in the tube must be about 10^{-5} mm. of mercury. A three-stage mercury diffusion pump with mechanical fore-pumps is used to exhaust the microscope compartments. To facilitate removal of the specimen or of the photographic plate, air locks are used. The specimen may be moved by a handle into a small region which is simultaneously isolated from the rest of the tube. Air is admitted and the object may then be adjusted or removed. To replace the specimen the air lock chamber is evacuated by an auxiliary pump and then placed in communication with the main compartments again. In this way adjustments may be made without the tedious process of evacuating the whole tube.

Power supply. The electrode potentials and coil currents must be kept constant within very close limits; for a resolving power of 10 A.U. the maximum permissible fluctuations are:—

accelerator voltage	0.015%
objective lens current	0.008%
eyepiece lens current	0.07%
condenser lens current	0.1%

To obtain stability of this order with the customary power supply arrangements (p. 63) would necessitate rigid regulation of the mains voltage—which is difficult and requires inconveniently large smoothing chokes and condensers. For this reason the power supply is derived from an oscillator working at 32 kilocycles/sec. and is then rectified and smoothed in the usual way. At this frequency the smoothing network need not be very bulky and screening is also facilitated, so that the power pack may be mounted close to the rest of the instrument.

INDEX