AN INTRODUCTION TO ELECTRONICS

AN INTRODUCTION TO ELECTRONICS

By

J. YARWOOD

M.Sc., A.Inst.P.

Senior Assistant in Physics

The Polytechnic, Regent Street, London, W.1

LONDON

CHAPMAN & HALL LTD.

37 ESSEX STREET W.C.2

1950

First published, 1950

Catalogue No. 253/4

PRINTED IN GREAT BRITAIN BY
JARROLD AND SONS LTD., THE EMPIRE PRESS, NORWICH

PREFACE

The theory and practice of electronics have been described extensively in many periodicals and text-books, especially in those admirable volumes that frequently appear from the American publishing firms. The author feels, however, that there is yet a place for an English text-book on the subject, particularly to serve the needs of those physics and engineering students who, on leaving college to enter industry, so often find that they need an account of the subject at an intermediate standard, which grows out of the knowledge of electricity and magnetism which they have already acquired. At the same time, it is hoped that this book will help the students at the Universities who are finding that more and more questions on electronics are appearing in their final examinations; and that it will also serve to guide those lecturers who have been giving special courses on this most important subject at technical colleges.

The author makes no claim that original work appears in the text, but has adopted new methods of approach to many topics of fundamental importance. Much has been gleaned from original papers, and the standard works on electronics; a great deal has been learnt from experimental work; and experience has been gained by giving many lectures on electronics and radio-communication to technical college students.

A grateful acknowledgment has to be paid to Mrs. G. Allvey for typing the manuscript, to Mr. L. Holloway, who drew all the diagrams, and to Mr. R. Hercock, B.Sc., for many suggestions.

J. YARWOOD.

CONTENTS

CHAPTER 1

*Fundamental Electricity and Magnetism**

The Electrostatic Field. If two metal plates are arranged parallel to one another, but insulated from each other, and a potential difference is applied across them, as in fig. 1, then a small, positively charged body placed in the space between the plates will move, or tend to move, towards the negative plate.

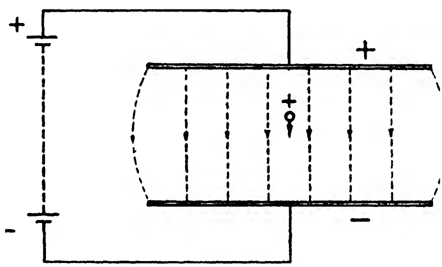


Fig. 1. The Electrostatic Field.

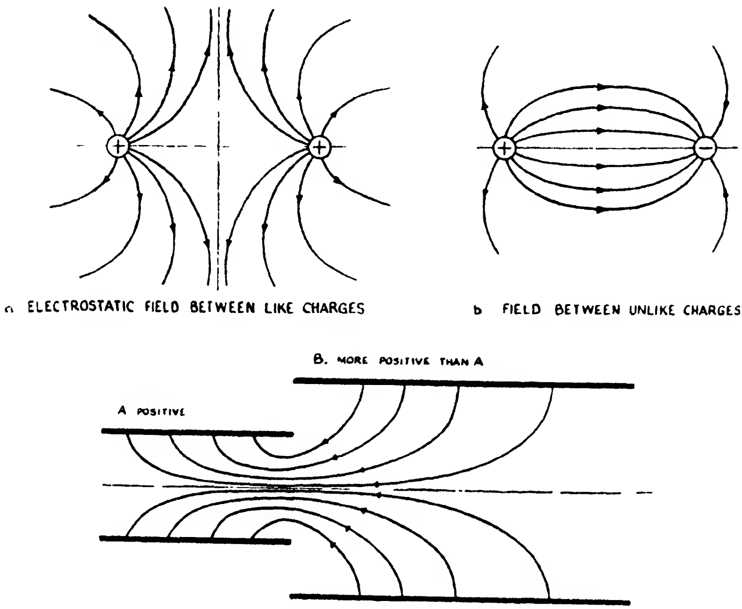
This arrangement is basically similar to that adopted by Millikan† in measuring the charge on an electron. In that case the charges introduced into the field between the plates were induced on small oil droplets ejected from an atomiser.

The force produced on such a charge is due to the electric strain set up between the plates, which is conveniently represented by a pattern of lines of force in the electrostatic field in the space concerned, the direction of a line of force being such that its tangent at any point along it is in the direction of the electric intensity at that point, i.e. the direction in which a free unit positive charge placed at this point will move. In the case considered such lines will be straight, equidistant, parallel lines stretching in the normal direction from the positive to the negative plate; the field will be uniform, except near the edges of the plates. In the general case, the distribution of such lines of force will

* It is assumed that the reader has already a fair knowledge of electricity and magnetism, but a number of important conceptions, which are of particular value in electronics, are collected together in this chapter.

† R. A. Millikan, *Phil. Mag.*, 34, 1, 1917.

depend on the geometry of the electrodes, and the charges upon them, some typical cases being shown in fig. 2.



a ELECTROSTATIC FIELD BETWEEN LIKE CHARGES

b FIELD BETWEEN UNLIKE CHARGES

c ELECTROSTATIC FIELD IN DIAMETRAL PLANE BETWEEN TWO CO-AXIAL CYLINDERS

FIG. 2. Some Electrostatic Field Distributions.

Force Between Charges. Coulomb established the law that two electric charges, of magnitudes q_1 and q_2, placed at a distance d apart in a non-conducting medium would exert a force on one another determined by the relationship

$$F=\frac{q_1 q_2}{kd^2}, \qquad . \quad . \quad . \quad . \quad (1)$$

where k is the dielectric constant of the medium.

Like charges (i.e. both positive or both negative) will repel one another, whereas unlike charges (one positive and the other negative) will attract. This leads to the definition of *unit electric charge* as being such a charge that, placed one centimetre away from an equal charge *in vacuo*, the force of repulsion will be one dyne. Equation (1) can then be written with k unity, i.e. provided the units adopted (see p.16) are correct.

Thus $$F=\frac{q_1 q_2}{d^2}. \quad . \quad . \quad . \quad . \quad . \quad (2)$$

The *electric intensity*, or *electric field strength* at any point in an electrostatic field is the force in dynes which would act on a unit positive charge placed at that point.

The Structure of the Atom. The smallest part of an element that can exist, yet still retain the properties of the element, is termed an atom. It consists of a positively charged nucleus surrounded by negatively charged electrons rotating with large velocities in specific orbits. The number of these electrons, and

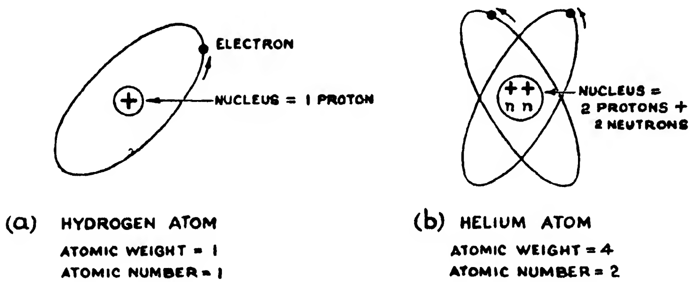


FIG. 3

their orbital paths, determines the nature of the material, i.e. whether it is hydrogen, helium, copper, or gold, or any one of the ninety-two* elements.

The positive nucleus is made up of protons and neutrons. A proton may for the present purposes be considered simply as the smallest possible amount of positive electricity. An electron is likewise the smallest possible negative charge, its charge being equal and opposite to that of the proton. A neutron is electrically neutral: it has no charge, and is therefore not affected by an electric field. It was shown by Moseley† that the number Z, of positive charges on the nucleus, i.e. the number of constituent protons, called the atomic number, was fundamental in determining the nature of the atom, and its behaviour in elements and compound.

Normally an atom is neutral. If the atomic number is Z, then the number of orbital electrons must also be Z. Reference to the Appendix, p. 319, will show, for example, that the copper atom

* There are ninety-six elements if the recently discovered radio-active substances, neptunium, plutonium, curium and americium are included.

† H. G. Moseley, *Phil. Mag.*, **26**, 1024, 1913, and **27**, 703, 1914.

possesses 29 orbital electrons, silver 47, hydrogen 1, aluminium 13; there are definite values of Z for the whole of the ninety-two atoms in existence. Figs. 3*a* and 3*b* show projections on a plane through the atoms of hydrogen and helium, indicating the nature of the orbital arrays.

If a neutral atom loses one or more of its orbital electrons due to the action of some external agency like X-rays, or bombarding electrons, then it becomes positively charged, and is called a positive ion. Both positive and negative ions exist. A singly charged positive ion, e.g. mercury, Hg^{+}, is an atom which has lost one electron, a doubly charged ion, e.g. Hg^{++}, has had its atomic structure robbed of two electrons. Negative ions are usually electrons, but an atom which temporarily gains an electron may also exist as a negative ion. See page 41.

There are some 10^{23} atoms per cubic centimetre of material in a metal. The effective diameter of an atom is between $2{\cdot}5 \times 10^{-8}$ cm. and 5×10^{-8} cm. An electron has a diameter of about 2×10^{-13} cm., whilst the proton is considered to be slightly smaller. An atom has, therefore, a diameter some 100,000 times greater than that of its constituents

The mass of an electron is $9{\cdot}04 \times 10^{-28}$ gm. The proton is 1845 times heavier than the electron, and the neutron 1846 times as heavy. It therefore follows that practically the entire mass of an atom resides in its nucleus, and that a positive ion is much heavier than the usual negative ion,* or electron.

An electric charge is due to the accumulation of a surplus or deficit of electrons on a body, depending on whether the charge is negative or positive. See fig. 4.

The presence of an electric charge, due to the dislodgment of the outer-orbital electrons of some of the atoms of material of the body on which the charge resides, produces a state of strain in the surrounding free space, described as an electrostatic field. If the electron moves, with a consequent attendant motion of the surrounding electric field, then an attendant state of strain in the surrounding space, called a magnetic field, is also produced.

Conducting substances, examples being all the metals and carbon, are those which contain so-called free electrons in their structure. There are one to three such electrons per atom of

* Negative ions are frequently electrons, but it is, perhaps, preferable to use the term negative ion only in the case of an atom, molecule, or group of molecules, which has gained one or more electrons.

material. An insulator, such as sulphur, mica, or ceramic, will contain very few such free electrons for a large number of atoms. The free electrons in a conductor are readily set in motion by the application of a potential difference from a source of electricity across the conductor. On moving, such electrons produce the familiar electric current, with a magnetic field surrounding it.

Dielectrics. Whatever the arrangement of electric charges in a neutral atom may be, they do not produce a resultant field.

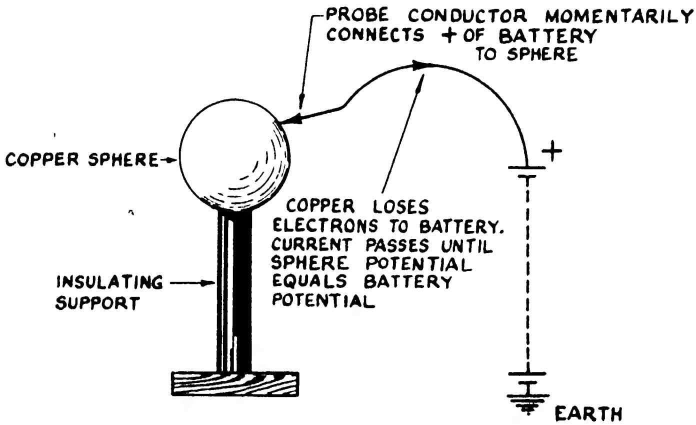


FIG. 4. Formation of an Electric Charge on a Conductor.

However, an external electric field applied to the atom will displace the positive charges in the direction of the field, and the negative electrons in the opposite direction. The extent of these displacements is much restricted by the powerful forces, due to the attraction of its nucleus, which bind the electrons in the atom. The effect of such small displacements is to electrify the atom, producing a *dipole* or electric doublet. Such a dipole will then produce its own electric field in its neighbourhood.

Consider a block of insulator, or dielectric material, e.g. glass, paraffin wax, paper, mica, or aluminium oxide, situated in an electric field X. Since the atoms become dipoles owing to the action of the external field, so the electrostatic field produced by the insulator will be the resultant effect due to the dipoles within it. The insulator is thus electrically polarised, and the extent of the polarisation is directly proportional to the external field.

Thus $$P = cX, \qquad (3)$$

where P is the polarisation of the medium, and c depends on the insulator material, being zero for a vacuum.

In fig. 5, the presence of the external field causes a positive charge to be produced over the surface AB, and a negative charge over CD. If σ is this charge per unit area, then σ is equal to P, the electric polarisation.

If the dielectric were absent, the electric field intensity would

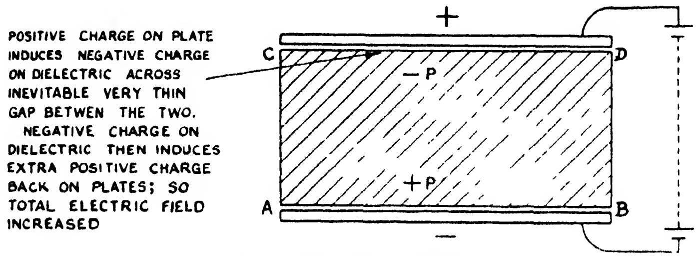


Fig. 5. Electric Polarisation.

be X, but owing to the presence of the dielectric dipoles produced, the intensity within the dielectric has increased to X_D, where

$$X_D = X + aP, \qquad (4)$$

where a is a constant.

X_D is called the *electric induction* within the dielectric.

Potential. The difference of potential between two points A and B is defined as the work done by the electric field on a unit positive charge in moving it from A to B. Suppose the electric intensity is X, then the work done in moving unit charge through a small distance dx is Xdx. The path taken by the moving charge is immaterial as regards the work done on it, since if we assume that in taking a path from A to B more work is done than in going from B back to A by a different path, then work has been gained with the original conditions unchanged. Since this is inconceivable, so the path is immaterial, and it may conveniently be assumed that in moving from A to B the charge traverses a path of which the direction is along the line of force, then

$$\text{work done} = -\int_A^B Xdx = V_B - V_A, \qquad (5)$$

where V_B and V_A are the potentials at points B and A with respect to a point at infinity, the potential at any infinitely distant point from any system of charges being zero.

Thus $$X = \frac{-dV}{dx}, \quad (6)$$

implying that the electric field strength is the rate of change of potential in any direction required.

Let points P_a, P_x and P_b be at distances a, x and b respectively from $+q$. Then the electric intensity at x is $X = q/x^2$.

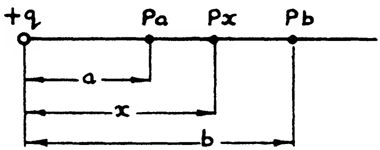


FIG. 6. Potential due to a Point Charge.

Thus the work done in moving unit positive charge from b to a against the force due to q is

$$V_a - V_b = -q\int_b^a \frac{dx}{x^2} = q\left[\frac{1}{x}\right]_b^a = \frac{q}{a} - \frac{q}{b}, \quad (7)$$

where $V_a - V_b =$ P.D. between points a and b.

Since all points at infinity are considered as at zero potential, so the absolute potential at a is $q/a - q/\infty = q/a$.

In electronics, zero potential is usually taken as earth potential, and other potentials are relative to earth as zero.

Gauss's Theorem. This states that the *total normal electric flux over any closed surface drawn in an electric field is* 4π *multiplied by the total electric charge inside the surface.*

The electric flux is defined as the normal component of electric induction multiplied by the area it permeates. If the area considered is small, then the induction can be said to be constant. Electric induction is k times the electric intensity, where the multiplication factor k is brought about by the electric charge being situated in a dielectric of constant k, instead of in vacuum.

In relation to fig. 7 consider the proof of Gauss's theorem. At any point P on the closed surface over which the flux is to be computed, let area ds be small enough to warrant that the electric intensity X is constant over it. Suppose the charge inside the surface is situated at a point O, and the distance OP is r. Let the direction OP make an angle α with the normal to the surface at P, then the electric flux is $X \cos \alpha \,.\, ds$. But X, the force on unit charge, is q/r^2, so that the electric flux equals $q \cos \alpha \,.\, ds/r^2$.

The solid angle $d\omega$ subtended at q by the surface ds is

$$ds/4\pi r^2 \,.\, 4\pi \cos \alpha = ds \cos \alpha / r^2.$$

Therefore the flux for ds is $q \,.\, d\omega$, and since the total solid angle about a point is 4π, so the total flux is $4\pi q$.

If a dielectric other than empty space is considered, then the

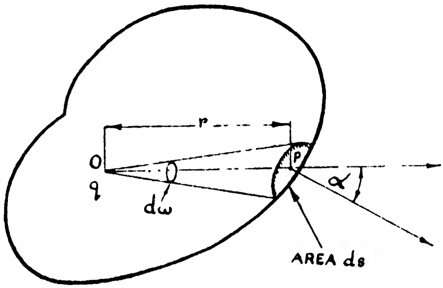


FIG. 7. Gauss's Theorem.

flux is $k \,.\, q/(kr^2) \,.\, \cos \alpha \,.\, ds = q \,.\, d\omega$, which is the same as before, i.e. the presence of a dielectric of constant k makes no difference.

We have assumed here that the charge inside the closed surface is situated at a point. If it is dissipated over a considerable region in space, the result is not vitiated because any charge Q, occupying a finite space, or formed of a number of separate point charges, can always be considered as the sum of a number of point charges, where $Q = \Sigma q$, so that the Gauss rule can be stated as a general law that the total electric flux is 4π times the total charge within the surface.

The electric intensity X can also be quoted in terms of the number N of lines of force, or Faraday tubes, which pass through unit area of a plane drawn at right angles to the direction of the field at the point. Applying Gauss's theorem it can readily be shown that

$$X = 4\pi N, \qquad . \quad . \quad . \quad . \quad . \quad (8)$$

where one Faraday tube arises from each unit charge on the conductor surface.

The Equipotential Surface. A surface drawn to include all points at the same potential in an electric field will be at right angles to the direction of electric intensity at any point since, otherwise, the electric intensity would have a resolved component along the equipotential surface, implying a change of potential within it.

Using Cartesian coordinates, at any point, x, y, z, the potential will be some function of x, y, z, so $V=f(x, y, z)$, and for an equipotential surface

$$V=f(x, y, z)=C, \qquad (9)$$

where C is a constant.

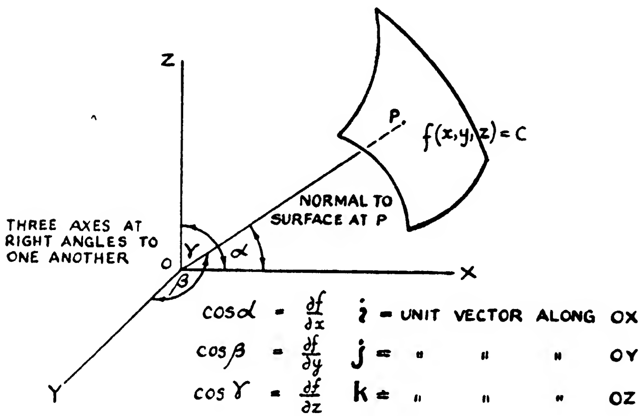


FIG. 8. The Gradient of a Function.

The Gradient of a Function. If any surface in three-dimensional space is represented by $f(x, y, z)=C$, then the variation of constant C will produce a family of surfaces.

In fig. 8 the direction-cosines of the normal to the surface at the point (x, y, z) are proportional to $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$. If $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$* are considered unit vectors in three directions at right angles to one another, then

$$\frac{\partial f}{\partial x}\mathbf{i}+\frac{\partial f}{\partial y}\mathbf{j}+\frac{\partial f}{\partial z}\mathbf{k} \qquad (10)$$

represents the vector in the direction of the normal at (x, y, z).

* Heavy type is used to denote a vector quantity, i.e. a quantity having direction as well as magnitude.

This expression (10) is a vector called the *gradient* of f, written grad. f, or ∇f, where ∇ is called "*nabla*", thus

$$\nabla = \boldsymbol{i}\frac{\partial}{\partial x} + \boldsymbol{j}\frac{\partial}{\partial y} + \boldsymbol{k}\frac{\partial}{\partial z}. \quad . \quad . \quad . \quad (10a)$$

The Electrostatic Potential Gradient. Bearing in mind equation (6) $X = -dV/dx$, it will be realised that such a simple expression is only valid for the variation of potential in one direction in a plane surface. If variation of the potential in three dimensions is considered, then applying equation (10)

$$\boldsymbol{X} = -\text{grad. } V \quad . \quad . \quad . \quad . \quad (11)$$

represents the rate of change of potential in the direction normal

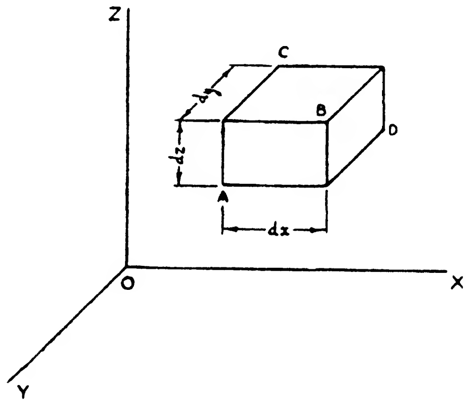


FIG. 9. Concerning Poisson's Equation.

to the direction of the electric intensity, in a three-dimensional system.

Poisson's and Laplace's Equations. In fig. 9 let $ABCD$ be a small rectangular block with sides of lengths dx, dy and dz parallel to three mutually perpendicular axes OX, OY and OZ. Let $\boldsymbol{X}$ be the electric intensity at the point A.

Suppose $\boldsymbol{X}$ is resolved into three components X_x, X_y and X_z respectively, parallel to the axes OX, OY and OZ. The space rate of variation of X_x in the direction OX can be written as dX_x/dx, so that its value at the face BD will be $[X_x + (dX_x/dx)dx]$.

The normal flux over the face AC is $kX_x \,.\, dy\, dz$, where k is the dielectric constant, the area of the rectangular face *AC* being $dy\,dz$. Over the face *BD*, the normal flux will be

$$k[X_x+(dX_x/dx)dx]dy\,dz,$$

and the difference between these two will be the contribution of these faces *AC* and *DB* to the total normal flux over the whole surface of the block. This difference is

$$k[X_x+(dX_x/dx)dx]dy\,dz-kX_x dy\,dz=k(dX_x/dx)dx\,dy\,dz.$$

In the same way, the contributions to the total normal flux of the faces *AB* and *CD*, and *AD* and *BC* will be respectively

$$k(dX_y/dy)dx\,dy\,dz \quad \text{and} \quad k(dX_z/dz)dx\,dy\,dz.$$

But from Gauss's theorem, the total normal electric induction over the block surface must be $4\pi\rho dx\,dy\,dz$, where ρ is the volume density of electric charge, and $dx\,dy\,dz$ is the volume of the block. So equating the total inductions worked out in the two different ways gives, on dividing by the volume, $dx\,dy\,dz$,

$$k\left(\frac{dX_x}{dx}+\frac{dX_y}{dy}+\frac{dX_z}{dz}\right)=4\pi\rho$$

or

$$\frac{dX_x}{dx}+\frac{dX_y}{dy}+\frac{dX_z}{dz}=\frac{4\pi\rho}{k}, \quad . \quad . \quad . \quad (12)$$

and if there is no charge within the block, then

$$\frac{dX_x}{dx}+\frac{dX_y}{dy}+\frac{dX_z}{dz}=0. \quad . \quad . \quad . \quad . \quad (13)$$

Equation (12) is known as Poisson's equation. Equation (13) is Laplace's equation.

The quantity $\frac{dX_x}{dx}+\frac{dX_y}{dy}+\frac{dX_z}{dz}$ is called the divergence of the electric field vector $\boldsymbol{X}$, so equation (12) can be written div. $\boldsymbol{X}=4\pi\rho/k$.

From equation (11), $\boldsymbol{X}=-$grad. V.

$\therefore$ div. $\boldsymbol{X}=-$div. (grad. V)

$$=-\left[\frac{\partial\,(\text{grad. }V)_x}{\partial x}+\frac{\partial\,(\text{grad. }V)_y}{\partial y}+\frac{\partial\,(\text{grad. }V)_z}{\partial z}\right]$$

and since $(\text{grad. } V)_x = \frac{\partial V}{\partial x}$, $(\text{grad. } V)_y = \frac{\partial V}{\partial y}$ and $(\text{grad. } V)_z = \frac{\partial V}{\partial z}$.

$$\text{div. } \boldsymbol{X} = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = \frac{-4\pi\rho}{k}.$$

The Laplacian operator $\nabla^2 = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)$, so equation (12) can be written

$$\nabla^2 V = \frac{-4\pi\rho}{k}, \quad . \quad . \quad . \quad . \quad . \quad (14)$$

and if the charge density is zero, then

$$\nabla^2 V = 0. \quad . \quad . \quad . \quad . \quad . \quad (15)$$

Magnetism. The law of force between two magnetic poles m_1 and m_2, distance r apart, is exactly comparable with the law in the case of electric charges,

$$\text{Force} = A \frac{m_1 m_2}{r^2}. \quad . \quad . \quad . \quad . \quad (16)$$

where A is a constant.

Likewise, a unit magnetic pole is one which placed 1 cm. away from an identical pole in air, repels it with a force of 1 dyne, so (16) becomes

$$F = \frac{m_1 m_2}{r^2}. \quad . \quad . \quad . \quad . \quad . \quad (17)$$

using these units.

If the poles are situated in any other medium, then equation (17) becomes $F = m_1 m_2/\mu r^2$, where μ is the magnetic permeability of the medium.

The force in dynes on a unit north pole placed at a point in a magnetic field is the magnetic field strength, or magnetic intensity at that point, quoted in oersted. Thus a field of strength H exerts a force on a magnetic pole of m units given by

$$F = Hm. \quad . \quad . \quad . \quad . \quad . \quad (18)$$

Note that H, like the electric intensity X, is a force, and therefore has direction as well as magnitude. It is thus a vector quantity, and should be denoted as $\boldsymbol{H}$.

Magnetic potential is again a conception arising like electric potential, defined as a quantity whose space rate of variation in any direction is the strength of the field in that direction. The potential V_B^A between two points A and B due to a north pole m at distances a and b from A and B respectively is proved, as in the electrostatic case, to be given by

$$V_B^A = \frac{m}{a} - \frac{m}{b}, \qquad (19)$$

whilst the potential relative to infinity at a point distant x from a pole of strength m is given by

$$V = \frac{m}{x}. \qquad (20)$$

Since Gauss's theorem in electrostatics rests fundamentally on the application of the inverse square law (1), and this law also applies in the magnetic case, so Gauss's theorem also applies in magnetism.

Thus the total normal magnetic flux over any closed surface drawn in a magnetic field is 4π multiplied by the total magnetic pole strength inside the surface, divided by μ, the magnetic permeability of the medium.

Moreover, for a magnetic field $\mathbf{H}$, with components in three directions at right angles to one another of H_x, H_y and H_z, equation (12) for the electrostatic case becomes, in the magnetic case,

$$\frac{dH_x}{dx} + \frac{dH_y}{dy} + \frac{dH_z}{dz} = \frac{4\pi\rho}{\mu}. \qquad (21)$$

Electromagnetics. If the magnetic flux around a conductor is changed in any manner, then an electromotive force e is induced in the conductor, the direction of the resultant current being such as to oppose the flux change (Lenz's law). Such electromagnetic induction can be expressed mathematically as

$$e = \frac{-dN}{dt}, \qquad (22)$$

where the magnetic flux N changes by an amount dN in the short time interval dt.

The direction of the induced electromotive force is related to that of the motion and the field. This direction can be conveniently remembered by Fleming's Right-Hand Rule, which states that if the thumb and first two fingers of the right hand are extended so as to be mutually at right angles to one another, then if the forefinger is in the direction of the magnetic field (from N. to S. poles), and the thumb in the direction of the motion of the conductor through the field, then the centre finger gives the direction of the current.

The motor principle concerns the reverse effect to that of electromagnetic induction. If a conductor situated in a magnetic field has an electric current passed through it, then the conductor will move. In this case, the direction of motion is given by a Left-Hand Rule, exactly comparable with the rule just quoted, except that the other hand is used. The direction of the current in this case is conventionally assumed to be from positive to negative. It must be remembered that in the case of an electron beam, corresponding to an electric current in a conductor, the action of a magnetic field will also produce movement, but the electrons, being negatively charged, are travelling in the opposite direction to that conventionally adopted for the current direction.

The Line Integral of Magnetic Field. This is defined as the work done in moving a unit magnetic pole along any path from one point to another in the field. If the field strength at any point of the path is $\mathbf{H}$, and the angle between the path of the pole and the line of force is θ, then the force produced on the pole by the field in the direction of motion of the pole is $\mathbf{H} \cos \theta$. In moving along a very short distance dl, the work done is $\mathbf{H} \cos \theta . dl$. In moving over any finite distance between two points A and B in the field, the work done is $\int_A^B \mathbf{H} \cos \theta . dl$. If the pole's motion is always along the lines of force, then this becomes $\int_A^B \mathbf{H} . dl$.

In the case of such a pole travelling round any closed path surrounding a conductor carrying a current i, then the work done can be shown to be $4\pi i$. This result is expressed mathematically as

$$\int \mathbf{H} \cos \theta . dl = 4\pi i, \qquad (23)$$

which can be written as

$$\text{curl } \mathbf{H} = 4\pi I, \qquad (24)$$

when the path encloses unit area and I is the current density.

Maxwell's Electromagnetic Field Equations. According to Maxwell,* the imaginary current consisting of changing electric strain in a dielectric produces a magnetic effect in the same way as a real current. He proved the following equations:

$$\left.\begin{aligned}\mu\frac{dH_x}{dt}&=\frac{dX_z}{dy}-\frac{dX_y}{dz},\\ \mu\frac{dH_y}{dt}&=\frac{dX_x}{dz}-\frac{dX_z}{dx},\\ \mu\frac{dH_z}{dt}&=\frac{dX_y}{dx}-\frac{dX_x}{dy},\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (25)$$

and

$$\left.\begin{aligned}-k\frac{dX_x}{dt}&=\frac{dH_z}{dy}-\frac{dH_y}{dz},\\ -k\frac{dX_y}{dt}&=\frac{dH_x}{dz}-\frac{dH_z}{dx},\\ -k\frac{dX_z}{dt}&=\frac{dH_y}{dx}-\frac{dH_x}{dy},\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (26)$$

where, in a medium of permeability μ and dielectric constant k, x, y and z are three directions mutually at right angles to one another. H_x, H_y and H_z are the components of the magnetic field in the directions x, y and z respectively, and X_x, X_y and X_z are corresponding components of the electric field.

These results were obtained by generalisation of the results (12), (21) and (22), quoted above, in conjunction with Maxwell's initial hypothesis. They show the manner in which the electric and magnetic variations set up in the medium are interdependent.

In the case of the propagation of a plane electromagnetic wave, the electric and magnetic intensities exist only in a plane at right angles to the direction of travel of the wave. It can be proved from Maxwell's equations that the electric and magnetic vectors are then necessarily at right angles to one another, and to the direction of propagation of the wave, which is proved to have a velocity of $1/\sqrt{(k\mu)}$, equal to c, the velocity of light in free space.

Two fundamental laws in electrodynamics are:

(*a*) The motion of an electric field of intensity $\boldsymbol{X}$ with velocity v perpendicular to $\boldsymbol{X}$ in a medium can be proved to produce a

* See S. G. Starling, *Electricity and Magnetism*, Longmans, 1937.

magnetic field perpendicular to $\boldsymbol{X}$, and the direction of motion of $\boldsymbol{X}$, of value $\boldsymbol{H}=k\boldsymbol{X}v$.

(*b*) Again, if the magnetic intensity $\boldsymbol{H}$ moves with velocity v perpendicular to $\boldsymbol{H}$, then it produces an electric field perpendicular to $\boldsymbol{H}$ and v, of value $\boldsymbol{X}=\mu\boldsymbol{H}v$.

Electrical Units. There are three systems of units commonly employed: (1) the electrostatic (es.), (2) the electromagnetic (em.) and (3) the practical systems.

The electrostatic system is based upon the definition of unit electric charge in accordance with equation (2) for the force between charges, whereas the electromagnetic system is founded on equation (17), relating the force between magnetic poles.

In the first case, let $k=1$ for free space, then $F=q_1q_2/r^2$, and unit charge becomes that charge which repels a similar one at a distance of 1 cm. with a force of 1 dyne. But $1/\sqrt{(k\mu)} =c=3\times10^{10}$ cm./sec. from electromagnetic theory, hence if $k=1$, $\mu=1/c^2=1/(9\times10^{20})$.

Likewise, if $\mu=1$ is put in equation (17), then the definition of unit magnetic pole follows, and $k=1/c^2=1/(9\times10^{20})$.

From a development of such reasoning, and the definitions of practical units, the following relationships can be proved:

E.M.F. 1 es. unit $=3\times10^{10}$ em. units $=300$ volts.

Current. 1 es. unit $=\dfrac{1}{3\times10^{10}}$ em. unit $=\dfrac{1}{3\times10^{9}}$ amp.

Resistance. 1 es. unit $=9\times10^{20}$ em. units $=9\times10^{11}$ ohms.

Capacity. 1 es. unit $=1$ centimetre $=\dfrac{1}{9\times10^{20}}$ em. unit $=\dfrac{1}{9\times10^{11}}$ farad.

Inductance. 1 es. unit $=9\times20$ em. units $=9\times10^{11}$ henrys.

Charge. 1 es. unit $=\dfrac{1}{3\times10^{10}}$ em. unit $=\dfrac{1}{3\times10^{9}}$ coulomb.

Magnetic Field Strength. The unit is called the oersted.* It is a field of such strength that a force of 1 dyne is exerted on unit pole situated in it. The name gauss is frequently used instead of oersted. The em. unit of field strength is equal to one oersted, equals 3×10^{10} es. units.

* In accordance with the International Convention, the oersted is employed as the unit of field strength, whereas gauss is the unit of magnetic induction. Popular usage has not yet standardised this practice.

CHAPTER 2

The Emission of Electrons from Conductors

The Free Electron. The passage of an electric current is explained as due to the movement of free electrons through a conductor, such electrons being an intrinsic part of the structure of the material. In the case of the metals and carbon, all of which are good electrical conductors, there are two or three free electrons per atom, whereas insulators like mica and bakelite possess very few free electrons amongst a large number of atoms. From this standpoint, logical theories and explanations of most of the ordinary direct and alternating current circuit phenomena can be developed. However, in the case of electronics, which deals primarily with the motions of these electrons through gases at reduced pressures under the influence of various electrostatic and magnetic fields, controlled by the effects of associated circuits, a more detailed study of the free electrons in conducting materials is necessary, so as to ascertain the factors which govern the release of electrons from materials into the surrounding space. The effect of the number of molecules of gas per cubic centimetre in this space, which is decided by the pressure, also needs consideration. The freedom of movement of electrons in a gas is decided by the degree of packing of the molecules.

Free electrons are readily set in motion from one atom to another of a conductor to form an electric current on the application of the smallest potential difference. It would seem that electrons could just as readily leave the conductor's surface and escape into the surrounding space. There is, however, a potential barrier at the surface of conductors which prevents this effect from taking place. Additional energy has to be supplied to the electrons to enable them to penetrate this barrier.

This energy is supplied by any of four methods in practical electron tubes:

(1) *Thermionic Emission.* By raising the temperature of the conductor.

(2) *Photo-electric Emission.* By electromagnetic radiation in the form of infra-red, visible or ultra-violet light, or X-rays.

(3) *Secondary Emission.* By irradiation of the conductor surface with other electrons moving with sufficient velocity through the surrounding space.

(4) *Cold Emission, or Field Emission.* By the application of an intense electric field at the conductor surface.

In the usual case, such methods of arranging to release the free electrons from a conductor by imparting to them sufficient energy to enable them to break through the surface barrier potential, are applied in a vacuum to allow the electrons a subsequent passage unimpeded by the presence of too many gas molecules.

The Potential Barrier. The charge on the nucleus of an atom is Ze, where Z is the atomic number, and e is the charge on one proton. The potential at a distance r from the nucleus is therefore Ze/r,* and the potential energy of an electron at such a distance from the nucleus will be $-Ze^2/r$, since the electron charge is $-e$. This is only a valid argument if the distance r is so small that, for the particular atom concerned, there are no other electrons within the distance r. In the case of an outer-orbital electron, considered as completely outside the other $(Z-1)$ electrons, the potential will be due to both the positive charge Ze on the nucleus, and the negative charge $-(Z-1)e$ on the other orbital electrons. The combination of these two effects gives a potential of $-e/r$, and the potential energy of this outer electron is simply $-e^2/r$. Clearly, as points are considered in various regions amongst the orbital electrons, there will be considerable variations in potential in the neighbourhood of the nuclei of a conductor. From the macroscopic point of view a conductor is considered as being at a uniform potential, but microscopically considered it is certainly not so, close to the nuclei.

Consider the effects at the surface of a conductor (fig. 10*a*). Suppose the point P_2 is just inside the surface, whereas P_1 is just outside. The potential energy at P_1 is practically zero because it is relatively distant from the nuclei, whereas the potential energies of the electrons inside the metal, near P_2, are negative. Work must necessarily be done to enable the electrons to escape through the surface barrier layer: their potential energies must be increased, and it is evident that a surface potential effect exists. Only those loosely bound electrons in the outer orbits of the atom,

* See p. 7.

having a small negative potential energy, will be able to receive sufficient energy by any of the four methods detailed above to enable them to escape. The inner orbital, or bound electrons, will have such great negative potential energies that they will not receive sufficient energy* by these methods to enable them to surmount the surface potential wall.

The Work Function. Fig. 10*b* indicates another way of regarding the effects of the potential barrier at a conductor surface.

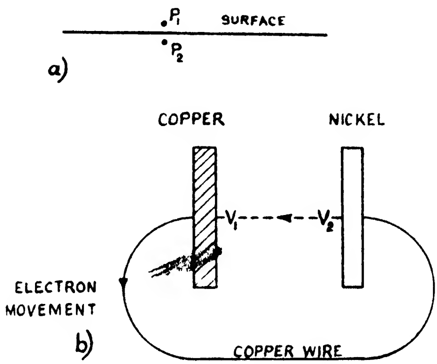


FIG. 10. *a*, The Potential Barrier at the Surface of a Conductor
b, The Work Function.

Here a relationship is established between barrier potential and contact potential difference. Suppose the two plates are of copper and nickel, and are connected by a copper wire. Consider that an electron is moved from the copper plate through the wire to the nickel plate, and then across the space between the plates back to the copper plate. Let V_2 be the potential just outside the nickel, and V_1 the potential just outside the copper. In accordance with the ideas of O. W. Richardson† let ω be the amount of work that has to be done on an electron to enable it to escape from the metal. This defines the term, work function. If e is the electron charge, and U the corresponding potential difference, then the necessary work is Ue. If ω_1 and ω_2 are the work functions, with corresponding potentials U_1 and U_2 for the copper and nickel in the above circuit, then since the total work done in moving an

* There is an exception in the case of X-rays, which are capable of causing the release of inner orbital electrons.

† See O. W. Richardson, *The Emission of Electricity from Hot Bodies*, Longmans, 1921.

electron round the path must be zero since it returns to its initial position, it follows that

$$e(V_1 - V_2) + (\omega_1 - \omega_2) = 0$$

or

$$e(V_1 - V_2) + e(U_1 - U_2) = 0, \quad . \quad . \quad . \quad (27)$$

i.e. the difference between the work functions, expressed as potentials of the two metals, is equal to the contact potential difference between them.

The work function, then, is the amount of energy in joules per coulomb that must be imparted to an electron to enable it to escape from the conductor surface. Since 1 joule/coulomb is equivalent to 1 V., so work functions are usually quoted in volts. For the various metals commonly employed in electronics the values are:

Tungsten	4·52	Platinum	4·40	Tantalum	4·30
Molybdenum	4·30	Carbon	4·00	Thorium	3·35
Nickel	2·77	Calcium	2·24	Strontium	2·00
Barium	1·70	Caesium	1·36		

In the following table a qualitative idea of contact potential difference is given on considering that any element in the series will be positive in potential to any element lower in the table, the value of the potential difference depending on how far apart in the series the two elements appear.

Caesium, rubidium, potassium, sodium, barium, strontium, calcium, magnesium, aluminium, zinc, iron, nickel, cadmium, tin, mercury, silver, gold, platinum, carbon, selenium.

A somewhat different manner of regarding the conception of work function is to consider that kinetic energy must be imparted to an electron within the conductor before it can surmount the potential barrier at the surface. This kinetic energy will be equal to $\frac{1}{2}mv^2$, where m is the mass of the electron, and v is the minimum velocity which the electron must have before it can get to the outside of the conductor. If, however, an electron of mass m and charge e undergoes motion due to a fall through potential of amount U then the velocity it acquires is obtainable by equating the potential energy in the field to the kinetic energy gained by the electron, thus

$$\tfrac{1}{2}mv^2 = Ue. \quad . \quad . \quad . \quad . \quad (28)$$

Instead of quoting the energy necessary in joules, it can be given in volts, where it is understood that the potential quoted is that through which the electron would have to fall to acquire the velocity and kinetic energy specified. This manner of regarding work function, and its consideration as a voltage, gives the same results as the previous method. In a like manner, electron velocities are often quoted in volts, or electron-volts, which may not be a particularly good practice, since the volt is not a unit of velocity or of energy, but which is a very convenient one (see p. 44).

The Kinetic Theory of Gases. Before proceeding with the various methods of producing electron emission, it is best to consider the physics of the gas into which the electrons are released. Moreover, the physical concepts involved in the kinetic theory as applied to gas molecules apply to a considerable extent in the case of a cluster of free electrons in a vacuum: the kinetic laws to be deduced will be applicable to both cases.

The molecules in a gas are considered to move in all directions, the effects they produce at the walls of the container being the pressure of the gas. Simplifying assumptions are made that these molecules are infinitesimally small compared with the distances between them; also that collisions between molecules, and of the molecules with the walls of the container are perfectly elastic so that no energy is lost at such collisions. Further, it is assumed that when a molecule of gas strikes the container wall at an angle, then the angle of reflection equals the angle of incidence. This last assumption is not at all justified in the case of a single molecule, but considering the average effects for a very large number, then the assumption of equality of these angles is valid.

Let a molecule have mass m and velocity v. By virtue of the above conditions, v will remain constant throughout the motion.

Consider a cubical vessel of sides d containing the gas, and relate the molecular motion to three Cartesian coordinate axes OX, OY, OZ parallel to the three pairs of faces of the cube.

Suppose the direction of motion of the molecule considered is such that it makes angles with the three axes which have direction-cosines of l, m and n relative to OX, OY and OZ respectively (fig. 11).

In unit time, the distance traversed in the x-direction$=vl$, in y-direction$=vm$, and in z-direction$=vn$.

The number of collisions in x-direction in unit time will be vl/d, since d is the distance, in the x-direction, between the cube faces.

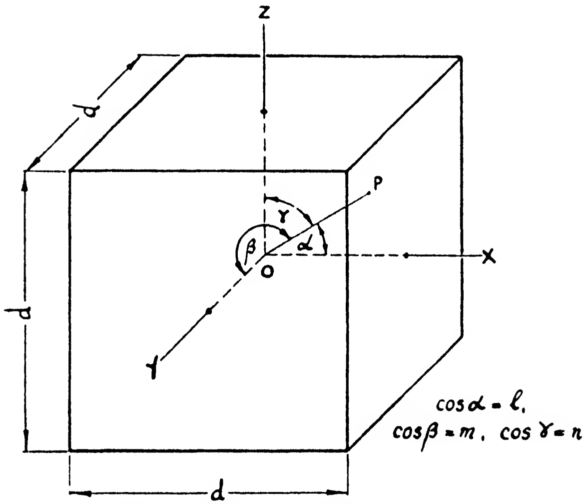


FIG. 11. The Kinetic Theory of Gases

The transfer of momentum on impact between one molecule and the face of the vessel $=2mvl$.

$$\text{Total for all impacts in unit time}=2mvl.\frac{vl}{d}$$

$$=\frac{2mv^2l^2}{d}.$$

Similarly for the molecular motion in the y- and z-directions so that, in unit time, the total transfer of momentum at all six faces of the cube is

$$\frac{2mv^2}{d}\left(l^2+m^2+n^2\right)=\frac{2mv^2}{d}, \qquad . \quad . \quad (29)$$

since the sum of the squares of the direction-cosines is unity (a standard trigonometrical formula).

Consider, now, a number of molecules of masses m_1, m_2, m_3, etc., and velocities v_1, v_2, v_3, etc., respectively.

The summation for the total momentum transfer for all such molecules gives $2/d(m_1v_1^2+m_2v_2^2+m_3v_3^2+\text{etc.})$.

Dividing by the area of the faces involved ($6d^2$) gives the total force per unit area, or the gas pressure, p.

$$p=\frac{1}{3d^3}\left(m_1v_1^2+m_2v_2^2+m_3v_3^2+\text{etc.}\right)=\frac{1}{3d^3}\sum_{n=1}^{n=\infty} m_n v_n^2, \quad (30)$$

i.e. pressure = two-thirds of kinetic energy per unit volume, since $\frac{1}{2}m_1v_1^2$, $\frac{1}{2}m_2v_2^2$, etc., are the kinetic energies of the molecules.

Note that the pressure p changes inversely as the volume d^3 provided that the kinetic energies, i.e. the temperature of the gas, remains constant, which is known as Boyle's law.

Put $\bar{V}^2$ = average value of v^2 for all the molecules in the gas, then

$$m_1v_1^2+m_2v_2^2+m_3v_3^2+\text{etc.}=(m_1+m_2+m_3+\text{etc.})\bar{V}^2.$$

If ρ = gas density, and v the volume, then $m_1+m_2+m_3+\text{etc.}=\rho v$.

So $$p=\tfrac{1}{3}\rho\bar{V}^2. \quad . \quad . \quad . \quad . \quad . \quad (31)$$

This equation enables the average molecular velocity $\bar{V}$ to be calculated if the gas pressure and density are known.

When two gases are brought into contact, if there is no nett transfer of energy from one to the other, implying that the two gases must be at the same temperature, then the mean translational energies of their individual molecules must be equal.

Hence $$\tfrac{1}{2}m\bar{V}^2=\tfrac{1}{2}m'\bar{V}'^2. \quad . \quad . \quad . \quad . \quad (32)$$

where m and $\bar{V}$ apply to one gas, and m' and $\bar{V}'$ to the other.

If the pressures of the two gases are also equal, and n and n' are the numbers of molecules in unit volume of the two gases, then from equation (31) $mn\bar{V}^2=m'n'\bar{V}'^2$, since $\rho=mn$; combining with equation (32) gives $n=n'$.

So any two perfect gases at the same temperature and pressure contain the same number of molecules per unit volume. This is Avogadro's law.

Avogadro's number is the number of molecules in a gram-molecule, 1 gram-molecule being the molecular weight in grams of the gas. This quantity can be determined in various ways, but probably the most accurate is to determine the charge carried by 1 gram-molecule of singly ionised molecules in solution, and divide by the charge on the electron, as determined by Millikan. Avogadro's number, which is constant in accordance with the kinetic theory, is found to be $6{\cdot}06 \times 10^{23}$, denoted by N.

If V is the volume occupied by a gram-molecule of gas, then from equation (31)

$$pV = \tfrac{2}{3} \cdot \tfrac{1}{2} mN\bar{V}^2.$$

By experiment it is known that

$$pV = RT, \qquad . \quad . \quad . \quad . \quad (33)$$

where R is the gas constant, p, V and T being respectively the pressure, volume and absolute temperature of a perfect gas. R is proved to have the same value for all such gases.

From the last two equations it is seen that

$$\tfrac{3}{2} RT = \tfrac{1}{2} mN\bar{V}^2. \qquad . \quad . \quad . \quad . \quad (34)$$

Thus the gas constant R is equal to two-thirds of the total kinetic energy of the molecules in 1 gram-molecule of a gas at temperature of 1 degree absolute (1° K.).

By means of equation (34) the value of $\bar{V}$, the root mean square velocity of the gas molecules, can be calculated at a given temperature merely by knowing the empirically determined value of R, which is found from measurements of the gas pressure, volume and molecular weight.

For example, for nitrogen the molecular weight is 28, and the root mean square velocity calculated from the foregoing is found to be 58,400 cm./sec. at 0° C., and 76 cm. pressure.

The gas constant per molecule is $R/N = k$. k is known as Boltzmann's constant, equals $1{\cdot}37 \times 10^{-16}$ erg/degree.

So far only the average velocity of the molecules in a gas has been considered. It is necessary to find the most probable distribution of velocities amongst the gas molecules to which the gas will tend when an equilibrium is established.

If there are n molecules, of which one has velocity components u, v and w parallel to the three mutually perpendicular axes, then the number of molecules having component velocities which lie between the limits u and $u+du$, v and $v+dv$, w and $w+dw$ is

$nf(u, v, w)\, du\, dv\, dw$, where f is some function of u, v and w to be determined.

Maxwell* has shown that this function f has the form $A\varepsilon^{-hm(u^2+v^2+w^2)}$, A and h being constants, and ε being the exponential base 2·71828.

This result can be stated in more convenient form in terms of the number of molecules whose velocities lie between c and $c+dc$, where c is the resultant velocity having components u, v and w, and is given by

$$4\pi n A^3 \varepsilon^{-hmc^2} c^2 \,.\, dc, \qquad (35)$$

n being the number of molecules of mass m in unit volume of gas at temperature T.

A can be shown to equal $(hm/\pi)^{\frac{1}{2}}$, and $h=N/2RT=1/2kT$, k being Boltzmann's constant.

Therefore formula (35) becomes

$$4\pi n\left(\frac{hm}{\pi}\right)^{3/2} \varepsilon^{-hmc^2} c^2 \,.\, dc$$

$$=4\pi n\left(\frac{m}{2\pi kT}\right)^{3/2} \varepsilon^{-mc^2/2kT} c^2 \,.\, dc. \qquad (36)$$

If values of the function, equation (36), are plotted against c a curve exhibiting a maximum is obtained.† This maximum is the most probable velocity, here called α, which is given by differentiating (36) with respect to c, and equating to zero.

$$\therefore \quad d/dc(\varepsilon^{-hmc^2} c^2)=0,$$

$$\therefore \quad c^2 \varepsilon^{-hmc^2}(-2hmc)+\varepsilon^{-hmc^2} 2c=0,$$

$$\therefore \quad c=0, \quad \text{or} \quad c=\infty, \qquad \text{giving minima,}$$

or

$$c(-2hmc)+2=0,$$

$$\therefore \quad c=\sqrt{(1/hm)}=\alpha, \qquad \text{giving the maximum,} \qquad (37)$$

where α is the most probable speed.

The *mean free path* of a gas molecule is the average distance a molecule travels in a gas at a given pressure before it collides with another. If d is the diameter of a molecule, then if two molecules approach so that their centres are at distance d apart, they have collided. Thus the target area of a molecule is πd^2. If there

* See J. H. Jeans, *The Dynamical Theory of Gases*, Cambridge, 1925.

† See fig. 12, p. 31.

are n molecules per unit volume, then the number in a thickness t of cross-section S is ntS, and their target area is πnd^2tS. When $\pi nd^2tS=S$, the target formed by the molecule fills up the whole area S, so that no molecule can traverse the distance t without making a collision.

When $\pi nd^2tS=S$,

$$t=\frac{1}{\pi nd^2}, \quad . \quad . \quad . \quad . \quad . \quad (38)$$

where t is the mean free path.

Applying the results of Maxwell's consideration of the distribution of velocities amongst the molecules, a more accurate result for the mean free path can be shown to be

$$L=\frac{1}{\sqrt{(2)}\,.\,\pi nd^2}. \quad . \quad . \quad . \quad . \quad (39)$$

For the important case of an electron passing through the gas, the mean free path is considerably increased because the diameter of an electron is much smaller than that of a gas molecule. So the mean free path L_e for an electron becomes

$$L_e=5{\cdot}66L. \quad . \quad . \quad . \quad . \quad . \quad (40)$$

For nitrogen at 760 mm. Hg pressure, $L=9{\cdot}44\times10^{-6}$ cm.

But
$$L\propto\frac{1}{\text{pressure}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (41)$$

Therefore at pressure$=7{\cdot}6\times10^{-3}$ mm. Hg, $L=1$ cm. approx.

In a vacuum tube it can be seen from these considerations that the pressure needs to be reduced to below 10^{-3} mm. Hg if an electron is to travel any distance in the gas without collision with a gas molecule, and the creation of an ion. At 10^{-3} mm. Hg the electron free path L_e is given by

$$L_e=5{\cdot}66\times\frac{760}{10^{-3}}\times9{\cdot}44\times10^{-6}=43 \text{ cm.}$$

This distance is greater than the average electrode spacing in a small receiving valve. In larger vacuum tubes, like transmitter valves, cathode-ray tubes, electron microscopes, etc., the necessarily longer electron paths, combined with the greater accelerating potentials employed, enhance the probability of the electron causing ionisation of the gas by collision, so the pressure needs to be reduced to less than 10^{-4} mm. Hg.

In the case of modern thermionic and photo-cathodes, free alkali-earth elements such as barium and strontium appear within the cathode surface. The readily oxidisable nature of these metals makes it imperative to avoid even very minute traces of gas in the electron tube, so that pressures lower than 10^{-4} mm. Hg become necessary.

Thermionic Emission. In connection with thermionic emission a formula due to the work of Richardson* and later Dushman† is

$$I_0 = A_0 T^2 \varepsilon^{-b/T}. \quad . \quad . \quad . \quad . \quad . \quad (42)$$

where I_0 = emission current in amp./sq. cm. of emitting surface,
T = absolute temperature of emitter,
b = constant depending on the work function of the emitter,
A_0 = constant, largely independent of nature of the emitter if it is a pure metal,
ε = exponential base 2·71828.

A proof of this formula, based on the thermodynamical similarity between the emission of electrons from a heated conductor and the evaporation of gas from a heated liquid, is here considered. Other proofs of the equation, particularly more difficult statistical methods, give similar results in a more rigid manner.

If a liquid at a given temperature is within a closed vessel then a state of equilibrium exists in which the number of molecules leaving the liquid surface per second is equal to the number returning to the liquid. A vapour cloud forms above the liquid surface, the pressure of which is a constant at the stated temperature. The same state of affairs exists in the case of a conductor at a sufficiently elevated temperature inside an evacuated vessel. Electrons are emitted from the conductor, and an equilibrium will exist where the number of electrons leaving the conductor equals the number arriving back at the conductor per second. The pressure of the electron gas outside the conductor is a constant

* See O. W. Richardson, *The Emission of Electricity from Hot Bodies*, Longmans, 1921.
† S. Dushman, *Phys. Rev.*, 21, 623, 1923.

depending on the temperature and nature of the emitting conductor. Again, in both these cases work is done when the particle, be it molecule or electron, leaves the surface. In the liquid evaporation case the familiar latent heat effect is involved; in the electron emission case the corresponding effect is involved in the work function energy. Hence the idea of the latent heat of vaporisation of electrons is a useful one. So the Clausius-Clapeyron* thermodynamic equation is applicable to the electron emission problem, provided suitable provisos are specified.

Let p be the pressure of the electron atmosphere when equilibrium is attained; V and U the volume of 1 gram-molecule of electrons in the free space and in the metal itself. One gram-molecule is the molecular weight in grams of a gas. In the case of a gram-molecule of electrons, the mass of the electron is 1/1846 of the mass of the hydrogen atom, and since the molecular weight of hydrogen is 2·016, where there are two atoms in the molecule, so the "molecular weight" of an electron is 1·008/1846. One gram-molecule of electrons is that quantity of electrons which has a weight of 1/2000 gm. approximately.

According to the law of Avogadro, the number of molecules in 1 gram-molecule of a perfect gas is a constant irrespective of the nature of the gas; it is called N. We are therefore concerned with N electrons.

Let L = energy absorbed in the emission of these N electrons, then according to Clausius-Clapeyron

$$L=T\frac{dp}{dT}(V-U). \qquad . \quad . \quad . \quad (43)$$

where T is the absolute temperature of the emitter.

The volume U of the electrons in the metal will be very small compared with their volume V in the free space, so that U is negligible compared with V in the above equation.

The energy L is used in two ways: first, to overcome the work function effect, which is ω per electron, or $N\omega$ per gram-molecule of electrons; secondly, in doing work against the pressure in the atmosphere of electrons into which the emitted electron evaporates, which is pV.

This electron gas can be said to obey the perfect gas laws, since

* See H. S. Allen and R. S. Maxwell, *A Textbook of Heat*, Part II, p. 634, Macmillan, 1939.

the mutual repulsion effects between electrons due to their like charges are negligible forces compared with those that pertain in accordance with their motion as considered in accordance with kinetic theory. It follows that the usual perfect gas equation applies

$$pV=RT, \qquad . \quad . \quad . \quad . \qquad (33)$$

where R is the gas constant for a gram-molecule, which can be put $=Nk$, where N is Avogadro's number, and k is Boltzmann's constant (see p. 24).

$$\therefore \quad pV=NkT.$$

$$\therefore \quad L-N\omega=NkT.$$

Substituting for L from equation (43), therefore

$$VT\frac{dp}{dT}=N(\omega+kT),$$

or, since $V=\frac{NkT}{p}$, therefore

$$\frac{NkT^2}{p}\cdot\frac{dp}{dT}=N(\omega+kT).$$

$$\therefore \quad k\frac{dp}{p}=\frac{(\omega+kT)}{T^2}dT.$$

Integrating, $k\log_e p=\int\frac{\omega dT}{T^2}+\int\frac{kdT}{T}+C$, where C is constant.

$$\therefore \quad \log_e p=\int\frac{\omega dT}{kT^2}+\log_e T+C.$$

$$\therefore \quad p=A_1 T\,\varepsilon^{\int\omega dT/kT^2}. \qquad . \quad . \quad . \quad . \qquad (44)$$

where A_1 is constant.

Since the kinetic theory of gases applies in the case of an electron gas, so the number of electrons n_0 crossing unit area of cross-section in the gas per second will be

$$n_0=DnT^{1/2}. \qquad . \quad . \quad . \quad . \qquad (45)$$

where D is a constant, and n is the number of electrons per unit volume, since the root mean square velocity is dependent on the square root of the temperature (see equation 34).

In equation (44) we can write $p=EnT$, where E is a constant, and n is again the number of electrons per unit volume.

$$\therefore \quad n=A_2\varepsilon^{\int \omega dT/kT^2} \quad . \quad . \quad . \quad (46)$$

A_2 being a new constant.

From (45) and (46)

$$\therefore \quad n_0=A_0T^{\frac{1}{2}}.\varepsilon^{\int \omega dT/kT^2}. \quad . \quad . \quad . \quad (47)$$

If it is assumed that the electron has no energy within the metal, then the total work done in extracting it from the metal must include the energy $3/2kT$ which it has when outside the metal, in accordance with equation (34).

Hence ω takes the form

$$\omega=\omega_0+3/2kT, \quad . \quad . \quad . \quad (48)$$

where ω_0 is constant.

Substituting this value for ω in equation (47)

$$n_0=A_0T^{\frac{1}{2}}.\varepsilon^{\int(\omega_0+3/2kT)/kT^2.dT},$$

$$\int\frac{\omega_0+3/2kT}{kT^2}.dT=\omega_0\int\frac{dT}{kT^2}+\frac{3}{2}\int\frac{dT}{T}$$

$$=\frac{-\omega_0}{kT}+\frac{3}{2}\log_\varepsilon T.$$

$$\therefore \quad n_0=A_0T^{\frac{1}{2}}.\varepsilon^{-\omega_0/kT+3/2\log_\varepsilon T}=A_0T^2.\varepsilon^{-\omega_0/kT}. \quad . \quad (49)$$

Since each electron carries a charge e, the current is $I_0=n_0e$ per unit area.

$$\therefore \quad I_0=A_0T^2.\varepsilon^{-\omega_0/kT}=A_0T^2.\varepsilon^{-b/T}, \quad . \quad (50)$$

giving the Richardson-Dushman form of the thermionic emission equation.

Velocity Distribution amongst Emitted Electrons. The electrons emitted from a conducting thermionic cathode will have a velocity distribution following Maxwell's equations for the molecules of a gas. Therefore the speeds of the electrons will be distributed according to a curve obtained from function (36)

$$=4\pi n\left(\frac{m}{2\pi kT}\right)^{3/2}\varepsilon^{-mc^2/2kT}c^2.dc. \quad . \quad . \quad (36)$$

The electrons inside the cathode, before emission, will not have velocity distributions according to Maxwell, but obey a statistical law due to Fermi and Dirac.* To avoid the difficulty of considering a new conception in mathematical physics, it is assumed here that the electrons within the metal also obey Maxwell statistics. This assumption leads to incorrect quantitative predictions, but simplifies the treatment, and leads to useful ideas.

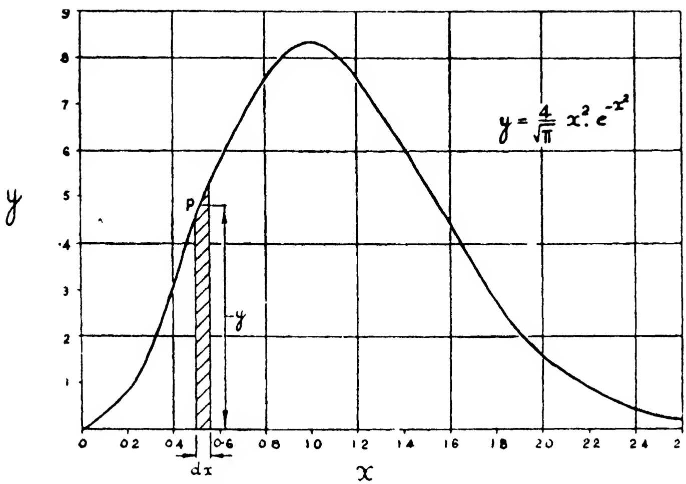


FIG. 12. Distribution of Velocities amongst Electrons.

To simplify (36), put $mc^2/2kT=x^2$ and differentiating this,

$$x\,dx=(cdc\,.\,m)/2kT.$$

$$\therefore\quad dc=\frac{x\,dx}{c\,.\,m/2kT}=\frac{x\,dx}{m/2kT\,.\,x/\sqrt{(m/2kT)}}=\frac{dx}{\sqrt{(m/2kT)}}.$$

Substituting in (36) gives

$$\frac{4}{\sqrt{\pi}}e^{-x^2}\cdot\frac{x^2}{m/2kT}\cdot\left(\frac{m}{2kT}\right)^{3/2}\cdot\frac{dx}{\sqrt{(m/2kT)}}\cdot n$$

$$=n\,.\,\frac{4}{\sqrt{\pi}}\,.\,x^2\,.\,e^{-x^2}dx. \qquad (51)$$

Hence a curve of y plotted against x from the equation $y=(4/\sqrt{\pi})x^2e^{-x^2}$ gives the required velocity distribution (fig. 12).

* See *Electronics* by Millman and Seely, McGraw-Hill Co., 1941, or *Thermionic Emission* by Jones, Methuen, 1936.

This curve can be interpreted as follows. At any point P on it consider an element dx along the x-axis, then the area $y\,dx$ is the fraction of the whole number of electrons which, at any time, have velocities between x and $(x+dx)$ times the most probable speed, the most probable speed being for $x=1$ in fig. 12.

But the most probable speed is given by (37) as $\alpha=1/\sqrt{(hm)}$, where $h=1/2kT$.

$$\therefore \quad \alpha^2=\frac{2kT}{m}=\frac{2RT}{M}, \qquad (52)$$

where R is the universal gas constant $=8{\cdot}3\times10^7$ ergs/gram-molecule. M is the molecular weight number $=1/1850$ for electrons.

$$\therefore \quad \alpha^2=\frac{16{\cdot}6\times10^7T}{1/1850}=30{\cdot}5\times10^{10}T\ (\text{cm./sec.})^2.$$

Hence at 2000° C.=2273° K., the most probable (speed)^2 is 0·2 electron-volt (see p. 43).

For tungsten the work function is 4·52 V. Therefore the number which have (velocity)^2 greater than 4·5 V., i.e. $\sqrt{(4{\cdot}5/0{\cdot}2)}$ $=4{\cdot}7$ times the most probable speed, which decides that $x=4{\cdot}7$, will possess sufficient energy to escape from the metal into the surrounding space. From fig. 12 this fraction of the total number of free electrons within the tungsten is seen to be exceedingly small. On the other hand, if a material of lower work function, say 1 V., is used as an emitter, then electrons with a value of x exceeding $\sqrt{(1/0{\cdot}2)}=2{\cdot}2$ will be able to escape at 2000° C. From fig. 12 it is seen that the area under the distribution curve for values of x greater than 2·2 is very much greater than that area for x greater than 4·7, indicating the enormous benefit achieved by using an emitter of as low work function as possible.

Thermionic Emitters used in Practice. It is evident from a study of the Richardson emission equation (42), and from the previous paragraph, that the achievement of a high emission electron current from a heated conductor *in vacuo* demands either a high operating temperature or a material of low work function. Unfortunately one cannot have both, since the materials of low work function cannot be operated at excessive temperatures in a vacuum without volatilising. The material which was most used in the early days of valve manufacture was tungsten. Though this does not have a low work function, yet it possesses the advantage of being the metal with the highest melting-point, and so can be operated at a higher temperature *in vacuo* than any other suitable conductor. It has practically passed out of use for small types of valve with anode voltages less than 5000, since the high operating temperature demands excessive filament wattage if the emitting area of the cathode is to be at all considerable. The emission in mA./sq. cm./W. dissipated in the heating of the filament is low, implying low efficiency. However, tungsten filaments still find extensive application in the construction of transmitting valves, operating at anode voltages in excess of 5000. In this case, tungsten has a much greater life than any of the more efficient emitting substances, since the cathode is bombarded by the small number of positive ions formed in the inevitable residual gas in the valve. These ions do not cause harm if their energies are low, but when accelerated by potentials in excess of 5 kV. they impinge on the cathode with such vigour that all but a robust metal surface like tungsten will slowly disintegrate.

Two other types of thermionic emitters have been widely used:

(*a*) thoriated tungsten filaments,

(*b*) a barium oxide-strontium oxide mixture coated on to a directly heated filament, or indirectly heated cathode (see p. 79).

(*b*) is by far the most widely used and efficient emitter.

(*a*) Thoriated tungsten filaments are formed from tungsten wire containing 1 to 2% of thorium oxide. This filament is activated by a four-stage procedure: (1) the tungsten is cleaned and the oxide reduced to metallic thorium by electrically heating the filament, to 2800° K. whilst on the vacuum pumps. (2) The temperature is reduced to at 2100° K. for some minutes. Thorium diffuses to the filament surface, and some of it evaporates. (3) The temperature is reduced to 1600° K. and naphthalene vapour is admitted to form a layer of tungsten carbide on the filament surface. This reduces the electron emission somewhat, but effectively reduces the evaporation of thorium from the filament. (4) The normal operating temperature is 1900° K.

The benefit resulting from the thoriation of the tungsten is seen from the work-function figures: pure tungsten 4·52 V., thorium 3·4 V., thoriated tungsten 2·6 V. The presence of the electro-positive thorium layer on the tungsten reduces the work *function by virtue of its ability to lower the potential wall at the tungsten surface. Such thoriated tungsten can be operated for long periods at a temperature at which pure thorium would* quickly evaporate.

(*b*) Langmuir and his associates found as a result of a long series of experiments that an approximately fifty-fifty mixture of barium oxide and strontium oxide was capable of producing a greater thermionic electron emission than any previously discovered substance. This oxide coating can be prepared upon a pure metal like tungsten or nickel, but the emission is improved by the use of an alloy as base metal, such alloys usually containing nickel, cobalt and iron. The preparation is sprayed or deposited upon the supporting metal; the carbonates, nitrates, azides or oxides of the alkali-earth metals being used. An "activation" schedule, carried out during the time the electron tube is on the vacuum pumps, produces centres of the pure alkali-earth metal in a matrix of the oxides. The work function is about 1 V., and the operating temperature is usually between 800° C. and 1000° C.

Performances of Thermionic Emitters

	Emission in mA./cm.^{2}/W.	Work function in V.
Pure tungsten . .	2–10	4·52
Thoriated tungsten	5–100	2·6
Oxide-coated cathodes (indirectly heated) .	10–200	1·0
Oxide-coated filaments .	200–1000	1·0

The minimum and maximum figures given for the emission in mA./sq. cm. of cathode surface per watt of electrical power supplied for cathode heating,* depend on the temperature, and in the case of the oxide-coated types, on the activation schedule and base metal. Oxide-coated cathodes need to be operated within a fairly narrow temperature range: too low a temperature will seriously impair the emission, too high a temperature will volatilise the free alkali-earth metal too rapidly. An indirectly heated cathode valve with a nominal heater voltage of 6·3 should not have a voltage supply greater than 7·5 V. or less than 5·4 V. for long period operation.

* Cathode watts = heater volts × heater current.

Photo-electric Emission. The release of electrons from a conductor into the surrounding space can be caused by electromagnetic radiation in the form of gamma-rays from radioactive sources, X-rays, ultra-violet rays, visible light and infra-red radiation. Most of the common metals only exhibit this effect if the wave-length of the radiation is less than that for the violet light at the short-wave end of the visible spectrum. Thus experiments are best carried out using ultra-violet light from a quartz mercury-vapour lamp, or from a carbon arc.

The Lenard-Einstein* equation for photo-electric emission is

$$h\nu=\tfrac{1}{2}mv^2+\omega, \qquad . \quad . \quad . \quad . \quad (53)$$

where h is Planck's constant$=6{\cdot}62\times10^{-27}$ erg sec., and ν is the frequency of the incident radiation. In accordance with Planck's quantum theory, electromagnetic radiation is emitted and absorbed in discrete energy amounts, called light-quanta, or photons, and the smallest amount of energy associated with radiation of frequency ν is $h\nu$, where h is a universal constant.

Thus the emission of radiant energy takes place in the form of the transference of photons, comparable with the flow of electric current in the form of electrons. Though radiant energy is discontinuous in its nature, yet for most practical purposes the fundamental particles of energy involved are so small as to be insignificant.

In equation (53), $\tfrac{1}{2}mv^2$ is the kinetic energy of the photo-electrically released electron of velocity v and mass m. This is equal to Ve, where V is the negative potential which would be required external to the photo-cathode to prevent the electrons, each of charge e, from leaving.

The work function ω of the material is necessarily involved, and has the same value as in the thermionic emission case. In order that photo-electric emission occurs, $\tfrac{1}{2}mv^2$ must be positive in equation (53), so that $h\nu$ must be greater than ω.

If $\nu=\nu_0$, where $h\nu_0=\omega$, then ν_0 is the threshold frequency below which photo-emission will not take place. Since $\nu=c/\lambda$, where c is the velocity of light and λ is its wave-length, so

$$\lambda_0=\frac{hc}{\omega} \qquad . \quad . \quad . \quad . \quad (54)$$

gives the threshold wave-length.

* See A. Einstein, *Ann. der Phys.*, **17**, 132, 1905.

The threshold wave-length decreases with increase of the work function of the material. The lower the work function the greater the wave-length which can be employed to produce photo-emission.

It is remarkable that the only factor concerned in the velocity of the released photo-electrons is the wave-length of the incident radiation. The intensity of the radiation has no influence on these velocities. Moreover, the electrons are released for practical purposes as soon as the light is received by the photo-electric surface.* These experimentally ascertained results lent great support to the quantum theory, and could not be explained in accordance with the older classical mechanics. Thus light of increased intensity implies a greater number of photons, but not photons of greater energies. The photon energy can only be increased by using radiation of shorter wave-length.

Fortunately for the users of photocell measuring devices, the number of electrons released per second, hence the photo-electric current in a vacuum photocell, is directly proportional to the incident light intensity.

Another factor which influences the magnitude of the photo-electric effect is the ability of the surface to absorb light, and hence enable the light photon energies to be transferred to the loosely bound electrons within the surface atoms. In accordance with physical theory† light is absorbed because of the interaction with free electrons, so that the polished reflecting surface of a conductor, such as that of a metal, prevents the energy reaching the electrons. Again, a transparent insulating substance like glass exhibits low light absorption. The semi-conducting substances, particularly those with unpolished surfaces, are therefore the best photo-electric emitters.

If the photo-electric current produced per lumen, or per watt (see Appendix), by a scrupulously cleaned pure metal surface is plotted against the frequency of the incident radiation, it is noted that in some cases, e.g. the alkali-metals, a maximum effect is produced at a particular frequency (fig. 13). No satisfactory quantitative explanation of the shape of these spectral-distribution curves has yet been given.

* Time-lags between the instant of exposure of the surface to light and the release of photo-electrons are of the order of 10^{-9} sec.

† See Richtmyer and Kennard, *An Introduction to Modern Physics*, McGraw-Hill, 1942, and A. Sommer, *Photo-electric Cells*, Methuen, 1946.

A qualitative explanation of the existence of such maxima takes into account the following factors:

(*a*) As the frequency ν is increased, so the energy per photon $h\nu$ increases, and the electron velocities will rise in accordance with equation (53).

(*b*) The photo-current depends on the *quantum yield*, or the number of electrons released per photon. This ratio will depend

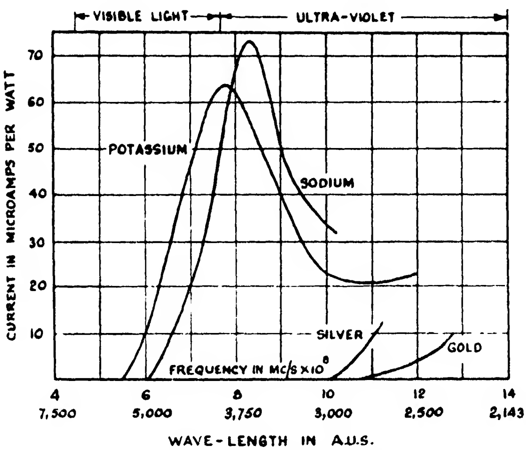


FIG. 13. Spectral Distribution of Photo-electric Effect for Some Pure Metals.

on the absorption of the radiation by the photo-electric surface, and the direction and polarisation of the incident photons. At least fourteen photons are needed per electron even in the most efficient types of surface. Now the number of photons per unit intensity of the incident beam will decrease with rise of frequency, since the greater the frequency the greater the energy per photon. Accordingly it would be expected that the photo-electric current would decrease with increasing frequency of the radiation, provided that the absorption and polarisation were constant.

Owing to their low atomic binding forces and consequently long threshold wave-lengths the alkali-earth metals, caesium, rubidium and the alkalis sodium and potassium play a large part in the manufacture of vacuum photocells.

The important role played by composite materials in producing an anomalous photo-electric effect, and the comparable effects of photo-conductivity, and the photo-voltaic effect are considered in Chapter 9.

Secondary Emission. If an electron impinges at adequate velocity on a positive conducting target (or is sufficiently accelerated towards a semi-conducting material) in a vacuum, its kinetic energy is capable of being transferred to the free electrons existing within the target material, so that they acquire sufficient velocity

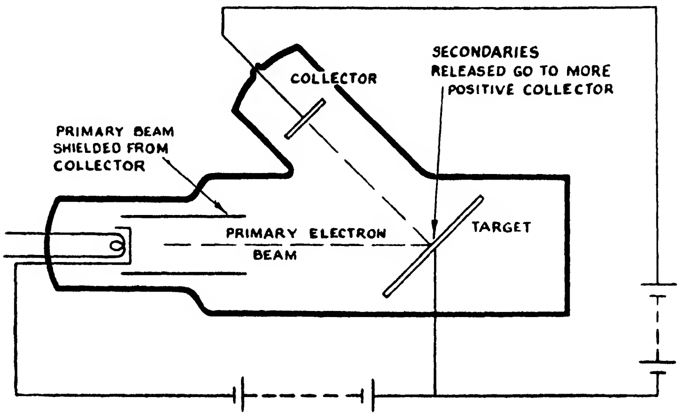


FIG. 14. Secondary Emission.

to overcome the work function effect of the material, and escape into the surrounding space.

Such electrons are called secondary electrons, and the ratio of number of secondaries per primary can be greater than unity. Indeed, making use of the low work function materials like caesium, a secondary emitting surface can be produced in which, at optimum primary electron velocities, this ratio may be as much as eight to ten, resulting in the development of the electron multiplier (p. 220), which is capable of enormous current magnifications.

Since the velocity of arrival of the primary electrons at a positive target depends only on the target potential in accordance with equation (28), so this potential is one condition in deciding the secondary emission effect. The other is the work function of the target material.

The target potential* must exceed a minimum value, which

* Note that the target potential is relative to the source of electrons. Accelerating potential is perhaps a better term. Thus the target may be earthed, and the emitter at a negative potential.

depends on the work function, to enable secondaries to be released at all, otherwise the arriving primaries will not acquire sufficient kinetic energy. This minimum potential is of the order of 15 to 30 V.

The secondaries leave the target with a maximum velocity of 25 V., and follow a Maxwell distribution law of velocities. They travel in random directions away from the target, and are confused with ordinary reflected primary electrons, especially if the primary beam makes a large angle of incidence with the target. Generally, highly polished metal surfaces yield more secondaries than rough surfaces. Contaminating films of foreign material usually increase the secondary emission ratio.

Zworykin, Morton and Malter* give the following table of secondary emission ratios for various materials at various positive potentials:

Voltage	Nickel	Molybdenum	Tantalum	Aluminium	Ag-Cs-CsO
100	0·7	0·9	0·5	1·5	5
200	1·0	1·1	1·3	2·0	7
400	1·3	1·2	1·3	2·3	8·3
800	1·2	1·0	1·4	2·2	7·8
1200	1·0	0·9	1·3	1·7	

It is noteworthy that there is an optimum target potential at which the secondary emission ratio is a maximum. This is at 300 to 400 V. An explanation is that as the primary electron velocities are increased so they penetrate more deeply into the target surface, and the secondaries are released at deeper levels within the conductor. They therefore undergo many more collisions with atoms in trying to escape from the conductor, so that many of them fail to be released. An optimum secondary ratio therefore depends on the distance within the target which the primary beam travels, giving an optimum primary velocity, and consequently a target potential at which the effect is a maximum. Most of the secondaries are emitted with small speeds of less than 3 electron-volts. Fewer than 10% of the total number of secondaries attain the maximum speeds of 25 electron-volts.

Cold Emission. This is also known as field emission, and auto-electric emission, and is obtained from metallic cathodes at room temperatures providing a powerful positive field is exerted at the cathode surface. Thus two metal plates separated by a few centimetres and sealed in a highly exhausted tube will

* See Zworykin, Morton and Malter, *Proc. Inst. Radio Eng.*, **24**, 355, 1936.

yield a small electron current if several thousand volts is applied between them. The potential field at the cathode surface must exert a sufficient attractive force to pull electrons through the metal potential barrier, where the free electrons inside the metal receive no additional aiding kinetic energy due to temperature increase. This effect is a bugbear to be avoided in some electron tubes, particularly photo-electric devices where the small, desired photo-emission may be confused by a cold emission effect if the operating anode potentials are high. Of recent years, however, cold cathode valves have been developed, such as the Ferranti K3 cold cathode triode, in which the cathode consists of a thin film of potassium on nickel; the ignitron using a mercury pool cathode; and the strobotron which has a cathode consisting of a nickel wire mesh coated with a caesium-aluminium mixture.*

An equation developed by Fowler and Nordheim gives the current density I in amp./sq. cm. of cathode surface as

$$I = 6{\cdot}2 \times 10^{-6} \sqrt{\left(\frac{W_m}{\omega}\right)} \cdot \frac{F^2}{W_m + \omega} \cdot e^{(-2{\cdot}1 \times 10^8 \omega^{\frac{3}{2}}/F)}. \qquad (55)$$

where ω is the work function of the metal in volts, F is the field at the cathode surface in V./cm., W_m is the maximum velocity attained by the electrons in the metal in electron-volts, being a constant at a given temperature.

Currents obtained rarely exceed 0·1 μA. in practice, but are increased if the metal surface is contaminated.

Ionisation by Collision. If electrons, produced by any of the means already discussed, pass to a positive electrode with sufficient velocity they will, under certain conditions, be capable of ionising residual gas molecules in the region they traverse. If L is the mean free path for an electron at the prevailing gas pressure, then the velocity it acquires between successive collisions with gas molecules in an electric field of strength X is given by Xet/m, where t is the time between collisions, which equals L/u, where u is the average electron velocity. The velocity acquired by the electron, or ion, between successive collisions must exceed a certain minimum, depending on the gas, before it is sufficient to disrupt the gas molecules to form further ions.

Since L is necessarily quoted as a mean value it follows, in

* See "Control Applications of Cold Cathode Valves," by L. Atkinson, Ch. VII, *Electronics*, edited by B. Lovell, Pilot Press, 1947.

practice, that ionisation will commence at values for a free path which is less than the mean value, and moreover, ionisation will not be complete until the value of the free path is much greater than the mean value.

It can be readily shown that the number of fresh ions formed per cm. by a given ion is obtained from the formula*

$$\text{No. of ions} = C \,.\, p \,.\, \varepsilon^{CTp/Xe}, \quad . \quad . \quad (56)$$

where p is the pressure, T is the minimum energy which must be exceeded before ionisation commences and C is a constant depending on the nature of the gas.

The ionisation potential of a gas is the potential difference through which an electron must fall to acquire sufficient energy to ionise the molecule.

The following are values for the commoner gases:

Gas	Ionisation Potential V.
Argon	15·4
Helium	24·6
Hydrogen . . .	16·1
Mercury	10·4
Neon	21·5
Nitrogen	16·0
Oxygen	16·0

When electrons pass through a gas, the possibility of the formation of negatively charged atoms, molecules or clusters of molecules depends on the ratio of the electric field strength to the gas pressure, i.e. X/p. If p is small, then the mean free path is great, hence with a large value of X, the velocity of an electron on hitting a gas molecule is high, so that elastic collision occurs, the electron not being captured. If, however, X is small, and p large, then, by similar reasoning, it follows that the electron impinges on a gas molecule with low velocity, and is therefore likely to be captured by the molecule to form a negative ion. The criterion given by Townsend is that if $X/p < 0{\cdot}1$, then negatively ionised atoms, molecules and clusters of as many as eight molecules are formed, whereas if X/p is $> 0{\cdot}1$, then the only negative ions present are electrons. X is measured in es. units and p in dynes per sq. cm.

* See J. A. Crowther, *Ions, Electrons and Ionizing Radiations*, Arnold, 1938.

CHAPTER 3

The Motion of the Free Electron in Electrostatic and Magnetic Fields

Uniform Electrostatic Field; Electrons Initially at Rest. An electric field is maintained between a pair of parallel plates, and electrons are released, by one of the methods discussed on

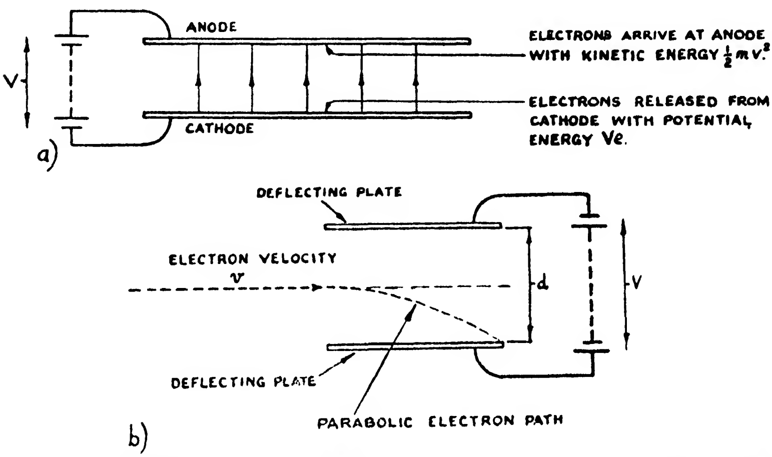


FIG. 15. *a*, Electrons moving along Lines of Electric Force. *b*, Electrons moving transversely to Lines of Electric Force.

p. 17, with insignificant initial velocity from the negative plate, or cathode.

By definition (p. 1) an electric line of force is the direction in which a free positive point charge placed in the field will move. Electrons, being negatively charged, will move in the opposite direction, from negative to positive plate, assuming that the presence of a number of electrons does not introduce a space-charge (p. 81) which would alter the configuration of the existing field.

From the definition of potential difference (p. 6), the work done on the electron in moving from the negative to positive plate (fig. 15*a*) will be Ve, where V is the P.D. concerned, and e the electron charge. This work will be transferred into kinetic energy gained by the electron, equal to $\frac{1}{2}mv^2$, where v is the velocity of the electron of mass m on reaching the positive plate.

$$\therefore \quad Ve=\tfrac{1}{2}mv^2. \qquad (57)$$

$$\therefore \quad v=\sqrt{\frac{2Ve}{m}}, \qquad (58)$$

giving the electron velocity in cm./sec. provided that V, e and m are quoted in absolute units.

Putting $e=4{\cdot}8\times10^{-10}$ e.s.u., $m=9{\cdot}1\times10^{-28}$ g.

$$\therefore \quad v=\sqrt{\left(\frac{2\times4{\cdot}8\times10^{-10}}{9{\cdot}1\times10^{-28}}\right)}\times\sqrt{V}=\sqrt{\left(\frac{9{\cdot}6}{9{\cdot}1}\times10^{18}\right)}\sqrt{V}$$

$$=1{\cdot}027\times10^{9}\sqrt{V}\text{ cm./sec.}$$

If V is put in practical units, volts, instead of electrostatic units, then $V_{\text{e.s.u.}}=300V_{\text{volts}}$ (p. 16).

$$\therefore \quad v=\sqrt{\left(\frac{V}{300}\right)}\times1{\cdot}027\times10^{9}=5{\cdot}93\times10^{7}\sqrt{V}\text{ cm./sec.},$$

where V is in volts.

For example, if the accelerating P.D. $=100$ V., then

$$v=5{\cdot}93\times10^{7}\sqrt{100}=5{\cdot}93\times10^{8}\text{ cm./sec.}$$

If the electron moves with a speed which is significant compared with c, the velocity of light ($=3\times10^{10}$ cm./sec.), then from the notion of the electromagnetic origin of the electron mass, combined with the theory of relativity, the mass increases in accordance with the formula

$$m=m_0[1-(v/c)^2]^{-\frac{1}{2}}, \qquad (59)$$

where m_0=rest mass of electron, and m=mass of the electron when it acquires the velocity v. Moreover, its kinetic energy will be $(m-m_0)c^2$, instead of $\frac{1}{2}mv^2$.

This correction is unimportant ($<1\%$) up to accelerating P.D.'s of 7000 V. At higher potentials equation (57) should be modified to give

$$Ve=(m-m_0)c^2=m_0c^2\left[\frac{1}{\sqrt{(1-v^2/c^2)}}-1\right]. \qquad (60)$$

In electronics, an increasing number of specialised pieces of apparatus are being developed requiring high operating potentials, and demanding this relativistic correction in their design formulae. Such are the electron microscrope, the high-voltage cathode-ray tube, and the cyclotron.

The *electron-volt* is frequently adopted as a unit of energy, or velocity in the case of moving electrons. Though a misnomer, yet it is a useful term, since an electron of energy n electron-volts, or such a velocity, is one which has acquired this energy or velocity by undergoing a potential fall of n volts, in accordance with equation (57).

Electron Path in a Transverse Electrostatic Field. If an electron enters the space between a pair of parallel plates, distance d apart, with a P.D. between them of V, then the electric field operating will be V/d, the potential gradient. The force on the electron will then be Ve/d. The corresponding acceleration will be $\frac{\text{force on electron}}{\text{electron mass}}=Ve/md=a$, say. In a time t after the entry of the electron into the field, it will therefore undergo a displacement $\frac{1}{2}at^2$ towards the positive plate (fig. 15*b*).

Suppose the forward velocity of the electron be uniform, and through the field centre, parallel to the plates. This velocity will be unaffected by the transverse deflecting force which acts. The problem is then dynamically equivalent to that of a small body, thrown horizontally forward, and subject to the uniform acceleration due to gravity: its path will be a parabola.

If this horizontal velocity is v, then in time t the horizontal displacement x executed is vt. In this same time t, the vertical displacement $y=\frac{1}{2}(Ve/md).t^2$. Combining these two equations to eliminate t gives

$$y=\frac{1}{2}\cdot\frac{Ve}{md}\cdot\left(\frac{x}{v}\right)^2=\frac{Ve}{2mdv^2}\cdot x^2. \qquad (61)$$

Therefore $y=$const. x^2, which is the equation for a parabola.

Electron Path in a Transverse Magnetic Field. An electron at constant velocity v is supposed to enter a uniform magnetic field H in vacuum. Let the initial electron path be at right angles to the magnetic lines of force (fig. 16*a*).

Since the initial electron velocity is v, its motion will correspond to a rate of displacement of charge, or electric current, ev. In accordance with the motor principle (p. 14), this electron will

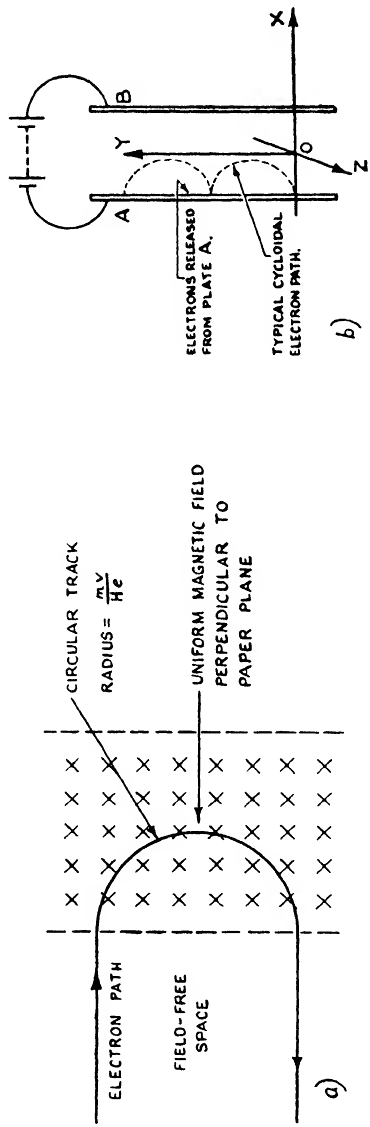


FIG. 16. *a*, Electron Trajectory in a Uniform Magnetic Field. *b*, Electron introduced into Joint Electric and Magnetic Fields.

experience a force acting at right angles to both its initial direction and to the direction of the lines of force, given by

$$F=Hev. \qquad (62)$$

Since the electron experiences a force always at right angles to its direction of motion at any moment, it is dynamically analogous with the case of a small mass being whirled round in a circle whilst tied to a thread along the radius. The deflecting force must be equal to the centripetal force, so

$$Hev=\frac{mv^2}{r}, \qquad (63)$$

$$\therefore \quad r=\frac{mv}{He}. \qquad (64)$$

The electron will therefore move in a circular path of radius given by (64), or along the arc of such a circle if it leaves the field before the circular path is completed.

A helical path will be followed by the electron if, in addition to the above forces, it experiences a forward motion along the magnetic lines of force due to an electric field component in that direction.

Electrons entering Joint Electrostatic and Magnetic Fields. If electrons of velocity v entering a magnetic field have, in addition, an electric field superimposed of which the lines of force are at right angles to both the magnetic lines, and the initial direction of the electrons, then the deflecting forces due to both electric and magnetic actions will be in the same plane, but may either aid or oppose one another depending on the polarity of the electric field. If they are in opposition then the fields may be adjusted so that the forces cancel one another, and no deflection of the beam occurs. Then

$$Xe=Hev, \quad \text{or} \quad v=\frac{X}{H}. \qquad (65)$$

Eliminating velocity v between (64) and (65), gives

$$\frac{e}{m}=\frac{X}{rH^2}. \qquad (66)$$

By measurement of X, H and r, the radius of the electron path in the magnetic field alone, J. J. Thomson found the value of e/m, the ratio of electron charge to mass.

A second interesting case occurs if the electrons do not traverse the crossed fields in a beam, but instead are released from the negative electric field plate with zero velocity (cf. the magnetron, p. 299). The electron trajectory can then be shown to be a cycloid (fig. 16*b*).

Consider the motion of the electron relative to three axes OX, OY and OZ at right angles to one another. Let the axis OZ be parallel to the magnetic lines of force, and the axis OX be parallel to the electric lines of force. At any instant the electron will have a velocity dx/dt parallel to OX due to the electric field, which will be accompanied by a mechanical force due to the magnetic field of $He(dx/dt)$ acting at right angles to both the magnetic field and to the electric field, i.e. in direction of OY.

$$\therefore \quad m\frac{d^2y}{dt^2}=He\frac{dx}{dt}, \qquad (67)$$

since d^2y/dt^2 will be the acceleration given to the electron in the direction of OY.

In the direction of OX there will act an electric force equal to Xe, and a mechanical force due to the magnetic field equal to $He(dy/dt)$, where dy/dt is the component of the electron velocity along OY.

$$\therefore \quad m\frac{d^2x}{dt^2}=Xe-He\frac{dy}{dt}. \qquad (68)$$

The solution of equations (67) and (68) is

$$\left.\begin{aligned} x&=\frac{X}{wH}(1-\cos wt),\\ y&=\frac{X}{wH}(wt-\sin wt), \end{aligned}\right\} \qquad (69)$$

where $w=He/m$.

These equations are those of a cycloid, the curve traced out by a point on the circumference of a circle which is rolling along a straight line; the path of the electron therefore being in the form of a series of loops, as shown in fig. 16*b*.

CHAPTER 4

Fundamental Alternating Current Theory

Simple Theory. The usual electrical data regarding alternating currents can be summarised as follows:

(*a*) An alternating current or voltage of single frequency is represented graphically by a sine wave, and has the mathematical

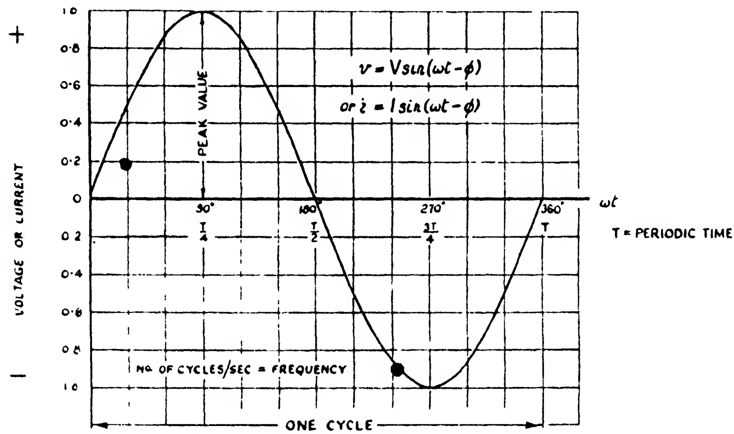


FIG. 17. Alternating Current or Voltage.

formula $i=I \sin \omega t$,* or $v=V \sin \omega t$, where i is the instantaneous value of the current at time t, I is the peak current in the cycle and ω, called the *pulsatance*, is 2π times the frequency, f. V and v are similarly the terms used in the case of an alternating voltage (fig. 17).

(*b*) The equivalent direct current, as regards heating effect, to an alternating current of peak value I, has a magnitude of $I/\sqrt{2}$, called the root mean square value (R.M.S.).

(*c*) When applied across a pure resistance an alternating E.M.F. produces an alternating current of the same wave-form and phase, where Ohm's law is obeyed, giving

$$i=\frac{v}{R}=\frac{V \sin \omega t}{R}.$$

* Here assumed that $i=0$ when $t=0$, otherwise $i=I \sin(\omega t-\phi)$, where ϕ is a phase-angle, and $i=0$ when $t=\phi/\omega$.

(*d*) A source of A.C. across a pure inductance produces an alternating current through the inductance which lags in phase by 90° on the applied voltage.

$$\text{R.M.S. current} = \frac{\text{R.M.S. applied voltage}}{\text{Reactance of inductance}},$$

where inductive reactance $X_L = \omega L$, L being in henrys, and X_L is given in ohms.

(*e*) The alternating current through a pure capacitance leads in phase by 90° on the applied voltage, and has a value given by $\text{R.M.S. current} = \frac{\text{R.M.S. applied voltage}}{\text{Reactance of capacity}}$, where capacitative reactance $X_C = 1/\omega C$, C being in farads, and X_C in ohms (fig. 18).

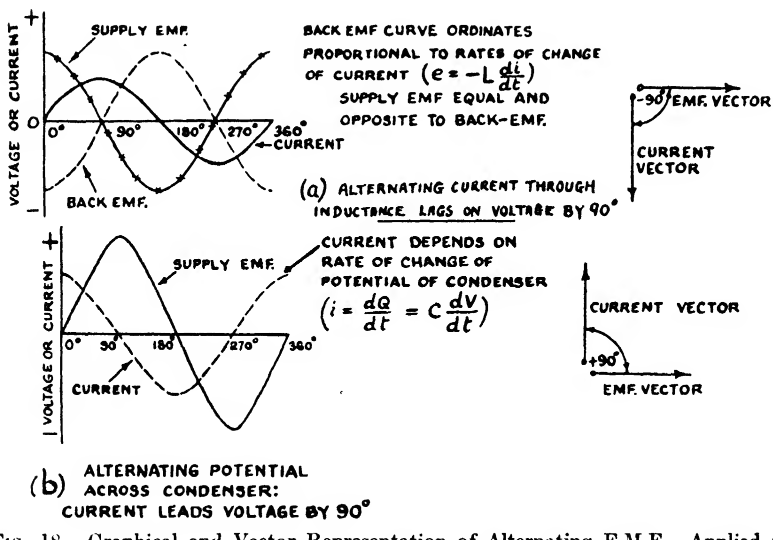


FIG. 18. Graphical and Vector Representation of Alternating E.M.F. Applied to (*a*) Inductance, (*b*) Capacitance.

"j" Notation. In A.C. theory it is much more convenient to represent phase differences of 90° by use of the operator "j", where $j = \sqrt{-1}$. If this is regarded as being a mathematical convention, then a simple meaning can be attached to j by considering fig. 19. In this diagram the quantity $+1$ is represented, as in normal graphical work, as being drawn one unit of scale to the right of the origin along the X-axis, whereas -1 is a similar distance marked off to the left of the origin. The convention adopted is that $+1$ is changed to -1 by rotation of the radius vector OA through 180°. After rotation through only 90°, this radius vector would be along the Y-axis, at OA'. Let a direction along the Y-axis be represented by j, or $\sqrt{-1}$, or, rather, suppose rotation of the radius vector through 90° is equivalent to

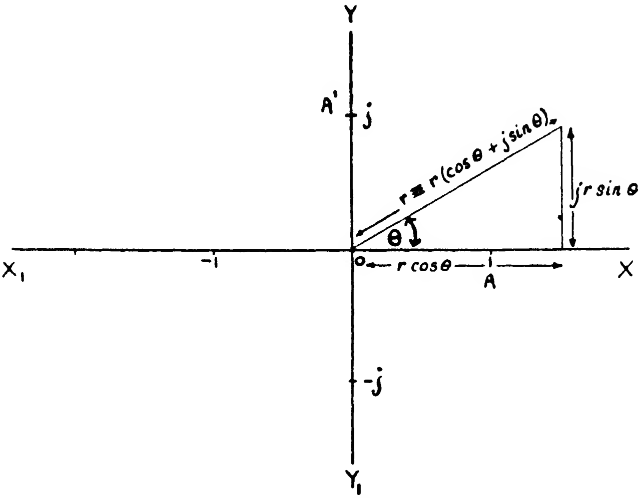


FIG. 19. Conventional Representation of Real and Imaginary Numbers, Use of Operator j.

multiplying by j. Rotation through 180° will then be equivalent to multiplication by j again, i.e. by j^2 altogether. But $j^2 = \sqrt{-1}^2 = -1$, which is in accordance with the original specification as to the direction given to -1.

Again, any radius vector at any angle θ to the X-axis, of length r will subtend lengths of $r\cos\theta$ along the X-axis, and $r\sin\theta$ along the Y-axis. Hence r is equivalent to $r(\cos\theta + j\sin\theta)$, since $r\cos\theta$ is in the direction of the real numbers, in the X-direction, whilst $r\sin\theta$ is in the direction of the imaginary numbers, along the Y-axis, and is therefore represented by multiplying by j.

In accordance with this notation, and by reference to fig. 18, it can be seen that the current through an inductance is best considered from the formula $i = v/j\omega L$, whereas through a condenser the current is given by $i = \dfrac{v}{-j/\omega C} = j\omega C$. Note here that

$-j$ is in the opposite direction to $+j$, and since $j^2=-1$ so $-j^2=+1$, and hence $-j=1/j$.

Circuit Consisting of Inductance and Resistance in Series. If V is R.M.S. value of supply E.M.F., and I the corresponding current, then

$$V=RI+j\omega L \,.\, I \quad . \quad . \quad . \quad . \quad (70)$$

on equating supply voltage to total voltage across circuit components (fig. 20a).

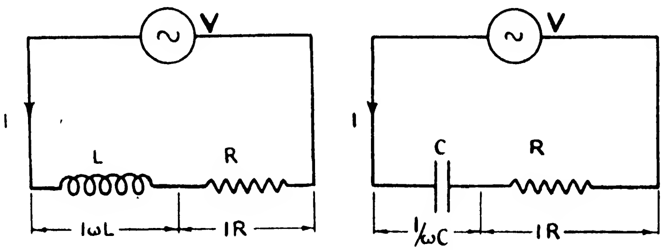


FIG. 20. a, Inductance and Resistance in Series. b, Condenser and Resistance in Series.

But $Z=V/I$ is the impedance of the circuit, defined practically as $\dfrac{\text{R.M.S. value of applied voltage}}{\text{R.M.S. current through the circuit}}$.

$$\therefore \quad Z=R+j\omega L.$$

Since it is known that j signifies that the reactive inductance opposition to current flow is 90° different in phase from the resistive opposition, so Pythagoras' theorem applies in obtaining the impedance:

$$Z=R+j\omega L=\sqrt{(R^2+\omega^2L^2)}. \quad . \quad . \quad . \quad (71)$$

Condenser and Resistance in Series. In this case (fig. 20b)

$$V=RI-j\,.\,I/\omega C. \quad . \quad . \quad . \quad . \quad (72)$$

$$\therefore \quad Z=V/I=\sqrt{(R^2+1/\omega^2C^2)}. \quad . \quad . \quad . \quad (73)$$

Power Factor. All condensers dissipate some energy when alternating current effectively flows through them. This is due to imperfect insulation of the dielectric, to dielectric hysteresis and to heating effects in the condenser plates and leads. This energy loss is represented by an equivalent series resistance R, where the power loss per cycle is I^2R, I being the R.M.S. current flowing. The ideal condenser, reactive only in its nature, does not introduce any power consumption in the circuit in which it is situated, since the voltage and current are exactly 90° out of phase. The condensers encountered in practice will, however, inevitably suffer some power loss because of the fore-mentioned imperfections. The power factor of a condenser is defined as the

$$\frac{\text{real power}}{\text{apparent power}} = \frac{\text{power dissipated in equivalent resistance}}{\text{apparent power supplied to the condenser}}.$$

This ratio $= \frac{I^2R}{I^2Z} = \frac{R}{Z} = \cos\phi$, where ϕ is actual phase angle between supplied volts and condenser current. If $\phi = 90°$, as in the ideal condenser, then $\cos\phi = 0$, and $R = 0$, so the real power supplied is zero.

Since the power factor, $\cos\phi = \frac{R}{Z} = \frac{R}{\sqrt{(R^2 + 1/\omega^2C^2)}}$, and R is usually much smaller than $1/\omega C$, so

$$\cos\phi = \frac{R}{\sqrt{(1/\omega^2C^2)}} = \omega CR \text{ approx.} \quad . \quad . \quad (74)$$

Likewise, an inductance inevitably has actual resistance, and has a power factor

$$\cos\phi = \frac{R}{Z} = \frac{R}{\sqrt{(R^2 + \omega^2L^2)}} = \frac{R}{\omega L} \text{ approx.} \quad . \quad (75)$$

A.C. Applied to Condenser, Inductance and Resistance in Series. In this case (fig. 21*a*)

$$V = j\omega L \, . \, I - \frac{j}{\omega C} \, . \, I + RI.$$

$$\therefore \quad V = j\left(\omega L - \frac{1}{\omega C}\right)I + RI.$$

$$\therefore \quad Z = \frac{V}{I} = j\left(\omega L - \frac{1}{\omega C}\right) + R$$

$$= \sqrt{\left\{\left(\omega L - \frac{1}{\omega C}\right)^2 + R^2\right\}}. \quad . \quad . \quad (76)$$

Examination of this impedance formula shows that Z has a minimum value of R when $\omega L = 1/\omega C$. The impedance of such a series circuit is therefore a minimum at a critical *resonant* frequency, given by $\omega L = 1/\omega C$.

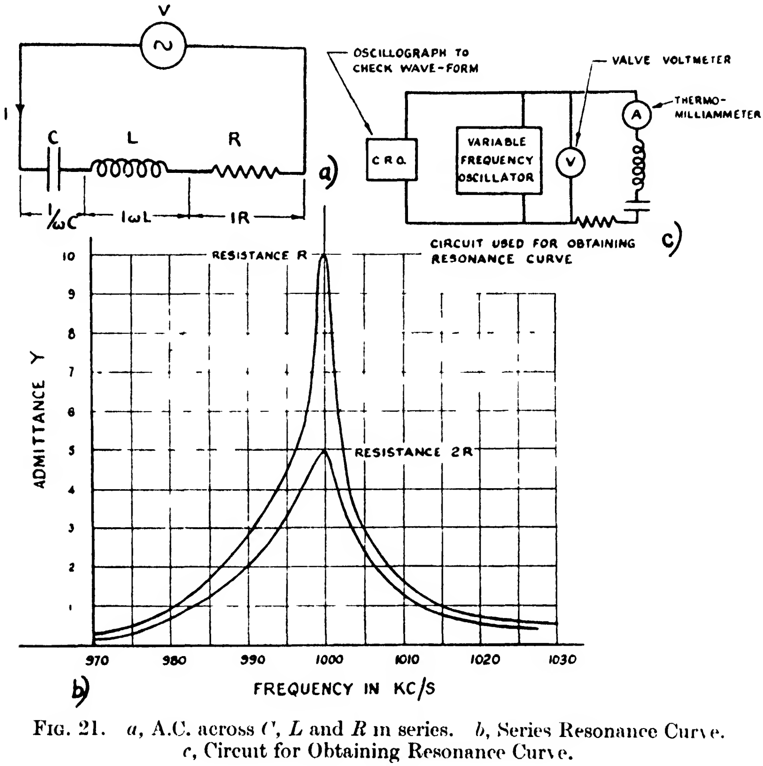


FIG. 21. *a*, A.C. across C, L and R in series. *b*, Series Resonance Curve. *c*, Circuit for Obtaining Resonance Curve.

$$\therefore \quad \omega^2 = \frac{1}{LC}, \qquad \omega = \frac{1}{\sqrt{LC}},$$

so the resonant frequency

$$f = \frac{1}{2\pi\sqrt{(LC)}} \quad . \quad . \quad . \quad . \quad . \quad (77)$$

The Acceptor Circuit. This last type of circuit is much used in electronic practice to give a maximum response at a particular frequency. Its commonest use is in the radio receiver, where it can act as a simple tuned circuit in that the setting of the value of the condenser C, which is made variable, can in conjunction with a particular value of inductance L enable the circuit to give a maximum response to one only of the radio frequency currents present in the receiver aerial. Not only does this circuit exhibit this necessary selective action, it also gives a voltage magnification effect at resonance.

Consider fig. 21*a*; at resonance the supply voltage V has simply to overcome the voltage across the resistance, IR. The inductance voltage V_L is equal to, but in antiphase with, the condenser voltage V_C, so that at resonance V_C and V_L cancel one another. Nevertheless, the voltage across the inductance, or across the capacity, can be tapped off and fed to a valve amplifier circuit, and this voltage will be $I \, . \, \omega L$. Hence,

$$\frac{\text{the output voltage across coil or condenser}}{\text{supply voltage}}=\frac{\omega LI}{V}$$

$$=\frac{\omega LI}{RI}=\frac{\omega L}{R}. \qquad (78)$$

Since $\omega L/R$ is a function of the coil used, it is referred to as the coil "Q" factor.

By a suitable choice of L, and the use of a coil of as small a resistance as possible, $\omega L/R$ can be made much greater than 1. Values up to 200 are obtainable at radio frequency. For medium wave-band working (1500 kc./s. to 500 kc./s.), $L=175$ μH., and C is variable from 0·00005 to 0·0005 μF.

Resonance Curves for Series Acceptor Circuit. If a source of A.C. of which the frequency can be continuously varied over a suitable range is applied to a series LCR* circuit, and the applied voltage and series current are measured, then a curve of frequency vs. admittance $I/V=Y$ has the form illustrated in fig. 21*b*.

These curves indicate a maximum admittance, and hence maximum current for a given input voltage, at resonance. This current decreases with an increase of the circuit resistance. The extent to which the current falls off as the frequency departs from the resonant value is a measure of the *selectivity* of the circuit.

The ratio of the current I_ω at a pulsatance ω to the current I_{ω_0} at resonance is

$$\frac{I_\omega}{I_{\omega_0}}=\frac{R_0}{R+j(\omega L-1/\omega C)}. \qquad (79)$$

* The abbreviation LCR is to be used to denote a circuit containing inductance L, capacity C and resistance R.

Note that R_0, the resistance of the circuit at the resonance frequency, is not the same as R, the resistance of the circuit at another frequency. This is due to the skin effect, whereby at high frequency an alternating current becomes primarily confined to the skin of the conductor, so that the resistance increases with frequency (see p. 72).

From (79) $$\frac{I_\omega}{I_{\omega_0}}=\frac{R_0}{R+j\,\{(\omega^2LC-1)/\omega C\}}.$$

Put $\delta=\dfrac{\omega-\omega_0}{\omega_0}=\dfrac{\text{actual frequency}-\text{resonant frequency}}{\text{resonant frequency}}.$

$$\therefore\quad \frac{I_\omega}{I_{\omega_0}}=\frac{R_0}{R+j\,\{[(1+\delta)^2-1]/(1+\delta)\}\,\omega_0L} \quad . \quad . \quad (80)$$

on putting $\omega=\omega_0\,(1+\delta)$, and $\omega_0L=1/\omega_0C$.

But $$\frac{R}{R_0}=\frac{\omega}{\omega_0}=(1+\delta),\ \text{approx.}$$

$$\therefore\quad \frac{I_\omega}{I_{\omega_0}}=\frac{1}{1+\delta+jQ\delta[(2+\delta)/(1+\delta)]}, \quad . \quad . \quad (81)$$

where $Q=\omega_0L/R_0$=circuit Q at resonant frequency, δ is the fraction of the resonance frequency by which the actual frequency departs from resonance. It is to be especially noted that selectivity, as does the voltage magnification, depends on Q; the higher the value of Q the more sharply selective is the circuit.

Since $Q=\omega_0L/R_0$, put $\omega_0=1/\sqrt{(LC)}$.

$$\therefore\quad Q=\frac{1}{R_0}\sqrt{\frac{L}{C}}. \quad . \quad . \quad . \quad . \quad . \quad (82)$$

To obtain a series circuit of high voltage magnification and selectivity, the circuit needs a low value of resistance and a high ratio of L/C. In practice the ratio L/C cannot be excessively large because a high value of L necessitates self-capacitance of the coil, increasing the effective value of C.

A.C. Applied to Condenser and Inductance in Parallel. In radio-frequency practice the power factor of a condenser is usually so small as to be negligible, hence the effective series resistance of the condenser is neglected. On the other hand the coil resistance is an important factor. So the circuit is best considered as an A.C. source across two circuit branches in parallel, one branch being capacitative only, whilst the other branch consists of an inductance and resistance in series (fig. 22*a*).

If I=R.M.S. current from supply source, I_L=current through inductance branch, I_C=current through capacitance branch, then

$$I=I_L+I_C. \quad \quad (83)$$

But $I=V/Z$, where V=applied E.M.F., and Z=the total impedance of the circuit; also $I_L=\frac{V}{j\omega L+R}$ and $I_C=\frac{V}{1/j\omega C}$.

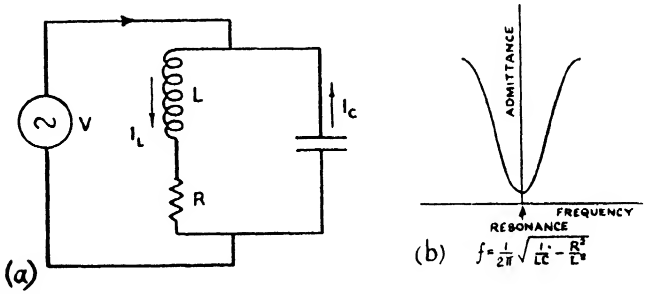


FIG. 22. *a*, A.C. across Condenser and Inductance in Parallel. *b*, Parallel Resonance Curve.

Substitution of these values for I, I_L and I_C in equation (83) gives

$$\frac{V}{Z}=\frac{V}{j\omega L+R}+\frac{V}{1/j\omega C}.$$

$$\therefore \quad \frac{1}{Z}=\frac{1}{j\omega L+R}+j\omega C=\frac{1+j\omega C(R+j\omega L)}{R+j\omega L}.$$

Since the appearance of j in the denominator is inconvenient, this expression is *rationalised* by multiplying both numerator and denominator by $R-j\omega L$. Such rationalisation is always performed on encountering such expressions, thus any term of the form $\frac{1}{A+jB}=\frac{A-jB}{(A+jB)(A-jB)}=\frac{A-jB}{A^2+B^2}$ in which the denominator is real.

$$\therefore \quad \frac{1}{Z}=\frac{[1+j\omega C(R+j\omega L)](R-j\omega L)}{R^2+\omega^2L^2}$$

$$=\frac{(R-j\omega L)(1+j\omega CR-\omega^2LC)}{R^2+\omega^2L^2}$$

$$=\frac{R+j\omega CR^2-\omega^2LCR-j\omega L+\omega^2LCR+j\omega^3L^2C}{R^2+\omega^2L^2}$$

$$=\frac{R+j(\omega CR^2-\omega L+\omega^3L^2C)}{R^2+\omega^2L^2} \quad . \quad . \quad . \quad . \quad (84)$$

This expression consists of two parts, one real, or in phase, i.e. resistive in its effect, and equal to $\frac{R}{R^2+\omega^2L^2}$; the other imaginary, or 90° out of phase, i.e. reactive in its effect, and equal to

$$\frac{\omega CR^2-\omega L+\omega^3L^2C}{R^2+\omega^2L^2}.$$

The definition of resonant frequency of a circuit is that it is the frequency at which reactive opposition to current flow vanishes. At resonance a circuit has an impedance which is purely resistive. Hence to find the resonant frequency of this parallel circuit, put $\omega CR^2-\omega L+\omega^3L^2C=0$,

$$\therefore \quad CR^2-L+\omega^2L^2C=0.$$

$$\therefore \quad \omega^2=\frac{L-CR^2}{L^2C}=\frac{1}{LC}-\frac{R^2}{L^2}.$$

$$\therefore \quad f=\frac{\omega}{2\pi}=\frac{1}{2\pi}\sqrt{\left(\frac{1}{LC}-\frac{R^2}{L^2}\right)} \quad . \quad . \quad . \quad (85$$

gives the resonant frequency.

$f=\frac{1}{2\pi\sqrt{(LC)}}$ if R is neglected compared with $\frac{1}{LC}$.

The Parallel Rejector Circuit. If this circuit is at resonance, then its impedance will be resistive only, and is therefore the inphase component of the equation (84).

Therefore at resonance, $\frac{1}{Z}=\frac{R}{R^2+\omega^2L^2}$.

$$\therefore \quad Z=\frac{R^2+\omega^2L^2}{R}.$$

But $\omega^2=\frac{1}{LC}-\frac{R^2}{L^2}$ from (85).

Therefore at resonance, $Z = R+\frac{L^2}{R}\left(\frac{1}{LC}-\frac{R^2}{L^2}\right)$

$$=R+\frac{L}{CR}-R=\frac{L}{CR}. \quad . \quad . \quad (86)$$

This circuit is called a *rejector* circuit in radio-frequency practice. Its impedance at resonance is L/CR, called the dynamic resistance. This impedance is greater than the impedance of the circuit at any other frequency. At resonance the circuit impedance is thus a maximum, and is resistive in its effect in that the voltage and current are in phase. Such a rejector circuit is commonly used as an anode load in a radio-frequency valve amplifier (see p. 141).

Whereas the acceptor circuit exhibits a voltage magnification effect at resonance, the rejector circuit gives a current magnification. The current through the condenser will be $V\omega C$ at resonance, whereas the supply current, called the "make-up" current, will be $\frac{V}{L/CR}$. The ratio of these currents is $\frac{\omega C}{L/CR}=\frac{\omega^2C^2R}{L}$, which is equal to $\omega L/R$ if R is neglected compared with LC in the resonance frequency formula.

Resonance Curve for Parallel Rejector Circuit. As in the case of the series LCR circuit, a variable source of A.C. is applied across a parallel circuit, the admittance $\left(\frac{\text{make-up current}}{\text{applied volts}}\right)$ being plotted against frequency. In this case the circuit exhibits a minimum admittance at resonance, i.e. the current is a minimum (fig. 22*b*).

Charge of a Condenser through a Resistance. Consider a steady source of E.M.F. connected across a condenser and resistance in series (fig. 23*a*).

At some instant of time t let the current be i and condenser charge $=q$.

Then $V=iR+q/C$.

Put $i=\frac{dq}{dt}$, therefore $V=R\frac{dq}{dt}+\frac{q}{C}$.

Rearranging, $$V-\frac{q}{C}=R\frac{dq}{dt}.$$

$$\therefore \quad \frac{dq}{V-q/C}=\frac{dt}{R}. \quad . \quad . \quad . \quad . \quad (87)$$

But $\dfrac{d}{dq}\left(V-\dfrac{q}{C}\right)=-\dfrac{1}{C}$, therefore $d\left(V-\dfrac{q}{C}\right)=\dfrac{-dq}{C}$.

Hence (87) becomes $-C\left(\dfrac{d(V-q/C)}{V-q/C}\right)=\dfrac{dt}{R}$.

Integrating, $\log_e(V-q/C)=-t/CR+a$, where a is constant.

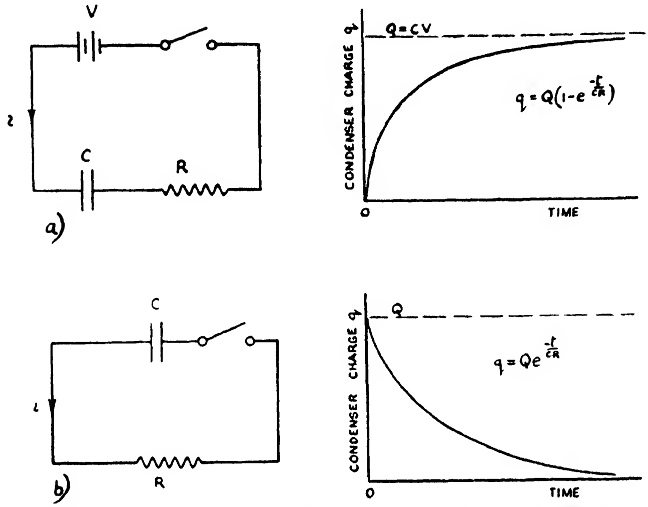


FIG. 23. *a*, Charge of a Condenser through a Resistance. *b*, Discharge of a Condenser through a Resistance.

When $t=0$, $q=0$, $a=\log_e V$.

$$\therefore \quad \log_e\left(V-\frac{q}{C}\right)=\frac{-t}{CR}+\log_e V.$$

$$\therefore \quad \log_e\left(\frac{V-q/C}{V}\right)=\frac{-t}{CR}.$$

$$\therefore \quad \frac{V-q/C}{V}=e^{-t/CR}.$$

$$\therefore \quad V-\frac{q}{C}=Ve^{-t/CR}.$$

$$\therefore \quad q=CV(1-e^{-t/CR})=Q(1-e^{-t/CR}), \qquad (88)$$

where Q is the final condenser charge.

Alternatively, this equation can be written $v=V(1-e^{-t/CR})$, where v is the potential across the condenser at the time instant t.

On plotting a graph of v, or q against t, an exponential curve of the form shown in fig. 23*a* is obtained.

Discharge of a Condenser through a Resistance. Consider a condenser C which is initially charged to amount Q, where the P.D. across its plates is V. This charged condenser is then connected across a resistance R so as to discharge (fig. 23*b*).

Then $0=iR+\frac{q}{C}$.

$$\therefore \quad 0=\frac{dq}{dt}R+\frac{q}{C}.$$

Rearranging, $$\frac{dq}{q}=\frac{-dt}{CR}.$$

Integrating, $\log_e q=\frac{-t}{CR}+b$, where b is constant.

When $t=0$, $q=Q$, hence $b=\log Q$.

$$\therefore \quad \log_e q=\frac{-t}{CR}+\log_e Q.$$

$$\therefore \quad \frac{q}{Q}=e^{-t/CR}, \quad \text{hence} \quad q=Qe^{-t/CR}$$

$$\text{or} \quad v=Ve^{-t/CR}. \qquad (89)$$

A graph of v, or q plotted against t gives the decay of charge of a condenser when shunted by a resistance, fig. 23*b*.

Time-Constant. In electronic practice, the rate at which the charging or discharging of a condenser through a resistance occurs is of great importance in many circuit devices, since widespread use is made of CR combinations to obtain electric actions of definite durations. A study of equations (88) and (89) in conjunction with the curves of fig. 23, shows that the time taken for a complete charging or discharging action of a condenser to occur is infinity, but with small values of C and R the time for the greater part of the charge to be lost, or gained, will be of short duration.

The time-constant, in either case, is given by the simple equation

$$T=CR, \qquad (90)$$

where T is the time in seconds, C is the condenser capacity in farads, and R the resistance in ohms.

In the charging case, put $t=CR$ in equation (88), then $q=Q(1-e^{-1})=Q(1-1/2\cdot72)=0\cdot63Q$. So the time-constant is the time taken for the condenser to acquire 0·63 of its total possible charge, or for its potential to be 0·63 of the total supply voltage.

In the discharge case, put $T=CR$ in equation (89).

$$\therefore \quad q=Qe^{-1}=Q\frac{1}{2\cdot72}=Q(1-0\cdot63).$$

Now the time-constant is the time taken for the condenser to lose 0·63 of its initial charge, or potential.

The Freely Oscillating Circuit. In communication practice the basic circuit for the production of radio-frequency alternating current consists of an initially charged condenser across an inductance, where the energy in the condenser charge alternates continually with the energy in the magnetic field associated with the inductance, the result being the production of an oscillatory current in the circuit of which the amplitude dies away more or less rapidly depending on the circuit resistance and the rate at which energy is dissipated in the form of electromagnetic radiation. The function of the valve in the valve oscillator is to restore this energy loss per cycle from the H.T. supply so as to maintain continuous oscillations (see p. 174). To enable important conceptions in electronic practice to be understood, a full study of this circuit is imperative.

A series circuit consists of a condenser C, inductance L and resistance R. The condenser is initially charged by connecting it momentarily to a battery of E.M.F. E on tapping the key K (fig. 24*a*). The subsequent action in the circuit is with the battery disconnected.

Let i be the current in the circuit at any instant of time t, and correspondingly q the condenser charge.

Since the circuit is closed, the sum of the potentials round the circuit is zero,

$$\therefore \quad -L\frac{di}{dt}-\frac{q}{C}=iR. \quad . \quad . \quad . \quad . \quad . \quad (91)$$

Since $L(di/dt)$ is the back-E.M.F. across the inductance L, due to a rate of change of current di/dt through it, q/C is the P.D. across the condenser, acting in the same direction as $L(di/dt)$, and the sum of these equals the P.D. across the resistance.

But $i=dq/dt$, since current is rate of change of charge,

$$\therefore \quad L\frac{d^2q}{dt^2}+R\frac{dq}{dt}+\frac{q}{C}=0. \qquad . \quad . \quad . \quad (92)$$

Dividing through by L, and putting $R/L=2b$, and $1/LC=k^2$,

$$\therefore \quad \frac{d^2q}{dt^2}+2b\frac{dq}{dt}+k^2q=0. \qquad . \quad . \quad . \quad (93)$$

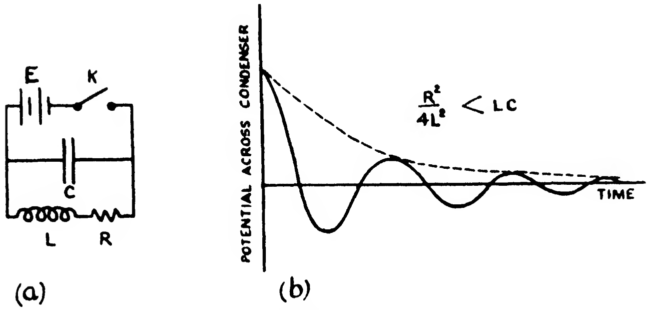


FIG. 24. *a*, The Freely Oscillating Circuit. *b*, Oscillatory Discharge ($R^2/4L^2 < LC$).

Put $q=e^{mt}$, therefore $dq/dt=me^{mt}$ and $d^2q/dt^2=m^2e^{mt}$. Then equation (93) becomes

$$e^{mt}(m^2+2mb+k^2)=0.$$

From the usual formula for the solution of a quadratic equation,

$$m=-b\pm\sqrt{(b^2-k^2)}.$$

$$\therefore \quad q=Ae^{[-b+\sqrt{(b^2-k^2)}]t}+Be^{[-b-\sqrt{(b^2-k^2)}]t} \qquad . \quad . \quad (94)$$

becomes a solution of the differential equation (93), two constants A and B being included because this equation is of the second order.

To evaluate the constants A and B, consider that when $t=0$, $q=q_0=CE$, the initial condenser charge.

Substituting in (94),

$$q_0=A+B. \qquad . \quad . \quad . \quad . \quad (95)$$

Also, when $t=0$, the current $dq/dt=0$.

But $dq/dt=[-b+\sqrt{(b^2-k^2)}]\,Ae^{[-b+\sqrt{(b^2-k^2)}]t}$
$+[-b-\sqrt{(b^2-k^2)}]\,Be^{[-b-\sqrt{(b^2-k^2)}]} \qquad . \quad (96)$

on differentiating equation (94).

Putting $dq/dt=0$ when $t=0$ in (96),

$$0=[-b+\sqrt{(b^2-k^2)}]A+[-b-\sqrt{(b^2-k^2)}]B. \quad . \quad (97)$$

Solving equations (95) and (97) simultaneously gives

$$A=\tfrac{1}{2}q_0\left(1+\frac{b}{\sqrt{(b^2-k^2)}}\right) \quad \text{and} \quad B=\tfrac{1}{2}q_0\left(1-\frac{b}{\sqrt{(b^2-k^2)}}\right),$$

and putting these values for A and B in (94) gives

$$q=q_0\left[\tfrac{1}{2}\left(1+\frac{b}{\sqrt{(b^2-k^2)}}\right)e^{[-b+\sqrt{(b^2-k^2)}]t} + \tfrac{1}{2}\left(1-\frac{b}{\sqrt{(b^2-k^2)}}\right)e^{[-b-\sqrt{(b^2-k^2)}]t}\right].$$

If $b>k$, so that $R^2/4L^2>1/LC$, then no further simplification of this equation is possible; the charge on the condenser falls off exponentially to zero.

On the other hand, if $b<k$, i.e. $R^2/4L^2<1/LC$, then a very different state of affairs exists, since $\sqrt{(b^2-k^2)}$ is then imaginary, and needs to be written as $j\sqrt{(k^2-b^2)}$.

Then $$q=q_0 \cdot e^{-bt}\left[\frac{e^{j\sqrt{(k^2-b^2)}t}+e^{-j\sqrt{(k^2-b^2)}t}}{2} + \frac{b}{\sqrt{(k^2-b^2)}}\cdot\frac{e^{j\sqrt{(k^2-b^2)}t}-e^{-j\sqrt{(k^2-b^2)}t}}{2j}\right],$$

and using the standard mathematical relationship that

$$\cos x=\frac{e^{jx}+e^{-jx}}{2}, \quad \text{and} \quad \sin x=\frac{e^{jx}-e^{-jx}}{2j}.$$

$$\therefore \quad q=\frac{q_0 \cdot ke^{-bt}}{\sqrt{(k^2-b^2)}}\cos\left[\sqrt{(k^2-b^2)}t-\theta\right]. \qquad (98)$$

where θ is such that

$$\tan\theta=\frac{b}{\sqrt{(k^2-b^2)}}, \quad \sin\theta=b/k \quad \text{and} \quad \cos\theta=\frac{\sqrt{(k^2-b^2)}}{k}.$$

On plotting the condenser charge q against time t from this equation (98), a curve of the form shown in fig. 24*b* is obtained, indicating the oscillatory character of the discharge action. Note that the amplitude of the oscillation diminishes with time at a rate governed by e^{-bt} or $e^{(-R/2L)t}$.

The current is given by differentiating (98) with respect to t, so

$$i=\frac{dq}{dt}=q_0ke^{-bt}.\sin[\sqrt{(k^2-b^2)}t-\theta]+\frac{q_0k.be^{-bt}}{\sqrt{(k^2-b^2)}}\cos[\sqrt{(k^2-b^2)}t-\theta]$$
$$-\frac{q_0ke^{-bt}}{\sqrt{(k^2-b^2)}}\{\sqrt{(k^2-b^2)}.\sin[\sqrt{(k^2-b^2)}t-\theta]+b\cos[\sqrt{(k^2-b^2)}t-\theta]\}$$
$$=\frac{q_0.k^2e^{-bt}}{\sqrt{(k^2-b^2)}}\sin\sqrt{(k^2-b^2)}t.$$

At times $t=0$, $\frac{\pi}{\sqrt{(k^2-b^2)}}$, $\frac{2\pi}{\sqrt{(k^2-b^2)}}$, etc., $i=0$.

Therefore time for one complete oscillation$=\frac{2\pi}{\sqrt{(k^2-b^2)}}$.

Therefore the frequency of the oscillatory current, f, is given by

$$f=\frac{1}{2\pi}\sqrt{(k^2-b^2)}=\frac{1}{2\pi}\sqrt{\left(\frac{1}{LC}-\frac{R^2}{4L^2}\right)}. \quad . \quad . \quad (99)$$

Since the energy originally stored in the condenser electrostatic field is interchanged during the oscillation cycle with the energy associated with the magnetic field round the inductance, then assuming no energy loss due to resistance and radiation of the circuit, the equation

$$\tfrac{1}{2}CE^2=\tfrac{1}{2}LI^2 \quad . \quad . \quad . \quad . \quad (100)$$

can be written, where $\frac{1}{2}CE^2$ is the electrical energy stored in a condenser, charged to potential E, and $\frac{1}{2}LI^2$ is the energy in the magnetic field round an inductance L, carrying current I.

$$\therefore \quad I=E\sqrt{\left(\frac{C}{L}\right)} \quad . \quad . \quad . \quad . \quad (101)$$

gives the peak current in the first cycle of oscillatory discharge.

The Logarithmic Decrement. The amplitude of the oscillating current in the discharge case falls exponentially. Let I_1 and I_2 be the values of the current at two instants of time t_1 and t_2 during the discharge.

Then $$\frac{I_1}{I_2}=e^{-b(t_1-t_2)}=e^{[-R(t_1-t_2)]/2L}=e^{[R(t_2-t_1)]/2L}.$$

Let $t_2-t_1=1/f$, where f is the oscillation frequency.

$$\therefore \quad \frac{I_1}{I_2}=e^{R/2fL}.$$

$$\therefore \quad \log_e\frac{I_1}{I_2}=\frac{R}{2fL}, \quad . \quad . \quad . \quad . \quad (102)$$

which is known as the logarithmic decrement.

Coupled Circuits. Consider two series *LCR* circuits coupled together by mutual induction M, as in fig. 25.

Let e be the instantaneous value of E.M.F. applied to the primary, which consists of L_1, C_1 and R_1 in series. The secondary is a closed circuit of L_2, C_2 and R_2 in series.

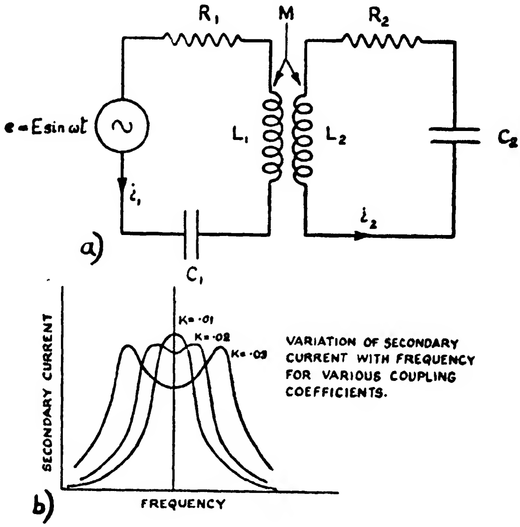


FIG. 25. Coupled Circuits.

The E.M.F. equation for the primary is

$$i_1R_1+ji_1X_1+j\omega Mi_2=e, \qquad (103)$$

where i_1=primary current, and $X_1=\omega L_1-1/\omega C_1$, the primary reactance.

For the secondary,

$$i_2R_2+ji_2X_2+j\omega Mi_1=0, \qquad (104)$$

where i_2=secondary current, and X_2=the secondary reactance.

From (104)

$$i_2=\frac{-j\omega Mi_1}{R_2+jX_2}=\frac{-j\omega Mi_1}{Z_2}, \qquad (105)$$

where Z_2=secondary impedance (when isolated from primary).

Substituting for i_2 in (103) from (105)

$$e=i_1R_1+ji_1X_1+\frac{\omega^2M^2}{Z_2}, \qquad (106)$$

$$\therefore \quad Z_1' = \frac{e}{i_1} = R_1 + jX_1 + \frac{\omega^2 M^2}{R_2 + jX_2}, \quad . \quad . \quad (107)$$

where Z_1' = effective primary impedance (including effect of coupled secondary circuit).

But $$\frac{\omega^2 M^2}{R_2 + jX_2} = \frac{\omega^2 M^2 (R_2 - jX_2)}{R_2^2 + X_2^2} = \frac{\omega^2 M^2}{Z_2^2}(R_2 - jX_2).$$

Therefore (107) becomes

$$Z_1' = R_1 + jX_1 + \frac{\omega^2 M^2}{Z_2^2}(R_2 - jX_2). \quad . \quad . \quad (108)$$

Equating real or resistive terms

$$R_1' = R_1 + \left(\frac{\omega^2 M^2}{Z_2^2}\right) R_2. \quad . \quad . \quad . \quad (109)$$

indicating that the effective resistance of the primary is increased by $(\omega^2 M^2 / Z_2^2) R_2$ as a result of the coupled secondary.

Also $$X_1' = X_1 - \left(\frac{\omega^2 M^2}{Z_2^2}\right) X_2, \quad . \quad . \quad . \quad (110)$$

showing that the effective reactance of the primary is decreased by $(\omega^2 M^2 / Z_2^2) X_2$ due to the coupling.

Optimum coupling between these circuits occurs for a value of mutual induction M for which the energy transfer from primary to secondary is a maximum, giving a maximum secondary current.

From (105) the secondary current $i_2 = \frac{-j\omega M i_1}{R_2 + jX_2}$ and substituting for i_1, from (107)

$$i_2 = \frac{-j\omega M}{R_2 + jX_2} \cdot \frac{e}{R_1 + jX_1 + (\omega^2 M^2 / R_2 + jX_2)}.$$

If it is considered that the two circuits concerned *are at resonance*, then $X_1 = X_2 = 0$.

$$\therefore \quad i_2 = \frac{-j\omega M e}{R_2 (R_1 + \omega^2 M^2 / R_2)}.$$

This will be a maximum when $di_2 / dM = 0$.

$$\therefore \quad \frac{d}{dM}\left(\frac{M}{R_1R_2+\omega^2M^2}\right)=0.$$

$$\therefore \quad \frac{R_1R_2+\omega^2M^2-M\,.\,2\omega^2M}{(R_1R_2+\omega^2M^2)^2}=0.$$

$$\therefore \quad R_1R_2=\omega^2M^2.$$

$$\therefore \quad \omega M=\sqrt{(R_1R_2)} \qquad . \quad . \quad . \quad . \quad (111)$$

gives the value of M required for optimum coupling between mutually coupled tuned circuits.

If a source of variable frequency A.C. is applied to the primary circuit and the variations of primary and secondary current are observed, then this coupled circuit will be found to exhibit *double-humping.* The currents will not be a maximum at the resonant frequency of either circuit considered separately, but at some frequency departing to a more or less extent from this value, the extent depending on the value of M.

Thus the frequencies at which resonance occurs are those at which the reactive opposition to current flow vanishes. Assume the coupled circuits have identical values of L and C, where $\omega_0=1/\sqrt{(LC)}$, the resonant pulsatance of either circuit separately. Then for the primary reactive component to vanish,

$$j\omega Li_1+\frac{i_1}{j\omega C}+j\omega Mi_2=0 \qquad . \quad . \quad . \quad (112)$$

and for the secondary,

$$j\omega Li_2+\frac{i_2}{j\omega C}+j\omega Mi_1=0. \qquad . \quad . \quad (113)$$

Multiplying (112) and (113) by $j\omega C$

$$i_1(1-\omega^2LC)-\omega^2MCi_2=0 \qquad . \quad . \quad (114)$$

and

$$i_2(1-\omega^2LC)-\omega^2MCi_1=0. \qquad . \quad . \quad (115)$$

From (114),

$$i_1=\frac{\omega^2MCi_2}{1-\omega^2LC},$$

substituting in (115)

$$\therefore \quad i_2(1-\omega^2LC)-\frac{i_2\omega^4M^2C^2}{(1-\omega^2LC)}=0.$$

Putting $LC=1/\omega_0^2$,

$$\therefore \quad \left(1-\frac{\omega^2}{\omega_0^2}\right)^2=\omega^4M^2C^2. \qquad . \quad . \quad (116)$$

By definition, the coupling coefficient k between two inductances L_1 and L_2, between which the mutual inductance is M, is $k=M/\sqrt{(L_1L_2)}=M/L$ when $L_1=L_2=L$.

Therefore (116) becomes

$$\left(1-\frac{\omega^2}{\omega_0^{\,2}}\right)^2=\omega^4k^2L^2C^2=\frac{\omega^4k^2}{\omega_0^{\,4}}. \qquad (117)$$

$$\therefore \quad 1-\frac{\omega^2}{\omega_0^{\,2}}=\pm\frac{\omega^2k^2}{\omega_0^{\,2}}.$$

$$\therefore \quad \frac{\omega^2}{\omega_0^{\,2}}(1\pm k)=1.$$

$$\therefore \quad \omega=\frac{\omega_0}{\sqrt{(1\pm k)}}. \qquad (118)$$

Hence the coupled system is resonant to two frequencies, f_1 and f_2 given by $f_1=\dfrac{f_0}{\sqrt{(1+k)}}$, and $f_2=\dfrac{f_0}{\sqrt{(1-k)}}$.

The band of frequencies which is passed by such a coupled circuit arrangement is therefore

$$f_0\left(\frac{1}{\sqrt{(1+k)}}-\frac{1}{\sqrt{(1-k)}}\right)$$
$$=f_0[(1+k)^{-\frac{1}{2}}-(1-k)^{-\frac{1}{2}}]=f_0[1+\tfrac{1}{2}k-1+\tfrac{1}{2}k]$$
$$=kf_0, \qquad (119)$$

if k is small, i.e. the coupling is loose.

Non-sinusoidal A.C. Wave-forms. In electronic practice the occurrence of alternating currents and potentials with a regular period, but a wave-shape other than sinusoidal, is very common. Examples are the currents produced by a microphone corresponding to musical sounds of various wave-forms, the output from a rectifier (see p. 88), saw-tooth and rectangular wave-forms (p. 190) generated for specialised applications. In dealing with electronic circuits of all kinds the use of inductance, condenser, resistance and valve combinations demands A.C. theory for computation work, employing impedance and other formulae based on the simplifying assumption that sinusoidal supplies of pulsatance $\omega=2\pi f$ are used. To obtain intelligent results when circuits are employed with non-sinusoidal supplies, it becomes necessary to relate their wave-forms to the sinusoidal case. Fortunately, except in the cases of currents with irregular discontinuities, or irregular period, the analysis of any periodic wave into a fundamental sinusoidal wave plus waves which have frequencies which are various multiples of the fundamental frequency, can be performed by the method due to Fourier.

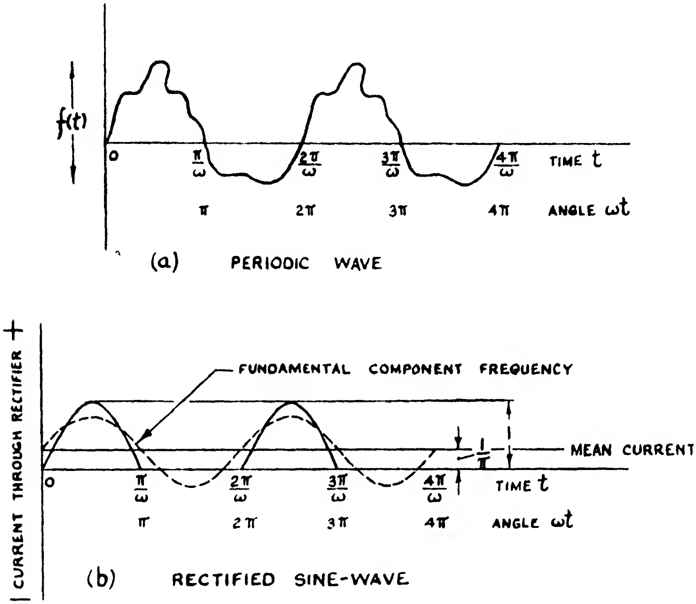


FIG. 26. *a*, Periodic Wave *b*, Wave-form for Rectifier Output

If the fundamental frequency, or first harmonic is f, then a frequency of $2f$ is the second harmonic, and in general a frequency of nf is the nth harmonic of f.

Fourier Analysis. Let a periodic wave (fig. 26) be such that its displacement at any time t is represented by $f(t)$, where f is any function which has only one value for a given value of t.

According to Fourier,

$$f(t)=A_1 \sin \omega t+A_2 \sin 2\omega t+A_3 \sin 3\omega t+ \ldots +A_n \sin n\omega t+ \ldots$$
$$+B_1 \cos \omega t+B_2 \cos 2\omega t+B_3 \cos 3\omega t+ \ldots$$
$$+B_n \cos n\omega t+ \ldots +k+ \ldots, \qquad (120)$$

where k is constant, $A_n \sin n\omega t+B_n \cos n\omega t$ representing completely the nth harmonic component.

It remains to find the actual values of A_1, $A_2 \ldots A_n$ etc., and B_1, B_2 etc., in order that the relative amplitudes of the component frequencies at the various harmonics are known. This is done by various trigonometrical methods, depending on the form of $f(t)$.

Formulae for A_n and B_n can be readily obtained by considering that

$$\int_0^{2\pi/\omega} f(t) \sin n\omega t \, dt = \int_0^{2\pi/\omega} [A_1 \sin \omega t + \ldots + A_m \sin m\omega t + \ldots + B_1 \cos \omega t + \ldots + B_m \cos m\omega t + \ldots + k] \sin n\omega t \, dt,$$

which is obtained from (120) on integrating (120) multiplied by $\sin n\omega t$ between the limits 0 and $2\pi/\omega$ corresponding to one period.

But $$\int_0^{2\pi/\omega} \cos m\omega t \sin n\omega t \, dt = \int_0^{2\pi/\omega} \sin m\omega t \sin n\omega t \, dt = 0$$

except when $m = n$.

$$\therefore \quad \int_0^{2\pi/\omega} f(t) \sin n\omega t \, dt = A_n \int_0^{2\pi/\omega} \sin^2 n\omega t \, dt = \frac{A_n \pi}{\omega}.$$

$$\therefore \quad A_n = \frac{\omega}{\pi} \int_0^{2\pi/\omega} f(t) \sin n\omega t \, dt, \qquad (121)$$

and, similarly, $$B_n = \frac{\omega}{\pi} \int_0^{2\pi/\omega} f(t) \cos n\omega t \, dt. \qquad (122)$$

So A_1, A_2, A_3, B_1, B_2, B_3 etc. can be found if $f(t)$ is known for a particular case, and n is put equal to 1, 2, 3, etc.

As an example of the use of Fourier's method, consider the case of half-wave rectification (p. 89). Inspection of fig. 26*b* shows that $f(t)$ has the form $V_R \sin \omega t$ for 0 to π/ω (half a period), and is zero from π/ω to $2\pi/\omega$, V_R being the peak voltage occurring across the load resistance.

$$\therefore \quad A_n = \frac{\omega}{\pi} \int_0^{\pi/\omega} V_R \sin \omega t \sin n\omega t \, dt + 0 \qquad \text{from (121)}$$

and $$B_n = \frac{\omega}{\pi} \int_0^{\pi/\omega} V_R \sin \omega t \cos n\omega t \, dt + 0 \qquad \text{from (122).}$$

Put $n = 1, 2, 3$, etc., to find the amplitudes of the 1st, 2nd, 3rd harmonics, etc.

Thus $A_1=\frac{V_R\omega}{\pi}\int_0^{\pi/\omega}\sin^2\omega t\,dt=\frac{V_R}{2}$, ... (123)

$$A_2=\frac{V_R\omega}{\pi}\int_0^{\pi/\omega}\sin\omega t\sin 2\omega t\,dt=0, \text{ and similarly for } A_3, A_4, \text{ etc.}$$

$$B_n=\frac{\omega}{\pi}\int_0^{\pi/\omega}V_R\sin\omega t\cos n\omega t\,dt.$$

Now $\sin\omega t\cos n\omega t=\frac{\sin(n+1)\omega t-\sin(n-1)\omega t}{2}$.

Therefore B_n is more conveniently written as

$$\frac{\omega V_R}{2\pi}\int_0^{\pi/\omega}\Big[\sin(n+1)\omega t-\sin(n-1)\omega t\Big]dt$$

$$=\frac{V_R\omega}{2\pi}\left[-\frac{\cos(n+1)\omega t}{(n+1)\omega}+\frac{\cos(n-1)\omega t}{(n-1)\omega}\right]_0^{\pi/\omega}$$

$$=\frac{V_R}{2\pi}\left[\frac{1-\cos(n+1)\pi}{n+1}-\frac{1-\cos(n-1)\pi}{n-1}\right].$$

When n is odd this reduces to zero, when n is even this reduces to $\frac{V_R}{2\pi}\cdot\frac{-2}{(n-1)(n+1)}$.

$$\therefore\quad B_2=-\frac{1}{3}\frac{V_R}{\pi}\quad\text{putting } n=2,$$

$$B_4=-\frac{1}{3\cdot5}\frac{V_R}{\pi}\quad\text{putting } n=4,$$

and so $$f(t)=\frac{V_R}{\pi}\left[k+\frac{\pi}{2}\sin\omega t-\frac{1}{3}\cos 2\omega t-\frac{1}{3\cdot5}\cos 4\omega t-\ldots-\frac{1}{(n-1)(n+1)}\cos n\omega t\right],$$

giving the amplitudes of the even harmonic components found in the output of a half-wave rectifier.

D.C. Equivalent to Mean Half-Wave Pulse of A.C. The average current $I_{D.C.}$ equivalent to a half-wave rectified current pulse of form $I_0\sin\omega t$, where I_0 is the peak current, is given by considering that during half a period π/ω the quantity of electricity concerned in one pulse is $\int_0^{\pi/\omega}I_0\sin\omega t\,dt$ which equals $I_{D.C.}\cdot\pi/\omega$, where $I_{D.C.}$ is the mean current.

$$\therefore \int_0^{\pi/\omega} I_0 \sin \omega t \, dt = I_0 \left[\frac{-\cos \omega t}{\omega} \right]_0^{\pi/\omega} = \frac{I_0}{\omega} \left[-1(-1)+1 \right] = I_{D.C.} \cdot \frac{\pi}{\omega}.$$

$$\therefore \quad I_{D.C.} = \frac{2I_0}{\pi}. \quad . \quad . \quad . \quad . \quad . \quad (124)$$

H.F. Resistance. At radio frequency the effective resistance of a conductor is greater than its value to direct current or low frequency alternating current. When current is passing through the cross-section of a conductor (fig. 27) it must be realised that those electrons which traverse the conductor core are surrounded by circular lines of magnetic force within as well as outside the conductor. So the alternation of current through the core of a conductor is accompanied by the variation of a greater number of surrounding lines of magnetic force than is the alternation of current through the skin of the conductor.

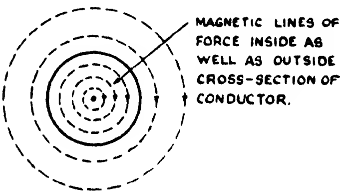


FIG. 27. The H.F. Resistance of a Conductor.

Consequently, the reactive opposition to the flow of rapidly alternating current through the core is greater than the impedance to the flow of such current through the surface layers. Alternating currents of frequencies of 1 Mc./s. and more therefore pass through the skin of the conductor rather than through the centre. As a result, the effective cross-section area of the conductor is reduced, with a consequent increase of its resistance.

This effect can be conveniently expressed by an approximate formula

$$\frac{R_{A.C.}}{R_{D.C.}} = d\sqrt{\left(\frac{\mu f}{\rho}\right)}, \quad . \quad . \quad . \quad . \quad (125)$$

where $R_{A.C.}$ = the resistance of the conductor to alternating current, $R_{D.C.}$ = the resistance of the conductor to direct current, d = the diameter of the conductor, μ = the permeability of the material of the conductor, ρ = the resistivity of the material of the conductor, f = the frequency of the alternating current.

Transmission Lines. Lines in the form of a parallel-pair, or coaxial cable, are extensively used to convey the energy from an oscillator to an aerial in transmitter practice. In ultra-high frequency practice they also play an important part as circuit elements, being in effect combinations of inductance and capacitance. A simplified treatment is considered here to establish an understanding of their action as circuit elements only.

Let Z=series impedance per unit length of line$=R+j\omega L$, where R=series resistance per unit length, and L=series inductance per unit length.

Suppose Y=shunt admittance per unit length$=G+j\omega C$, where G=shunt conductance per unit length, and C=shunt capacitance per unit length.

Referring to fig. 28a, suppose E is the instantaneous potential difference at a point x along the line, and I is the current. An alternator of pulsatance ω feeds the line, and distances x to any point on the line are measured from the receiving end.

Then $$\frac{dE}{dx}=IZ, \qquad (126)$$

and $$\frac{dI}{dx}=EY. \qquad (127)$$

$$\therefore \quad \frac{d^2E}{dx^2}=Z\frac{dI}{dx}=ZEY. \qquad (128)$$

(128) is the differential equation relating the current and potential in the line at any point x along it.

This equation can be shown to have a solution in the form

$$E=C\cosh\sqrt{(ZY)}x+D\sinh\sqrt{(ZY)}x. \qquad (129)$$

For present purposes, however, a simple mathematical treatment that is adequate can be considered for a low-loss line where R and G are neglected compared with the effects of L and C.

Then $$Z=j\omega L, \qquad (130)$$

and $$Y=j\omega C. \qquad (131)$$

$$\therefore \quad \frac{dE}{dx}=IZ=j\omega LI, \quad \text{and} \quad \frac{dI}{dx}=EY=jC\omega E.$$

$$\therefore \quad \frac{d^2E}{dx^2}=Z\frac{dI}{dx}=j\omega L\frac{dI}{dx}=j\omega L.j\omega CE=-E\omega^2LC.$$

$$\therefore \quad \frac{d^2E}{dx^2}=-E\omega^2LC. \qquad (132)$$

Put $E=A\cos\alpha x+B\sin\alpha x$.

$$\therefore\quad \frac{dE}{dx}=-A\alpha\sin\alpha x+B\alpha\cos\alpha x,$$

and
$$\frac{d^2E}{dx^2}=-A\alpha^2\cos\alpha x-B\alpha^2\sin\alpha x$$
$$=-\alpha^2(A\cos\alpha x+B\sin\alpha x).$$

Hence if $\alpha^2=\omega^2LC$, then

$$E=A\cos\alpha x+B\sin\alpha x \qquad . \quad . \quad . \quad (133)$$

is a solution of (132).

To evaluate the constants A and B, consider that, when $x=0$, $E=E_R$, where E_R=received voltage.

Substituting in (133), $x=0$ and $E=E_R$ gives $E_R=[A+0]$.

$$\therefore\quad A=E_R$$

Also
$$\frac{dE}{dx}=\frac{d}{dx}(A\cos\alpha x+B\sin\alpha x)$$
$$=(-\alpha A\sin\alpha x+\alpha B\cos\alpha x)=\alpha(-A\sin\alpha x+B\cos\alpha x). \qquad . \quad . \quad . \quad (134)$$

When $x=0$, $I=I_R$, the received current, and $\frac{dE}{dx}=Ij\omega L=I_Rj\omega L$.

Put $x=0$ and $\frac{dE}{dx}=I_Rj\omega L$ in (134).

$$\therefore\quad I_Rj\omega L=\alpha(0+B).$$

$$\therefore\quad B=\frac{I_Rj\omega L}{\alpha}=\frac{I_Rj\omega L}{\sqrt{(\omega^2LC)}} \text{ on substituting for } \alpha.$$

Therefore equation (133) can be written, on substituting for A, B and α as

$$E=E_R\cos\omega\sqrt{(LC)}x+jI_R\sqrt{(L/C)}\sin\omega\sqrt{(LC)}x, \qquad (135)$$

which is the fundamental equation relating the potential and current for a loss-less transmission line.

From electromagnetic theory, where the medium surrounding the lines is free space, so $\mu=k=1$ in the appropriate units, then $c=1/\sqrt{(LC)}$, where c is the velocity of light in free space.*

* See Willis Jackson, *High Frequency Transmission Lines*, Methuen, 1947.

But $c = \frac{\omega}{2\pi} \times \lambda$ (velocity = frequency × wave-length).

$$\therefore \quad \omega\sqrt{(LC)} = \frac{2\pi}{\lambda}.$$

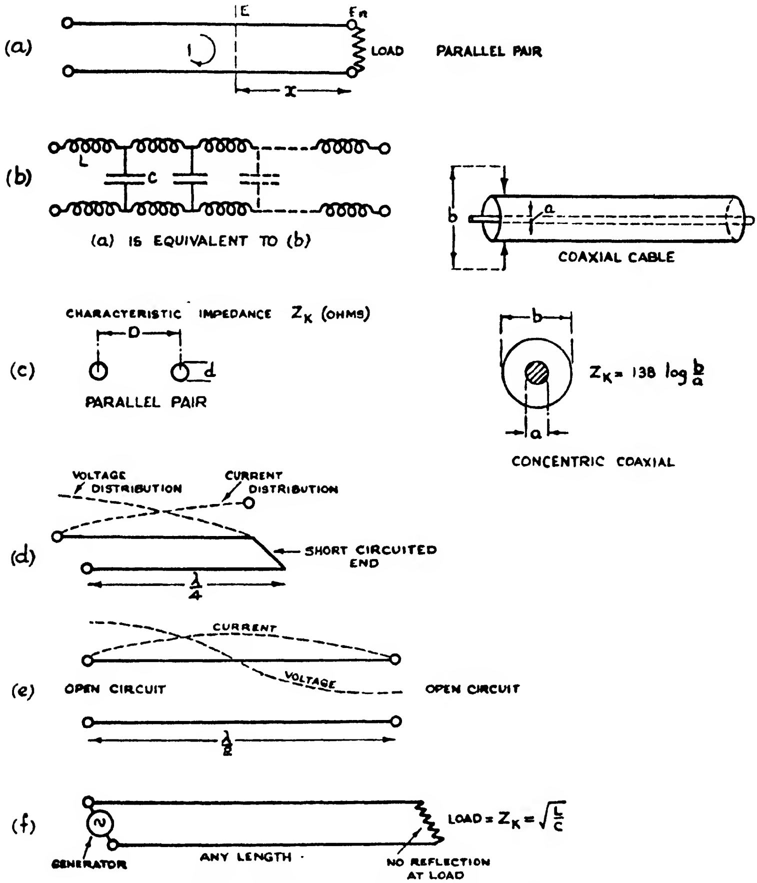


Fig. 28. Transmission-Lines as Circuit Elements.

Therefore equation (135) can be written in the more useful form:

$$E = E_R \cos \frac{2\pi x}{\lambda} + jI_R\sqrt{(L/C)} \sin \frac{2\pi x}{\lambda}, \qquad (136)$$

and

$$I = I_R \cos \frac{2\pi x}{\lambda} + j\frac{E_R}{Z_K} \sin \frac{2\pi x}{\lambda}, \qquad (137)$$

where I is similarly obtained by integrating the equation

$$d^2I/dx^2 = -\omega^2 LCI \quad \text{and} \quad Z_K = \sqrt{(L/C)}$$

is called the line characteristic impedance.

A transmission line terminated by a resistance equal to its characteristic impedance $\sqrt{(L/C)}$ in ohms acts as a standard resistance at high frequencies. All the energy transmitted down such a line is absorbed by the line: there is no reflection at the load, and no standing waves. This is the ideally matched case where there is no radiation of energy from the line due to standing current waves in it.

Thus, consider that a resistance of value $Z_K = \sqrt{L/C}\ \Omega$. is connected across the output end of a transmission line, as in fig. 28*f*. Then the received voltage divided by the received current must equal the load resistance, so $E_R/I_R = Z_K$. Inserting this relation in equations (136) and (137), it follows that

$$E = E_R \cos \frac{2\pi x}{\lambda} + jE_R \sin \frac{2\pi x}{\lambda}$$

and

$$I = \frac{1}{Z_k}\left(E_R \cos \frac{2\pi x}{\lambda} + jE_R \sin \frac{2\pi x}{\lambda}\right),$$

$$\therefore \quad \frac{E}{I} = \frac{\text{voltage at sending end}}{\text{current at sending end}} = Z_k,$$

irrespective of the length of the line, x.

Hence a transmission line terminated by a load equal to its characteristic impedance absorbs the same energy from the supply source as if the load were connected directly across the generator terminals. There is no energy radiated from the line.

Referring to fig. 28*c*, the characteristic impedance Z_K, for a parallel-pair of lines$=276 \log_{10} 2D/d\ \Omega$., and for a coaxial cable, $Z_K = 138 \log_{10} b/a\ \Omega$.

If a transmission line is short-circuited at the end, then the received voltage E_R must equal zero. Put $E_R = 0$ in equations (136) and (137), then the impedance at the generator end, a distance x from the receiving load, is given by

$$Z = \frac{E}{I} = \frac{jI_R Z_K \sin 2\pi x/\lambda}{I_R \cos 2\pi x/\lambda} = jZ_K \tan \frac{2\pi x}{\lambda}.$$

Since j is positive in this formula, the impedance must be an inductive reactance, and since the tangent of an angle varies from zero to infinity, so the magnitude of this impedance can have any value.

When $x=\frac{\lambda}{4}$, $Z=jZ_K \tan\frac{2\pi\lambda/4}{\lambda}=jZ_K \tan\frac{\pi}{2}=\infty$.

Therefore a line a quarter-wave-length long, short-circuited at its end, has an input impedance of infinity, neglecting losses. It compares as a circuit element with a rejector circuit, having a maximum input voltage with minimum current at a frequency decided by its length (see fig. 28*d*).

Similarly, an open-circuited line has a receiving-end current of zero, so $I_R=0$ (fig. 28*e*),

$$\therefore \quad Z=\frac{E}{I}=\frac{E_R \cos 2\pi x/\lambda}{j(E_R/Z_K) \sin 2\pi x/\lambda}=-jZ_K \cot\frac{2\pi x}{\lambda},$$

a capacitative reactance.

Put $x=\frac{\lambda}{4}$, $\therefore \quad Z=-jZ_K \cot\frac{\pi}{2}=0.$

This corresponds to zero input impedance, the generator voltage being a minimum when the current is a maximum, whilst the received voltage is a maximum with minimum current. It gives, therefore, a voltage magnification effect, comparable with that of an acceptor circuit at the resonant frequency.

In both cases, the opposite kind of reactance is obtained when the length of the section of line lies between $\lambda/4$ and $\lambda/2$, since the sign of the trigonometric function concerned changes.

Considering the transmission-line as a resonator somewhat more fully. As in the cases of ordinary acceptor and rejector circuits, a "*Q*" factor deciding the voltage or current magnification may be quoted. "*Q*" is defined as $\frac{\text{the reactive power in the line}}{\text{power loss}}$.

Assuming the distribution of the current along the tuned line to be sinusoidal, varying from a maximum at the shorted end to almost zero at the open end, then the power loss per $\lambda/4$ length of line$=P$, where

$$P=\left(\frac{I}{\sqrt{2}}\right)^2\frac{\lambda}{4}\cdot R=\frac{I^2R\lambda}{8},$$

I being the R.M.S. current in the line, and R=resistance per metre. The reactive power=$\omega L(I/\sqrt{2})^2 \, . \, \lambda/4$, where L=inductance per metre.

$$\therefore \quad Q = \frac{\omega L I^2 \lambda}{8} \div \frac{I^2 R \lambda}{8} = \frac{\omega L}{R}, \qquad (138)$$

a formula comparing immediately with that for the usual acceptor and rejector circuits.

A useful practical note here is that for a coaxial line used as a resonator, Q is a maximum when b/a=3·6 (see fig. 28).

CHAPTER 5

Diode and Triode Valves and Basic Associated Circuits

The Construction of a Diode. A diode valve consists essentially of a heated conductor emitting electrons which are attracted or repelled by a surrounding plate, or anode, depending on whether the potential on this plate is positive or negative with

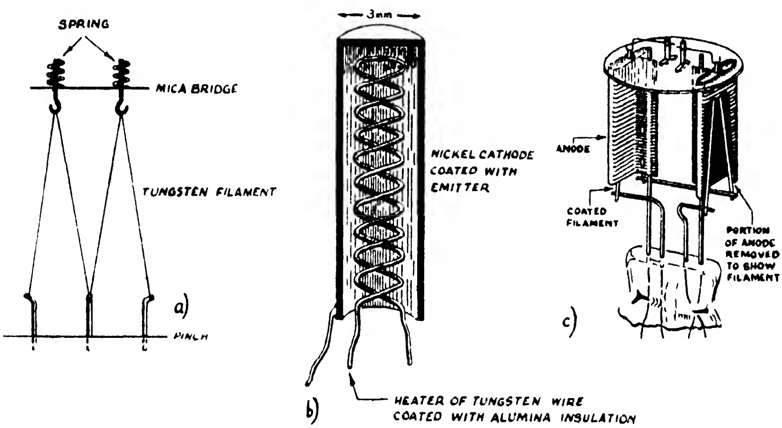


FIG. 29. *a*, A Filamentary Type Cathode. *b*, An Indirectly Heated Cathode. *c*, A Double-diode Valve.

respect to the emitter. The emitter, being usually at a negative potential with respect to the plate, is termed the cathode, and may be either of the filamentary type (fig. 29*a*), or the indirectly heated type (fig. 29*b*). In general, the directly heated filament is used when the heating current is steady D.C., from a battery, whereas the indirectly heated cathode is used when the heating current is alternating, supplied by an A.C. transformer.*

A glass pinch (fig. 29*c*) is formed by pressing glass on to lead-in wires consisting of nickel rods which are electrically welded to thin, copper-clad nickel-iron wires. Such copper-clad wire, when treated with borax flux, will make a vacuum-tight seal into glass.

* Exceptions occur in the case of valves such as the PX4, and PX25, which have filaments of sufficiently large thermal capacity, and so thermal lag, to permit direct heating by A.C.

The inverted V- or M-shaped filament is suspended by a hook under the tension of a spring from the top mica bridge, so as to be along the mid-plane of the nickel anode, and is welded at the bottom to the nickel rods which are fused into the pinch. A shaped, glass envelope is placed over the assembly, and fused to the flange at the pinch base by rotation of the valve inside a circular array of gas-air jets. The exhaust tube, through which the valve is evacuated by the pumping system, is joined to an aperture just below the pressed glass of the pinch. This tube is sealed off by applying a jet-flame near the base of the valve after exhaustion and activation are completed. The pinch wires are then threaded through the pegs of a moulded bakelite valve cap, and soldered to the tips of these pegs, the cap being joined by bakelite cement to the glass envelope. Thus a mechanically robust structure is obtained.

When a valve has been exhausted of air by a mechanical rotary vacuum pump it is baked in an oven at 300° to 400° C., the anode is heated to some 900° C. by an eddy-current furnace, the filament "flashed", and finally the "getter" volatilised. The getter is in the form of a magnesium strip, or copper-clad barium pellet supported by a metal disk which is placed near the valve pinch, away from the main electrodes. The temperature of this disk is raised to yellow heat by eddy-currents produced by a powerful radio-frequency oscillator so that the barium or magnesium volatilises and the active metal vapour produced combines chemically with any residual gas in the valve to "fix" it in the form of a low vapour-pressure magnesium or barium compound on the glass walls of the tube. The getter deposit appears as a silver mirror if it is magnesium, whereas a brownish-silver deposit is obtained if barium is used.

The final vacuum produced by this means is of the order of 10^{-4} mm. Hg. At such a low pressure the electrons emitted from the cathode will be permitted a free path of about 400 cm., unobstructed by gas molecules. Manifestly, there will be little possibility of ionisation of the residual gas by collision so that, for most practical purposes, such a vacuum is adequate.

The Action of a Diode. If a diode valve is placed in the electric test circuit of fig. 30*b*, its electrical behaviour can be investigated.

If the anode is positive, the negative electrons will be attracted to it to form an anode current, recorded by the series milliammeter.

If the H.T. battery is reversed, making the anode negative, then no anode current will flow since the emitted electrons are repelled. As a positive potential on the anode is raised the anode current increases until a saturation effect is produced, when further anode potential increase does not cause any further appreciable rise of anode current. The diode characteristic curve of anode current I_A plotted against anode voltage V_A, has the form shown in fig. 30*a*.

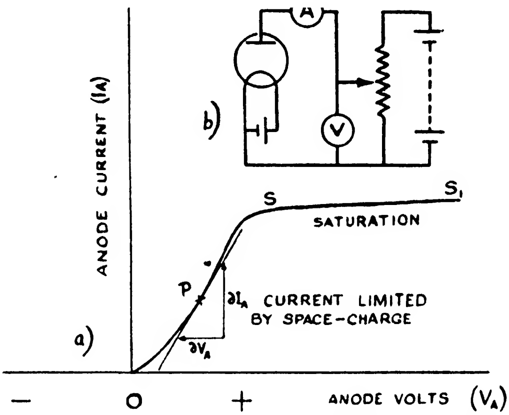


FIG. 30. *a*, Characteristic Curve of a Diode Valve. *b*, Circuit for obtaining the Diode Characteristic.

Explanation of the Shape of the Characteristic Curves. The factors determining the shape of this characteristic are (*a*) the total cathode emission, (*b*) the *space-charge* effect, and (*c*) the Schottky effect.

(*a*) The total cathode emission is determined by the insertion of the appropriate values of the constants A_0, T and b in equation (42), the material used as emitter being usually, nowadays, the barium oxide-strontium oxide mixture described on p. 34, operated at a temperature of approximately 800° C. (=1073° K.). The Richardson equation gives the emission in amperes per square centimetre, so it must be multiplied by the effective area of the cathode coating in a particular case. This cathode emission is, at specified operating conditions, constant. All the electrons emitted do not reach the anode, however, because of the space-charge effect.

(*b*) The space-charge effect is of paramount importance in deciding the action of all types of radio valve. Electrons, streaming from the cathode with an average velocity of 0·3 electron-volt imparted to them thermally, will not have achieved much acceleration towards the attracting anode when they have only just left the emitter. Being all charged negatively they will repel one another, and tend to repel further electrons being emitted from the cathode. This negative-charged electron cluster will also partially shield the effect of the attraction of the anode on the cathode.

Moreover, on leaving, an electron induces an equal positive charge on the cathode which tends to pull it back again.

This space-charge will control the number of electrons which reach the anode, since only those emitted with sufficiently high thermal velocities will be able to penetrate to the anode. As the anode potential is raised positively, so the space-charge effect will be reduced because the electrons receive greater accelerations towards the anode, tending to prohibit the formation of a space-charge. At the saturation anode-potential the space-charge will, for this reason, disappear, and all the thermally emitted electrons will reach the anode, giving the saturation anode current. A further increase of anode voltage cannot then produce any further anode current rise.

The potential distribution between a plane-parallel cathode and anode is represented by a straight line graph *ab* in fig. 31*b*, when the cathode is cold, and the anode positive. On raising the cathode temperature electrons are emitted, and with moderate anode potentials, a space-charge forms near the cathode. The potential in the field in the region near the cathode will then be more negative than before, the curved graph *acb* representing the distribution. Depending on the amount of emission and the anode potential, so the dip in the potential distribution curves will be more or less pronounced. If such curves as these are compared with the graph of fig. 12, representing the distribution of velocities amongst thermally emitted electrons at a particular temperature, it will be realised that only those electrons with sufficient thermally imparted speeds will be able to penetrate the space-charge region, and reach the anode. Since there are, depending on the prevailing conditions, a definite fraction of emitted electrons with velocities above a specified value, so the anode current will be a definite amount (cf. Maxwell's distribution law, p. 30). Increasing the

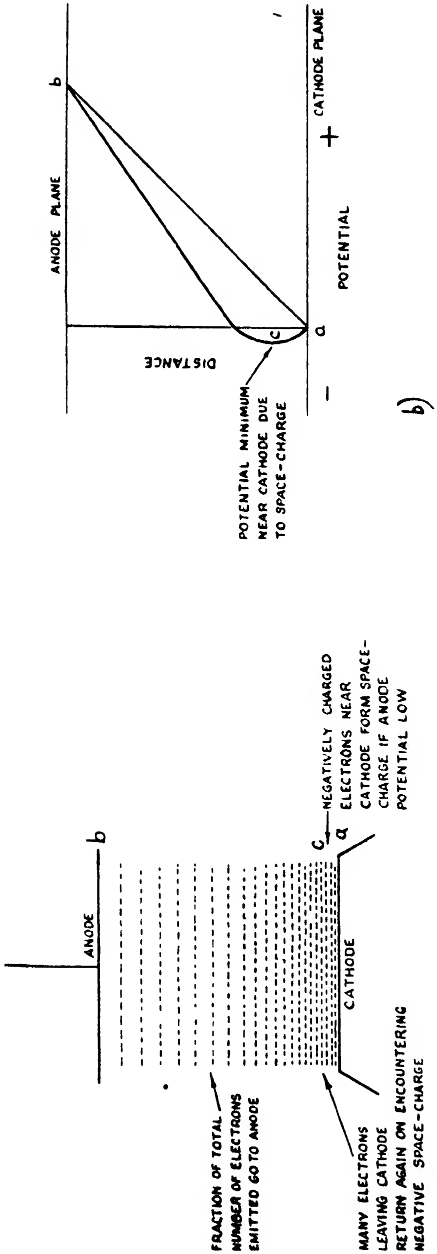


Fig. 31. *a*, The Space-charge Effect. *b*, Potential Distribution in a Diode.

cathode temperature will increase the anode current, though it should be borne in mind that, using modern oxide-coated cathodes, too high a temperature will volatilise some of the barium, tending to destroy the emitter.

A mathematical analysis of the space-charge effect should consider the distribution of velocities amongst the thermionically emitted electrons. Since such an approach leads to complicated expressions, a simpler method is presented here which, nevertheless, predicts results which are useful in practice.

The Poisson equation (see p. 11) is necessarily involved, since it defines the relation between charge and potential in an electrostatic field. Introducing the further simplification that a plane-parallel electrode system is adopted, so that the electrostatic field is uniform, then Poisson's equation reduces to the convenient form

$$\frac{d^2V}{dx^2} = -4\pi ne, \quad . \quad . \quad . \quad (139)$$

where V is the potential at any point distance x from the cathode in a direction perpendicular to the cathode-anode planes, and ρ, the volume density of charge in equation (14), here becomes $-ne$, where n is the number of electrons per cubic centimetre each of charge e. The y and z terms in the Poisson expression need not be taken into account since the electric field varies in the x-direction only. This is tantamount to considering the lines of force in the field as straight and parallel to one another, stretching normally from the cathode to the anode.

Assuming that the thermally imparted velocities are negligible, i.e. the electrons are emitted from the cathode with zero velocity, then from equation (57), p. 43,

$$\tfrac{1}{2}mv^2 = Ve, \quad . \quad . \quad . \quad . \quad (140)$$

where v is the electron velocity after undergoing a potential fall of V in the field.

The anode current

$$I_A = nev, \quad . \quad . \quad . \quad . \quad (141)$$

since current is dependent on the rate of movement of electric charge.

Substituting for ne in equation (139) from equations (141) and (140) gives

$$\frac{d^2V}{dx^2} = -4\pi I_A . \sqrt{\frac{m}{2eV}}. \quad . \quad . \quad (142)$$

By calculus, $2\frac{d^2V}{dx^2}=\frac{d}{dV}\left(\frac{dV}{dx}\right)^2$, so equation (142) integrates to give

$$\left(\frac{dV}{dx}\right)^2=16\pi I_A\sqrt{\left(\frac{m}{2e}\right)}V^{1/2}+C. \qquad (143)$$

The constant of integration C vanishes since the potential V, and the electric field dV/dx at the cathode are both zero.

The boundary conditions are $V=0$ when $x=0$, and $V=V_A$ when $x=a$, where V_A is the anode potential, and a is the cathode-anode clearance. Integration of equation (143) then gives

$$I_A=\frac{\sqrt{2}}{9\pi}\sqrt{\left(\frac{e}{m}\right)}\frac{V_A^{3/2}}{a^2}. \qquad (144)$$

If I_A is in amp./sq. cm., a in cm., and V_A in volts, then equation (144) becomes

$$I_A=\frac{2{\cdot}33\times10^{-6}}{a^2}V_A^{3/2} \qquad (145)$$

on substituting for e/m (see p. 43).

Study of this equation evinces that the anode-current depends on the three-halves power of the anode potential. The characteristic curve of fig. 30a will have a shape over the region OS corresponding to an equation of the form

$$I_A=KV_A^{3/2}, \qquad (146)$$

where K is a constant for a given cathode-anode clearance.

Also the space-charge limited current is inversely proportional to the square of the cathode-anode clearance in a plane-parallel type of electrode arrangement.

In the case of a diode valve in which the cathode and anode are coaxial cylinders, I. Langmuir* has expressed the space-charge equation in the form

$$I_A=\frac{2\sqrt{(2)}}{9}\sqrt{\left(\frac{e}{m}\right)}\cdot\frac{V_A^{3/2}}{r\beta^2}, \qquad (147)$$

where r is the anode radius, and β is a constant resulting from the integration of Poisson's equation expressed in cylindrical coordinates under the conditions prevailing. Langmuir expressed the constant β, which depends on the ratio r/a, where a is the cathode radius, in the form of a graph of r/a vs. β^2 (see fig. 32).

* I. Langmuir and K. Blodgett, *Phys. Rev.*, 22, 347, 1923.

This graph in conjunction with equation (147) is often of value when designing a valve, in estimating the effects produced by a change in the cathode or anode radius, since a fair number of diodes conform to the cylindrical shape. It is worth noting that, whereas in the plane-parallel diode the anode current varies inversely as the square of the anode-cathode clearance yet, in the cylindrical case, the anode current varies inversely as the first power of the anode radius. If a valve is required in which the

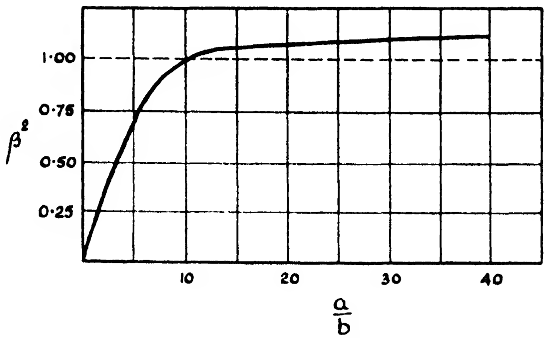


FIG. 32. Graph for Cylindrical Form Diode.

anode-cathode capacity has to be considered, this fact will often predetermine whether the plane-parallel or cylindrical form is to be used. The three-halves power law can be said to be correct, to a fair degree of accuracy, irrespective of the geometrical arrangement of the electrodes and is, therefore, useful in estimating the effects of anode potential change on the anode current.

P. S. Epstein*, in an analysis of the diode valve, considered the thermally imparted distribution of velocities amongst the emitted electrons. The equation given is

$$I_A = \frac{\sqrt{2}}{9\pi}\sqrt{\frac{e}{m}}\,\frac{(V_A - V_M)^{3/2}}{(a - x_M)^{3/2}}\left(1 + \frac{2{\cdot}66}{\sqrt{b}}\right), \quad . \quad (148)$$

where V_A, a and x have their former significance; V_M is the potential minimum between cathode and anode, produced by the space-charge effect at a distance x_M from the cathode; and $b = \frac{e(V_A - V_M)}{kT}$, where k is Boltzmann's constant, and T is the cathode absolute temperature.

* P. S. Epstein, *Deutsch Phys. Gesell. Verh.*, 21, 85, 1919.

(c) Schottky* has shown that the total electron emission from a thermionic cathode at a specific temperature is increased to a small extent by the application of a positive electric field at the cathode surface. In the case of a diode it is evident that, as a result, the saturation current obtained is not actually constant. The saturation anode current is obtained when the anode potential is sufficient to draw all the emitted electrons from the cathode to the anode so that the space-charge effect breaks down. According to simple theory no further increase of anode current can then occur if the anode potential is raised further, since all the emitted electrons are going immediately to the anode. However, because of the Schottky effect, a small increase of actual emission takes place as the positive field at the cathode is raised, so the portion SS_1 of the characteristic (fig. 30*a*) is not perfectly parallel to the anode volts axis, but rises slightly. This effect is somewhat more pronounced if an indirectly heated cathode is used, since then the cathode surface is comparatively rough, consisting of sprayed-on cathode coating containing many minute holes. As the positive anode potential is raised so the positive field at the cathode surface penetrates more and more into the small recesses in the cathode surface, and consequently causes the release of further electrons from these cavities.

The A.C. Resistance of a Diode. The variation of anode current produced by a small variation of anode potential is called the A.C. resistance of the valve. It must be measured at a specified mean anode potential, since it is not constant. From the characteristic curve of fig. 30*a* it is readily seen that the A.C. resistance at a potential decided by the point *P* on the curve depends on the reciprocal of the slope of the curve at point *P*.

$$\text{The A.C. resistance} = \frac{\text{a small change of anode voltage}}{\text{the corresponding small change of anode current}}.$$

If the current change is measured in amperes, then the resistance is given in ohms.

Symbolically, $$R_A = \frac{dV_A}{dI_A}. \quad \ldots \quad (149)$$

* W. Schottky, *Phys. Zeit.*, **15**, 872, 1914.

The A.C. resistance is also called the anode impedance, the internal resistance and the slope resistance by various authorities.

Since from equation (146), $I_A = KV_A^{3/2}$, differentiating with respect to V_A

$$\therefore \quad \frac{dI_A}{dV_A} = \frac{3}{2}K \cdot V_A^{1/2}.$$

$$\therefore \quad \frac{dV_A}{dI_A} = R_A = \frac{2}{3K} V_A^{-1/2}$$

$$= K_1 V_A^{-1/2},$$

where K_1 is a new constant.

But $I_A = KV_A^{3/2}$, therefore $V_A^{3/2} = \frac{I_A}{K}$ and $V_A^{1/2} = \frac{I_A^{1/3}}{K^{1/3}}$.

$$\therefore \quad V_A^{-1/2} = K_2 \cdot I_A^{-1/3},$$

where K_2 is a second new constant.

Hence $R_A = \frac{K_3}{\sqrt[3]{I_A}}$, where K_3 is constant, i.e. the A.C. resistance of a diode varies inversely with the cube root of the anode current. The greater the anode current the lower is the valve resistance.

Hence if a diode has an A.C. resistance of 10,000 Ω. when the anode current is 1 mA., then on raising the current to 8 mA. the resistance R_A will be decided by

$$\frac{R_A}{10{,}000} = \frac{1}{\sqrt[3]{8}} = \frac{1}{2}.$$

$$\therefore \quad R_A = 5000\ \Omega.$$

The Diode as a Rectifier. The chief function of a diode valve is as a rectifier of alternating current.

The alternating potential is put across the diode valve with the load resistance R in series. During positive half-cycles of the input, anode current will flow through R to build up a corresponding series of voltage half-wave pulses. During the negative half-cycles of the input no current will flow through the diode, the circuit will be open and no voltage pulses appear across R.

So the alternating potential input is transformed into a unidirectional pulsating D.C. potential across the load resistance.

This action is considered in the chart fig. 33*b*, by drawing the dynamic characteristic of the diode superimposed on which is a correctly orientated volts-time graph representing the input alternating potential. The dynamic characteristic is the characteristic of the valve with the load resistance in series. Such a characteristic will be straighter, and correspond to a higher total A.C. resistance (R_A+R) compared with the static characteristic for the valve alone. The improved linearity of the characteristic is often useful since it will mean less distortion of the wave-form of the output derived from the input.

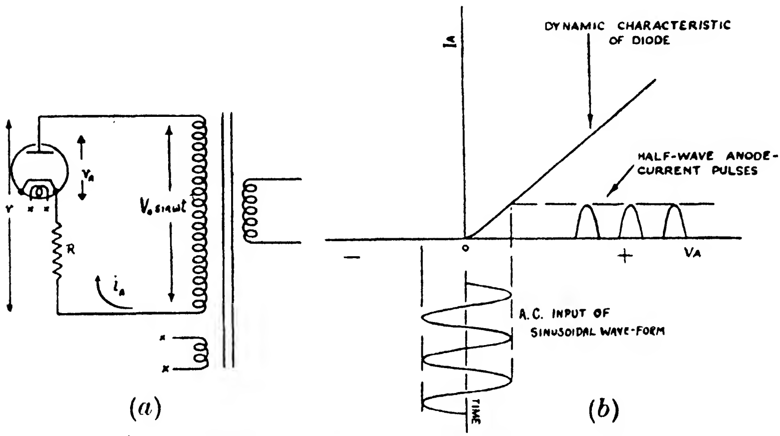


FIG. 33. *a*, The Simple Diode Rectifier. *b*, Action of the Diode Half-wave Rectifier.

In practice, it is usually required that the peak pulsating D.C. voltage across R be a considerable fraction of the peak input A.C. voltage. So R is generally greater than R_A.

Half-wave Rectifier. In fig. 33*a* let $V_0 \sin \omega t$ represent the input alternating potential. At any instant of time t during the action suppose v_A equal the voltage between anode and cathode of valve, and i_A be the corresponding anode current when v is the instantaneous total voltage applied across the valve and load resistance R in series.

Then $$v=V_0 \sin \omega t=v_A+i_A R. \qquad (150)$$

$$\therefore \quad V_0 \sin \omega t=i_A(R_A+R), \qquad (151)$$

since $v_A=i_A R_A$.

$$\therefore \quad i_A=\frac{V_0 \sin \omega t}{R_A+R}=I_A \sin \omega t, \qquad (152)$$

where I_A is the peak anode current.

$$\text{The efficiency of the rectifier} = \frac{\text{D.C. power output}}{\text{A.C. power input}} \times 100\%$$

$$= \frac{I^2_{D.C.} \cdot R}{I^2_{R.M.S.}(R_A + R)} \times 100\%. \qquad (153)$$

But $I_{D.C.}$=average value of half-wave rectified pulses of peak value $I_A = I_A/\pi$* (see p. 72).

The value $I_{R.M.S.}$=R.M.S. value of sinusoidal supply of peak value I_A which is $I_A/\sqrt{2}$. But the A.C. input only has to supply current, and hence power, for half the total input time since during the time when the diode anode is negative, the power drain on the input A.C. source is zero.

$$\therefore \quad \text{Input power is } \frac{1}{2}\left(\frac{I_A}{\sqrt{2}}\right)^2 . (R_A + R)$$

The efficiency of a half-wave rectifier is thus

$$\frac{(I_A/\pi)^2}{\frac{1}{2}(I_A/\sqrt{2})^2} \times \frac{100}{1+R_A/R} = \frac{400}{\pi^2} \times \frac{1}{1+R_A/R} = \frac{40 \cdot 6\%}{1+R_A/R}. \qquad (154)$$

The maximum possible efficiency is therefore 40·6%, when R_A/R tends to zero, i.e. the valve resistance is very low compared with the anode load resistance.

The maximum power output is obtained when the external resistance=internal resistance of supply.

In this case, therefore, $R = R_A$.

Then efficiency $= \frac{40 \cdot 6}{1+1}\% = 20 \cdot 3\%$.

Full-wave Rectification. By using a double-diode valve consisting of two separate diodes in the same envelope with the filaments, or cathode heaters wired in parallel, a type of circuit connection indicated in fig. 34*a* can be obtained whereby both positive and negative cycles of the input give rise to unidirectional pulses of current through the load. This is achieved by applying the A.C. input via a transformer with a centre-tapped secondary winding so that the inputs to the two valves are in anti-phase.

* Note that equation (124) p. 72, states that the average value of a half-wave rectified pulse is $2I/\pi$, but considering that the output current pulses flow for only half the time, and are separated by gaps where the current is zero, it is seen that the average current value in the half-wave case is actually I/π.

In this circuit note that when point A of the transformer secondary is at a positive potential with respect to centre-tap C, at earth potential, then point B is negative with respect to C by the same amount. When A is positive diode 1 conducts, whilst diode 2 does not since its anode is negative, so that electrons pass

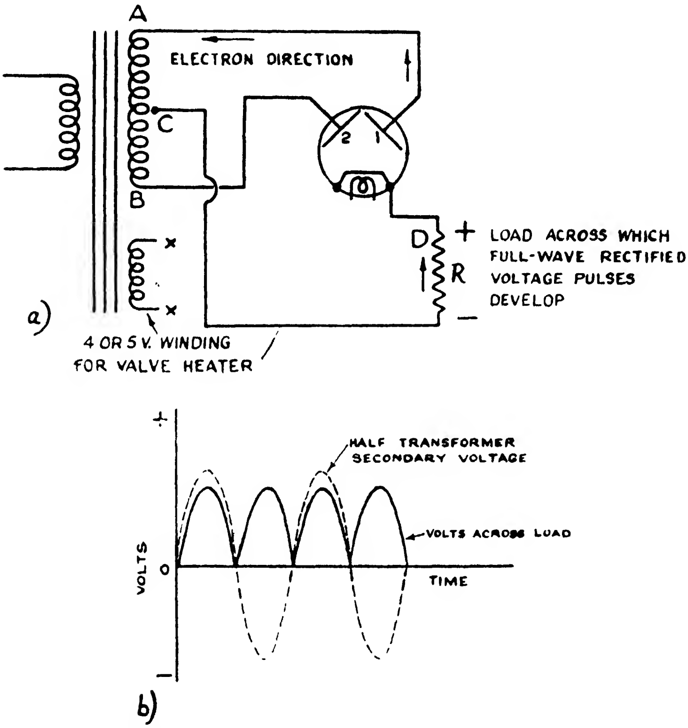


FIG. 34. *a*, Full-wave Rectification. *b*, Representation of Full-wave Rectification.

in direction from C to D through the load R. On the other hand, during the immediately succeeding half-cycle, A will be negative and B positive, so that diode 2 is conducting and diode 1 open-circuit, electrons are again drawn through R from C to D in the same direction as before.

From the chart representing the action of a full-wave rectifier it is readily seen that a pulsating D.C. output is obtained corresponding to both positive and negative cycles of the A.C. input. There are now no gaps occupying half a period between one output pulse and the next, as is the case in half-wave working.

Considering the efficiency of a full-wave rectifier in a similar manner to the way in which half-wave was discussed, comparison of the two methods shows:

(*a*) The average value of the D.C. output current is twice the value obtainable in the half-wave case, so equals $2I_A/\pi$;

(*b*) The A.C. input has to supply current during the whole input cycle, and not for only half the time of operation, since the two halves of the transformer secondary work alternately. So the input A.C. current has a R.M.S. value of $I_A/\sqrt{2}$.

Hence equation (154) becomes in the full-wave case

$$\frac{I^2_{D.C.}R}{I^2_{R.M.S.}(R_A+R)}\times 100\% = \frac{(2I_A/\pi)^2}{(I_A/\sqrt{2})^2}\times\frac{100}{(1+R_A/R)} = \frac{4/\pi^2}{\frac{1}{2}}\times\frac{100}{1+R_A/R}$$

$$= \frac{81\cdot 2}{1+R_A/R}\%.$$

The maximum efficiency when $R_A/R\to 0$ is therefore 81·2%, twice the maximum possible value in the half-wave circuit. Again, the maximum possible power-output when $R_A=R$ is obtained with an efficiency of $81\cdot 2/(1+1)=40\cdot 6\%$.

It is concluded therefore that assuming no transformer and valve heating losses, then the full-wave circuit has twice the possible power efficiency of the half-wave type.

The Gas-filled Diode. Considerations of the efficiency of rectifier circuits on the lines indicated above lead to the belief that the restriction on efficiency in practice is set by the necessarily finite A.C. resistance of the high vacuum diode. For this reason a gas-filled rectifier valve has been introduced which, owing to the ionisation of the gas by the emitted electrons, offers a very low resistance to current flow, and consequently lends itself to the design of circuits in which the maximum theoretical working efficiency is more nearly realised.

The hot-cathode mercury vapour rectifier is a much used type of valve in this category. This valve is able to give its full emission with an anode potential as low as 15 to 20 V.

The valve possesses a robust type of directly heated strip filament of nickel coated with the usual barium-strontium oxide mixture. The anode is placed near this filament, and has usually a top-cap connection. The anode-filament clearance is not at all

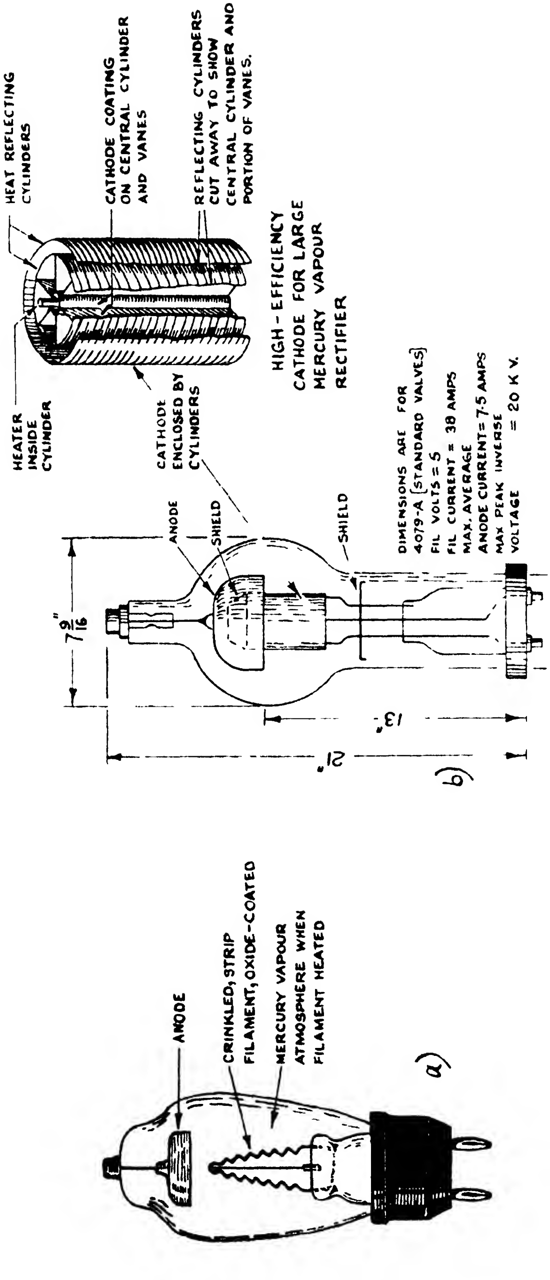


FIG. 35. *a*, The Mercury Vapour Rectifier. *b*, Mercury Vapour Rectifier of Large Current Capacity.

critical. After the valve has been pumped and activated a small pellet of mercury is introduced into the valve envelope, and remains there after sealing off from the pumps. This mercury is readily vaporised in the vacuum by the heating effect of the filament. Alternatively the inert gas argon is used.

Mercury vapour has an ionisation potential of 10·4 V. Consequently when the anode potential is raised to more than 10·4 V., the thermally emitted electrons drawn to it achieve sufficient energy to be able to ionise the mercury vapour atoms by collision. Both positive mercury ions, Hg^+ and Hg^{++}, and fresh electrons are produced.

With an anode potential of some 15 V. the collisions of electrons with mercury atoms are so numerous that a copious supply of mercury ions is available. The positive ions are $(600)^2$ times heavier than electrons, and consequently move 600 times more slowly in the electrostatic field. Whilst electrons are moving rapidly to the anode, these positive mercury atoms linger about, comparatively speaking, in the neighbourhood of the filament, where their positive charge neutralises the negative electron space-charge. The abolition of this space-charge permits practically the total number of emitted electrons to pass unobstructed to the anode, with a consequent flow of the saturation anode current with an anode voltage of only 15 to 20 V.

This desirably low resistance valve needs to be operated with some caution, otherwise its useful life will be seriously shortened. Precautions necessary are:

(*a*) The filament must be heated for some time, depending on the size of the rectifier, before the anode voltage is switched on. Otherwise the temperature within the bulb will be lower than the operating temperature, which is somewhat critical, and as a result the mercury vapour pressure will be too low to furnish an adequate supply of ionisable atoms within the space between cathode and anode. Consequently the internal resistance of the tube and the voltage drop across it will be high, so that positive mercury ions which are formed will speed to the cathode with such energy as to disintegrate the cathode coating. In a small type of rectifier, such as the GU 50, a preheating time for the filament of one minute is satisfactory. The larger types will need longer heating times, however, because of the greater thermal capacities of their cathodes, so that manufacturer's instructions regarding this precautionary time interval should be carefully observed. Using large rectifiers of this type it is customary to connect a time-delay switch in the anode circuit.

(*b*) A load resistance must be inserted in the anode circuit of sufficient value to prevent too high a voltage from being applied from the A.C. source to the rectifier anode, otherwise positive ions will again travel to the cathode with too high an energy for safety.

(*c*) The maximum inverse plate voltage which is specified for the valve, and which pertains during the negative, non-conducting half-cycle, must not be exceeded. This voltage value is somewhat lower than that which would obtain in the absence of the mercury vapour. An applied voltage in excess of this value will cause a spark-over between anode and cathode, destroying the valve.

(*d*) The bulb temperature must be maintained within definite limits. If it is too low then the mercury vapour pressure will be low, leading to the detrimental action indicated in (*a*). If the bulb is at too high a temperature, the vapour pressure of the mercury will be excessively high, so the maximum safe inverse peak voltage will be lowered.

Mercury vapour rectifiers with large current ratings have a cathode constructed in the form of a central, indirectly heated cylinder surrounded by vanes which are coated with the barium-strontium oxide mixture. A heat shield surrounds this arrangement, the outside of which is polished to prevent heat loss by radiation. This arrangement provides a large emitting area capable of supplying considerable electron currents, whilst the shielding conserves heat so well that the cathode efficiency, considered as amperes of emission current per unit area of cathode, is as much as five times greater than that obtainable from the open type of cathode necessary in the high vacuum valve. This arrangement is allowable since the positive ions penetrate into the remote pockets of the cathode, neutralising the space-charge, so that the full emission current is available to the anode, even though parts of the cathode surface may be in a plane at right angles to the anode plane. The anode is generally in the form of a cup which fits around the cathode, reducing the tendency to flash-back, and providing an efficiently electrically screened electron stream.

Though the mercury vapour rectifier is much more efficient than its high vacuum counterpart and consumes less heat power, yet it suffers from the disadvantages that damped high-frequency oscillations are likely to be generated within the electrons in the ionised gas, and damage to the cathode is more likely. This renders it unsuitable for use in radio receiver power supplies, but it is nevertheless frequently used in small transmitter power packs, and has also been developed for the large H.T. requirements of X-ray tubes.

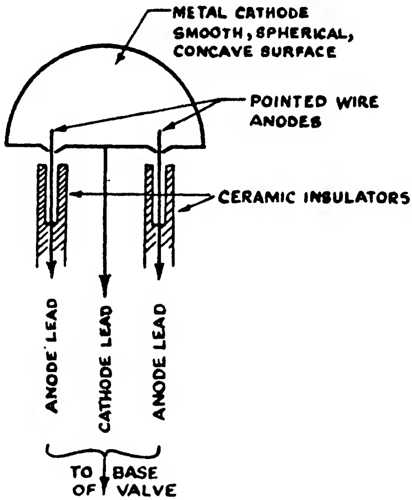


FIG. 36. The Cold-cathode Rectifier.

A gas filling of argon, or neon, may be used in place of mercury vapour in a gas-rectifier tube.

Cold-cathode Rectifier. If two electrodes situated in a gas are such that one has a plane or smooth surface whilst the other has a pointed structure, then when a P.D. is applied across them, the potential gradient immediately outside the sharply pointed electrode will be much greater than that outside the electrode with well-rounded contours. On the other hand, the current flowing through an ionised gas is a direct function of the area of the cathode surface, so that current will flow much more readily in the direction from the electrode of large surface area to the pointed electrode, than in the opposite direction.

A rectifier is constructed on this principle, in which a bowl-shaped cathode is used enclosing a pair of wire anodes. Such a tube can be used in the usual full-wave rectifier circuit, except that no filament circuit is necessary (fig. 36).

This type of valve is being applied to an ever-increasing extent for the supply of H.T. for radio receivers, and other small power electronic apparatus. It suffers from the disadvantage that the inverse peak voltage that can be applied without flash-over is small, being only a few hundred volts, so that this tube is not applicable to high voltage rectifier units. The gas pressure used is critical: too high a pressure makes the tube conduct much too readily in the inverse direction, when the anode is negative; too low a pressure will require too high an anode voltage to be established before conduction commences.

Smoothing Circuits. The output from a half-wave or full-wave rectifier circuit consists of a pulsating D.C. voltage, which is fluctuating from zero to a maximum value once in every cycle in the first case, and twice a cycle in the second case. Such big voltage variations would be intolerable for supplying H.T. to most types of electronic apparatus, so that some form of filter circuit is necessary to reduce these fluctuations to an insignificant ripple voltage. This is achieved by the use of filters of either the π-section or inverted L-section types, illustrated in fig. 37*a*.

Using the π-section filter, the condenser C_1 will charge for the time during which the rectifier is conducting, and discharge during the non-conducting period. Using C_1 only, as a reservoir condenser, the voltage fluctuation is reduced in the manner indicated in fig. 37*b*, provided C_1 has considerable capacity, and the rectifier current pulses are adequate to cope with the load current demanded.

This condenser is virtually a short-circuit across the valve during the conducting cycle, and therefore places a high demand on the valve. If this condenser capacity is made too large then the current pulses through the valve may be so excessive as to lead to premature cathode failure and overheating of the anode. This action restricts the use of such π-section filters to low power rectifiers, where they are preferred because of the higher output voltage they can supply. Such a π-section filter is out of the question for use with a mercury vapour diode, since the low internal resistance of this type of tube would involve such rapid supply of electric charge to the reservoir condenser that the large currents entailed would inevitably ruin the valve.

An L-section filter demands a continuous flow of current from the valve if the inductance is to counteract changes of current, and so bring about smoothing. It is therefore unsuitable for use in half-wave rectifier circuits.

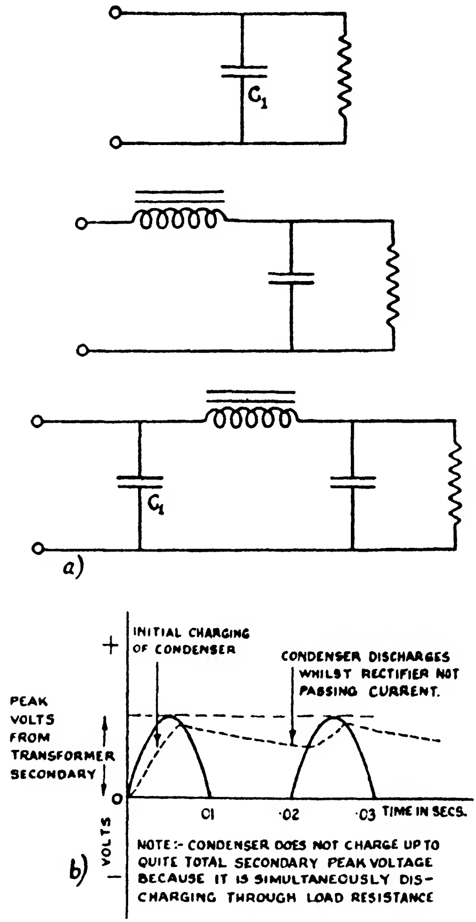


FIG. 37. *a*, Inverted L-section and π-section Filters. *b*, Smoothing Action of the Reservoir Condenser, C_1.

Design of Smoothing Filters. In the inverted L-section filter, the smoothing operation depends essentially on connecting across the rectifier output a choke and condenser in series. (See second of figs. 37*a*.) Since the pulsating D.C. supply to this arrangement consists of a wanted uniform D.C. component plus an A.C. component at the fundamental frequency, and harmonics of this frequency (p. 71), so the choke reactance is made high at the fundamental frequency compared with the reactance of the condenser. On the other hand, the D.C. resistance of the choke is kept as small as possible, whilst the D.C. resistance of the condenser is high (order of 0·1 MΩ. for electrolytic condensers, and practically infinity for a paper condenser). Therefore the choke-condenser combination acts as a potential-divider whereby most of the unwanted A.C. component appears across the inductance, whereas the required D.C. component is across the condenser and load resistance. If the load resistance is made too small, so as to be comparable in magnitude with the choke impedance, then the smoothing will be inadequate, since then some half of the A.C. component will appear across the condenser and load, and the rest across the choke. The smoothing will hence be good if the resistance load across the condenser is infinitely great, but will deteriorate as this load resistance is decreased.

In order to avoid over-complex mathematical analysis, simplifying assumptions are made which, nevertheless, enable useful design formulae to be proved. Suppose the A.C. voltage supplied is the rectified voltage output from the transformer, i.e. the voltage drop in the valve, and the drop in the reactance of the transformer secondary winding are neglected.

Then $$I_1=\frac{E_1}{\omega L}, \qquad (155)$$

where E_1=transformer A.C. voltage, R.M.S. value, I_1=A.C. component of the current through the filter, ω=pulsatance of A.C. supply. Frequencies at harmonics of ω will also be present, but if the design is adequate at the fundamental frequency, it will be more than adequate at harmonics of this value.

The reactance of the condenser C is kept so small compared with that of the choke, that it is neglected in this formula. However, the A.C. ripple voltage across the condenser, assuming the load resistance is infinite, will be E_R, where

$$E_R=\frac{I_1}{\omega C}. \qquad (156)$$

Substituting for I_1 from (155)

$$E_R = \frac{E_1}{\omega^2 LC}.$$

$$\therefore \quad LC = \frac{E_1}{\omega^2 E_R}. \qquad (157)$$

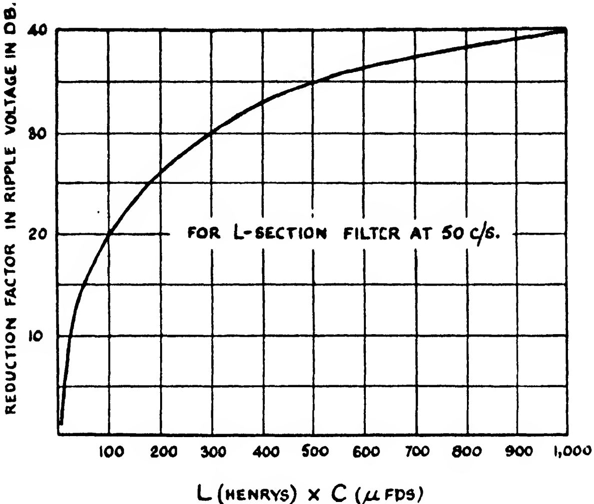


FIG. 38. Values of L and C for Inverted L-section Filter.

The minimum possible value of C, here called $C_{\text{min.}}$, is determined by the type of load supplied by the rectifier. Since a series inductance permits only slow changes in current through the filter, any rapid variations in load current will necessarily be supplied by the condenser. Also, if the current flow to the load is very small then the inductance will become ineffective in opposing current change, and hence in maintaining the current constant.

Let $I_{\text{min.}}$=minimum D.C. demand of load on rectifier. If the inductance is to act in smoothing the load current, then the alternating current peak value must never be less than the D.C. value $I_{\text{min.}}$, otherwise during the negative half-cycle* of the A.C.

* Note that the rectifier, of course, eliminates the negative half-cycle of the output from the transformer, but there will nevertheless be a negative half-cycle of the A.C. *component* of the pulsating D.C. output (see pp. 69 and 71).

component of the rectified output, the current will be temporarily at, or near, zero, and the inductance will then fail to act.

$$\therefore \quad \sqrt{(2)}I_1 = I_{\text{min.}} \quad \text{at limit.} \qquad (158)$$

$$\therefore \quad \frac{\sqrt{(2)}E_1}{\omega L} = I_{\text{min.}}$$

$$\therefore \quad L = \frac{\sqrt{(2)}E_1}{\omega I_{\text{min.}}} \qquad (159)$$

gives the minimum possible value for L.

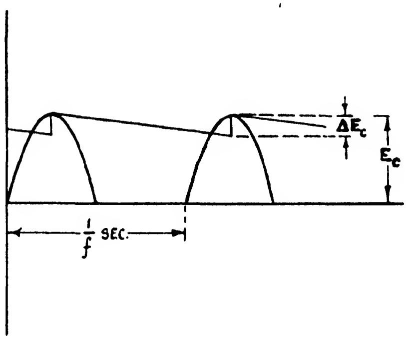


FIG. 39. Simplified Charge-Discharge Curve for Input Condenser of π-filter.

In the case of a π-section smoothing filter, see third of figs. 37*a*, consider the charge-discharge curve for the input condenser alone. The curve shown in fig. 37*b* can be simplified to that of fig. 39, where ΔE_C is the voltage change per cycle across the condenser for a half-wave rectifier circuit.

The change of condenser charge for a voltage change $\Delta E_C = (\Delta E_C)C$, where C=condenser capacity. This necessarily equals I_0T, where I_0 is the load current, and T is the time of one cycle, which equals $1/f$, where f is the frequency of the A.C.

$$\therefore \quad (\Delta E_C)C = I_0T = I_0\frac{1}{f}.$$

$$\therefore \quad \Delta E_C = \frac{I_0}{fC} = \frac{E_C}{RfC}.$$

Since $I_0 = \frac{E_C}{R} = \frac{\text{condenser voltage}}{\text{load resistance}}$, the percentage ripple voltage across the condenser and load is

$$100 \cdot \frac{\Delta E_C}{E_C} = \frac{100}{RfC} \qquad (160)$$

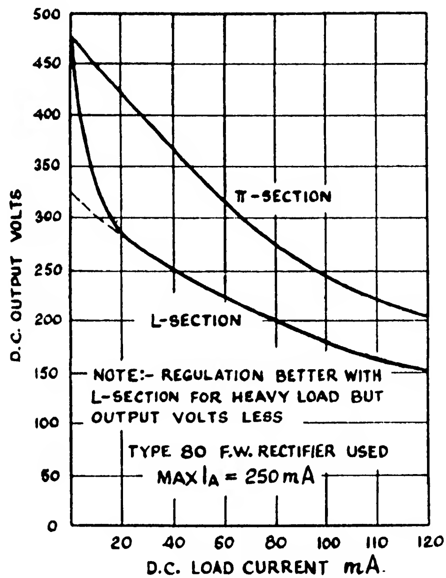


FIG. 40. Regulation of Rectifier Circuit with π- and Inverted L-section Filters.

If a full-wave rectifier circuit is used instead of a half-wave, then $T = 1/2f$, and % ripple $= 100/2RfC$, where f is the frequency of the A.C. supply. Hence with a given condenser capacity and load, the full-wave circuit produces half the ripple that the half-wave circuit does.

Evaluating a numerical example. Suppose a half-wave circuit supplies 300 V. D.C. with a load current of 60 mA. This is equivalent to a load resistance value of $300/0{\cdot}06 = 5000\ \Omega$. If the supply frequency is 50 c./s., and the condenser capacity is 16 μF., then

$$\text{Percentage ripple} = \frac{100}{RfC} = \frac{100}{5000 \times 50 \times 16 \times 10^{-6}} = \frac{100}{4} = 25\%.$$

If a full-wave circuit were used, the ripple would be only $12\frac{1}{2}\%$.

This comparatively large ripple voltage which would be obtained using a single condenser for smoothing will, using a π-section filter, be reduced by the following series and capacity in a similar manner to that occurring in an inverted L-section filter.

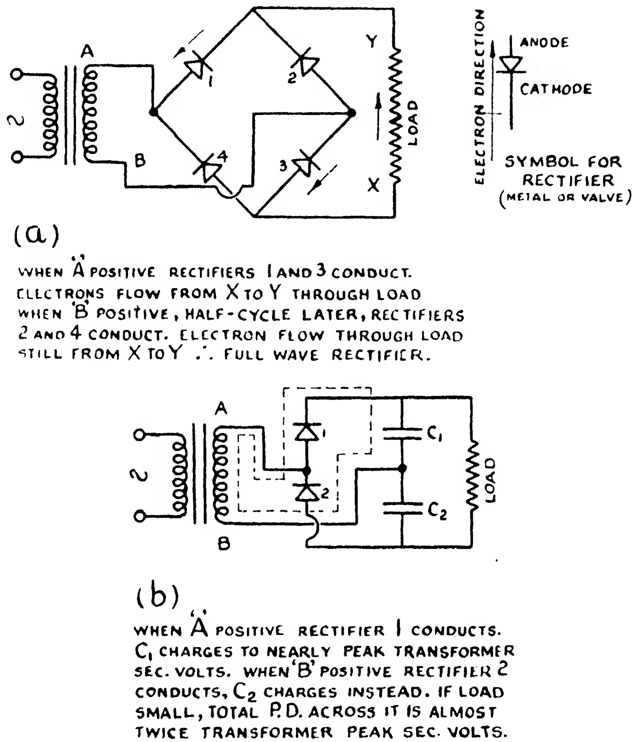


FIG. 41. *a*, Bridge Rectifier Circuit. *b*, Voltage Doubler Rectifier Circuit.

The effectiveness of L-section compared with π-section filters may be considered by reference to fig. 40. The voltage regulation, or maintenance of output voltage as the load current is increased, is better with L-section types except when the load current is very small, but the output voltage is lower. If a π-filter is used, and the load current is small, then the output D.C. voltage will approach the peak value of the applied A.C. voltage for small load currents.

The arrangement and working of bridge and voltage doubler rectifier circuits is considered in fig. 41.

The Triode Valve. In the three-electrode thermionic vacuum tube, or triode, a grid is inserted between the cathode and anode of a diode, near the cathode, giving the type of construction shown in fig. 42*a*. This grid has various forms, fig. 42*c*, but the flat-type of double-sided construction consisting of several turns of parallel molybdenum wires wound round nickel supporting rods is the commonest form.

The area of the spaces in this grid is usually much greater than the total area of the turns of wire, so that electrons find a ready path through the grid spaces to the anode. However, the number of these electrons which do penetrate the grid will depend on the grid potential relative to the cathode since the grid is situated in the electron space-charge region around the cathode.

If the grid is maintained at a positive D.C. potential then it will reduce the negative space-charge around the cathode, so that more electrons will be able to reach the anode per second; the anode current will rise. A negative potential on the grid will correspondingly increase the space-charge effect so that less electrons reach the anode per second; the anode current will fall. A sufficiently negative potential on the grid will prevent all the electrons emitted from going to the anode: the attractive effect of the positive anode is then completely offset by the negative grid potential. The negative grid potential necessary for this to occur is the *cut-off bias*. The greater the positive anode potential, the greater will be the necessary cut-off bias for a given triode valve.

The electrostatic effect of the grid potential is illustrated by considering the electric lines of force in a triode at various grid potentials, fig. 43.

Characteristic Curves for Triodes. The anode current in the triode will depend on two factors for a given amount of cathode emission: (1) the anode potential, (2) the grid potential. It is hence necessary to draw a family of curves to represent the behaviour of the valve. These can be either anode current vs. anode volts for various fixed values of grid bias, or anode current vs. grid volts for various fixed values of anode potential. The first set are called anode characteristics, the second set mutual characteristics.

A test circuit with which these characteristics can be obtained is shown in fig. 44*c*. The anode potential can be varied at will by

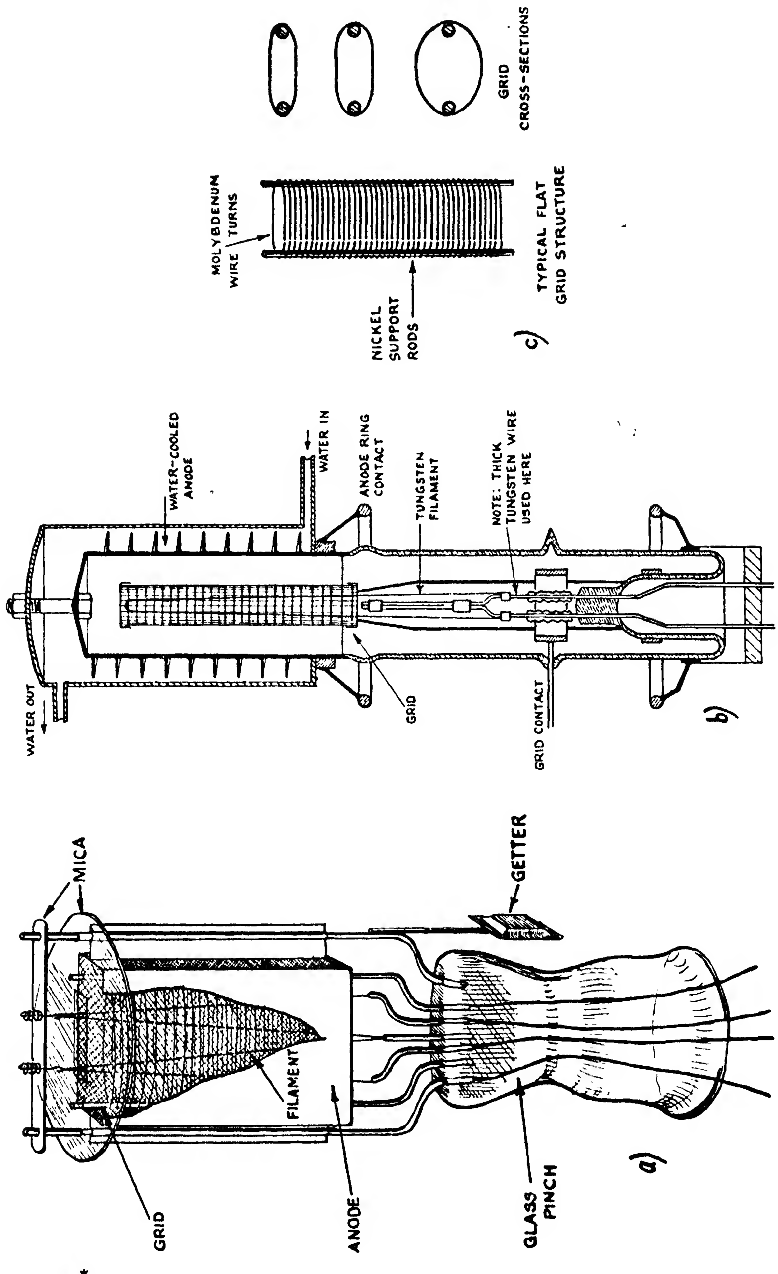


FIG. 42. The Triode Valve. *a*, Receiving Type. *b*, Transmitting Type. *c*, Grid Structure.

4*

the H.T. battery potentiometer. A grid bias battery potentiometer allows the grid potential to be set. Meters are provided for measuring anode voltage, grid voltage, series anode current and grid current.

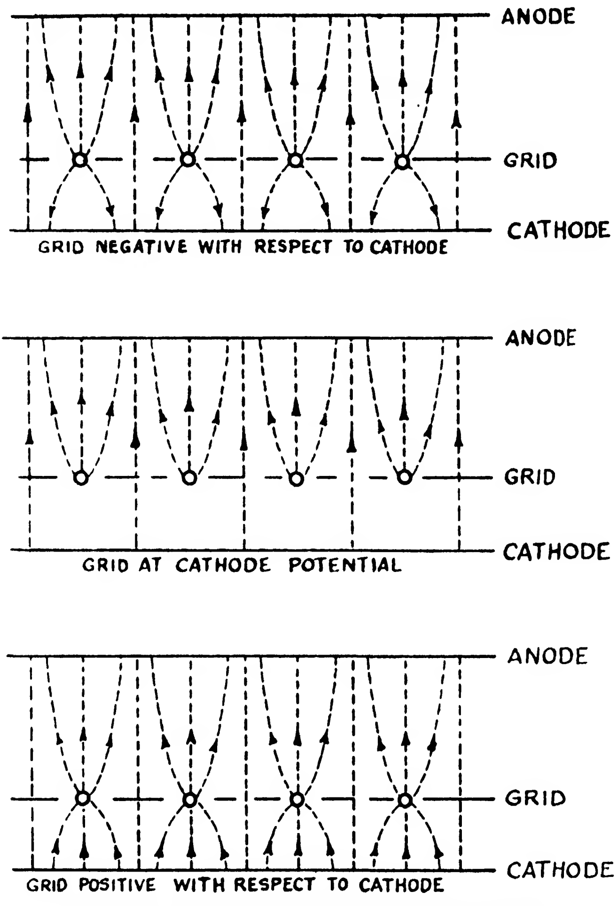


FIG. 43. The Action of the Grid in a Triode Valve.

The grid bias battery connection is such that either a positive or a negative potential can be impressed upon the grid.

Three Valve Constants. To define sufficiently the electrical action of a triode in a circuit three constants are used, and catalogued. These are:

(*a*) The *A.C. resistance*, defined as in the case of the diode but for a specified grid potential.

Values encountered in practical triodes range from 500 to 100,000 Ω.

(*b*) The *mutual conductance* (g_m) of the valve, defined as

$$\frac{\text{a change of anode current, } \partial I_A}{\text{small change of grid bias which produces the anode current change, } \partial V_G},$$

or the anode current change for one volt change of the grid volts. Specified at a given anode voltage, and mean grid bias. Quoted in milliamperes per volt (mA./V.), or amp./V.=mhos.

Symbolically, $g_m = \left(\frac{\partial I_A}{\partial V_G}\right)_{V_A}$.

Values encountered in small triodes range from 0·5 mA./V. to 3 mA./V., i.e. 0·0005 to 0·003 mhos.

(*c*) The *amplification factor* (μ), defined as

$$\frac{\text{the change in anode voltage required to produce a small anode current change}}{\text{the change in grid voltage required to produce the same anode current change}}$$

specified at mean anode and grid voltage values. Quoted simply as a ratio. The amplification factor is essentially a measure of the relative efficiencies of the grid and anode in changing the anode current. Since the grid is nearer the cathode its electrostatic influence on emitted electrons is much greater than the anode effect, so the amplification factors encountered in practice range from 3 to 100 for triodes.

In symbols, $\mu = \left(\frac{\partial V_A}{\partial V_G}\right)_{I_A}$.

The prime function of a triode valve is to amplify D.C. or A.C. voltages impressed on its grid relative to the cathode so as to produce a greater D.C. or A.C. voltage or power in the valve anode circuit. This action is considered in detail in the next chapter.

The Practical Determination of the Valve Constants. These constants are determined in practice using the circuit, fig. 44*c*. They may be evaluated directly, or from the characteristic

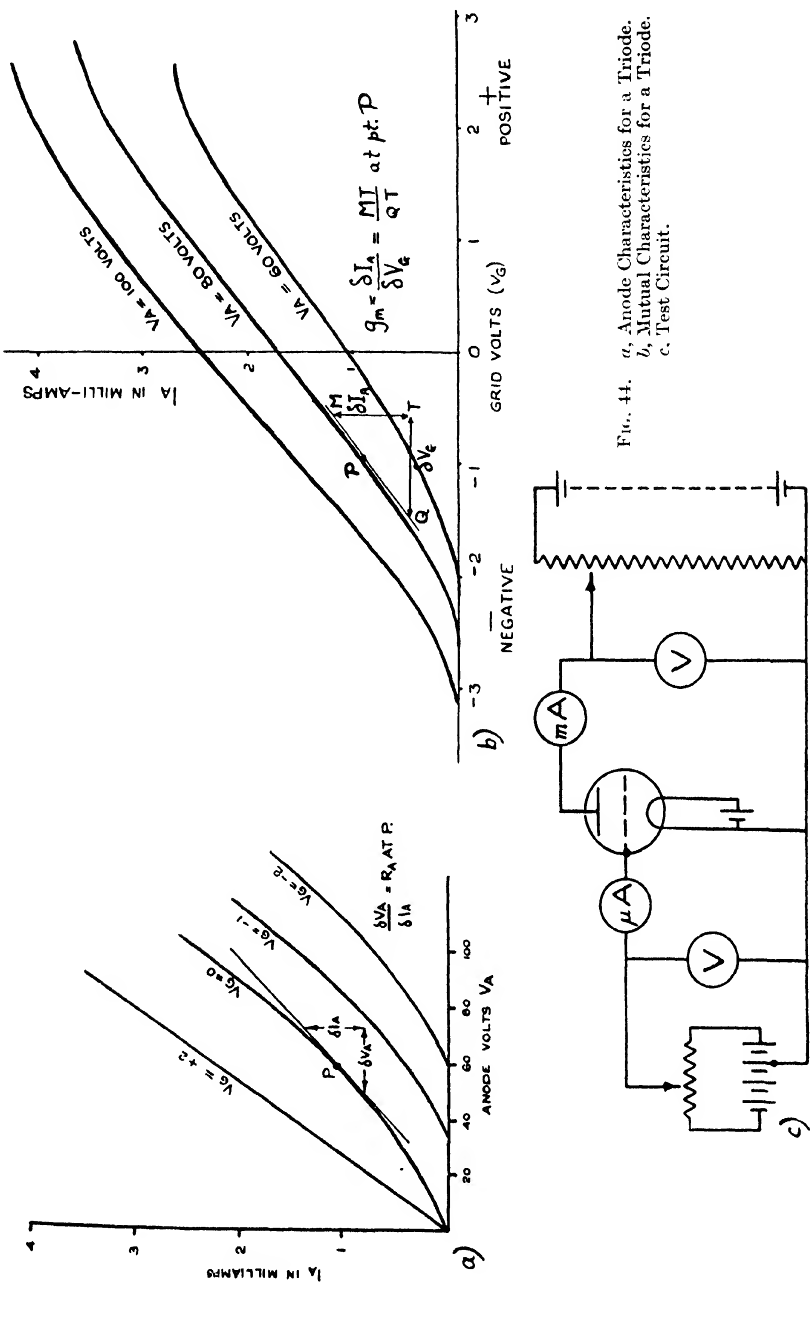


FIG. 44. *a*, Anode Characteristics for a Triode.
b, Mutual Characteristics for a Triode.
c, Test Circuit.

curves obtained using this circuit. Thus if the mutual conductance is required, then the anode voltage and grid bias are set at the values at which the constants are required, and the anode current variation in milliamperes obtained for a small grid volts change is recorded, a suitable small change being $\frac{1}{2}$ or 1 V. Thus if a change of bias from -2 to $-2 \cdot 5$ V. changes the anode current by 2 mA., then the mutual conductance ($g_m = \partial I_A / \partial V_G$) is 2 mA./$\frac{1}{2}$ = 4 mA./V.

Using the mutual characteristic, fig. 44*b*, then the mutual conductance will be the slope of the curve at the values of grid bias and anode potential specified. In this connection notice the change of slope, and hence change of mutual conductance, with grid potentials near the cut-off value; whereas in the region of grid volts equal to zero the characteristic is nearly straight, so that the mutual conductance is practically constant irrespective of the bias. The ratio of lengths TM/QT gives this slope at the point P on the graph. For a member of the family of mutual characteristics corresponding to a higher anode potential, the curves show the higher cut-off bias required, and also the increased slope or mutual conductance.

Approximately, the mutual conductance of a valve varies as the cube-root of anode current, i.e.

$$g_m \propto I_A^{1/3}. \qquad . \quad . \quad . \quad . \quad . \quad (161)$$

The specified grid bias and anode potentials at which the constants are determined define the *operating point*, which is determined by the conditions operating in the valve circuit (see below).

In a like manner, reference to fig. 44*a* will readily show how the anode impedance can be obtained from the anode ($I_A - V_A$) characteristics. Again a small change of anode volts, say 5 V., which produces a change of anode current, say 2 mA., will give an anode impedance of $5/0 \cdot 002 = 2500\ \Omega$ (note denominator representing current must be quoted in amperes if answer in ohms is required).

The amplification factor is determined, at a specified operating point, by first varying the grid potential by, say, 1 V., and recording the anode current change, then this potential is brought to its original value, and the anode potential is varied instead, and a measure of what anode potential variation is necessary to produce the same change of anode current as was produced by the grid potential variation is determined.

Since $g_m = \frac{\partial I_A}{\partial V_G}$, $\mu = \frac{\partial V_A}{\partial V_G}$ and $R_A = \frac{\partial V_A}{\partial I_A}$ then

$$g_m \times R_A = \frac{\partial I_A}{\partial V_G} \times \frac{\partial V_A}{\partial I_A} = \frac{\partial V_A}{\partial V_G} = \mu, \qquad (162)$$

so if two of the constants are determined the value of the third is known.

In the above practical determinations of valve constants it has been stressed that a small variation of the potentials should be made. A large variation is tolerable if it is known that in the region of such variation the characteristic is a straight line.

The Gas-filled Triode, or Thyratron. A robust type of grid construction is used in a triode valve operated, like the gas-filled rectifier, with a filling of mercury vapour or inert gas such as argon. The electrode spacings are not critical, but the design needs to take into account the high anode currents which are produced.

Action of a Gas-filled Triode. Consider such a triode in a test circuit as in fig. 45*b*, in which the anode and grid voltages can be varied. A protective resistance needs to be inserted in the anode circuit.

Suppose a negative D.C. potential, $-V$, is placed on the grid. Let $V.A$ be the anode voltage necessary to overcome the effect of this negative grid potential, where A is greater than 1. Suppose also that $V.A$ is larger than 10·4 V., the ionisation potential for mercury, or more than 15·4 V. if argon filling is used.

With the valve operating at grid bias $-V$, and anode volts just below $V.A$, no anode current will flow, since no electrons can pass the negative grid to the anode. If, however, a small positive D.C. potential is put in series with the existing supply to either the grid or the anode, then the cut-off effect of the grid will be relaxed, and some electrons will travel to the anode. With an anode potential greater than the ionisation potential of the gas concerned, these electrons will travel with such velocity, and have such kinetic energy as to be able to ionise the mercury or argon atoms by collision. A bluish-white glow will appear in the tube if mercury-filled, indicating that "striking" has occurred. The

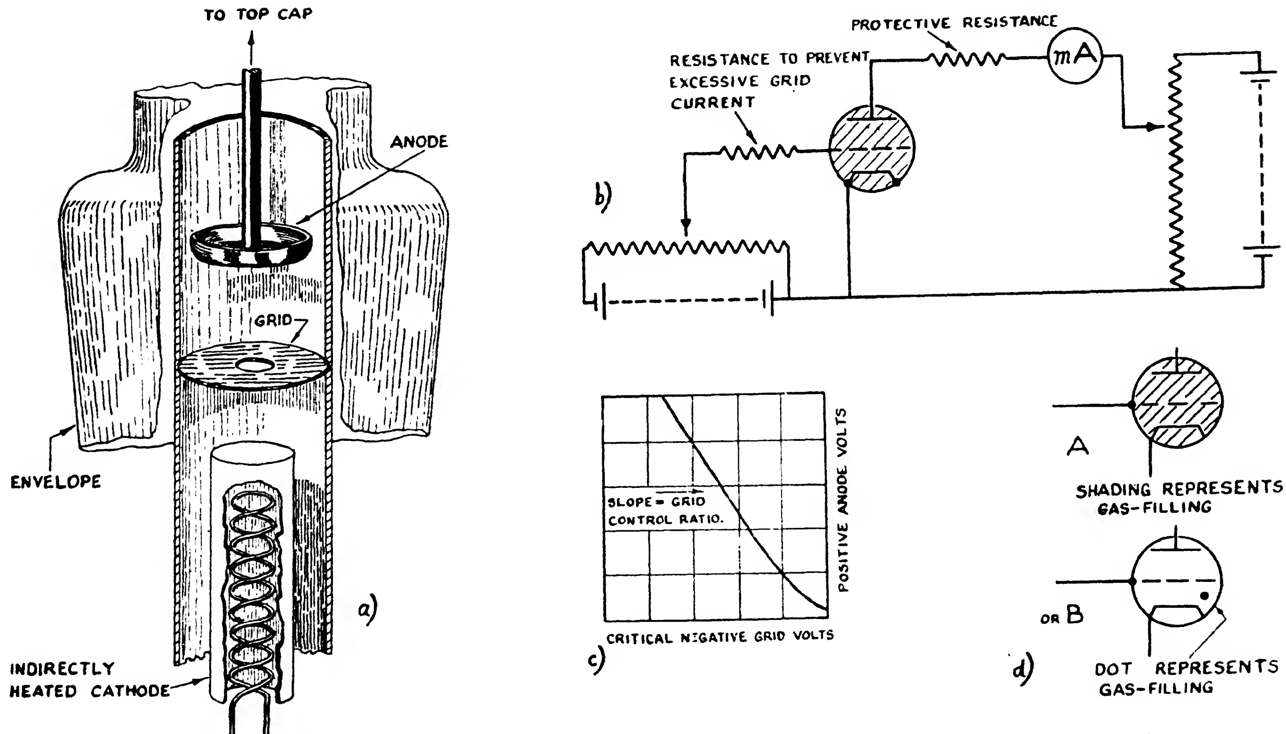


FIG. 45. *a*, Construction of a Gas-filled Triode. *b*, Circuit for obtaining Control Characteristics of Gas-filled Triode. *c*, Control Characteristics of Gas-filled Triode. *d*, Circuit Symbol.

heavy positive ions will neutralise the space-charge round the grid, as in the case of the mercury vapour rectifier, so that the grid, which influences the space-charge, is rendered ineffective. With anode potentials of some 20 V., the anode current will reach almost saturation value: i.e. a large flow of anode current capable of actuating a circuit requiring considerable power is effectively "triggered" by a small positive grid potential of which the power requirements are negligible. This valve therefore acts as a very effective relay.

Once the anode current is thus produced, and the grid becomes inoperative, the current can only be stopped by reducing the anode potential to less than the ionisation potential for the gas. Though the small additional voltage added at the grid effectively switches on the anode current flow, yet further potential variations impressed on the grid cannot switch the current off again: such cessation can only be produced by a sufficient reduction of the anode potential.

The gas-filled triode compares essentially with a relay of the electromagnet type. In the latter, a magnet closes the switch in a circuit carrying a large current, whereas the magnet is actuated by a much smaller current in a surrounding solenoid consisting of several turns of fine wire. The thyratron offers several advantages, however. Its action is very rapid, the time of successive operations being, in the end, limited by the finite recombination time of the ions, which is of the order of 0·0001 sec. The sensitivity and efficiency of the thyratron are very high, since it can be actuated by a fraction of a volt at practically zero power applied to the grid, whereas the anode power can be several watts.

Gas-filled triodes of various current ratings from 100 mA. upwards are available from the manufacturers. The upper limit of possible anode power is very high, particularly since grid-controlled mercury rectifiers controlling several thousand amperes in engineering practice can be considered as coming within this class of valve. Examples of commercial gas-triodes at opposite ends of the anode current rating scale are:

(1) The Osram GTIC has an indirectly heated cathode rated at 4 V. 1·3 amp., with a pre-heating time of 30 sec. It gives a peak anode current of 1 amp., and has a mean anode current rating of 300 mA. The anode potential can be 500 V. in the forward

or reverse cycles. The voltage drop across the valve is 16 V., since it is argon-filled. Its grid control ratio is 28 (see fig. 45*c*).

(2) The B.T.H. Co.'s BT 29 has an indirectly heated cathode operating at 5 V. 20 amp., with a pre-heating time of 5 min. This valve gives a peak anode current of 75 amp., with a mean rating of 12·5 amp. The maximum safe anode potentials forward and reverse are 2000 V. Being a mercury-filled valve it has an anode-cathode drop of 12 V. Its grid-control ratio is 70.

The cathode of a gas-triode is constructed considerably differently from that used in a high vacuum valve. Since the grid is inoperative once the tube has "fired", it is not necessary to preserve any particular geometrical relationship between the grid structure and the cathode. The cathode merely has to furnish a copious supply of electrons as efficiently as possible.

In design, therefore, prime attention is paid to conserving the cathode heat. Many types of heat-shielded cathode have been introduced to this end. Generally, a structure consisting of several nickel vanes spaced radially between concentric nickel cylinders, the whole of the inner part of the assembly being coated with the usual oxide mixture, is adopted in the case of gas-filled triodes with large anode current ratings. Further heat shields, polished on their inside surfaces to act as good heat reflectors, may surround such assemblies. Even in small valves, the use of such constructional techniques leads to an emission current per watt of cathode power ratio which is five to six times that obtainable in the necessarily open types of structure employed in vacuum valves, whilst in large gas-triodes, an improvement of cathode efficiency of as much as twenty times is obtainable by the careful consideration of the thermal efficiency of the cathode member.

The anode must not emit electrons, and must remain as cool as possible whilst furnishing large currents to the external circuit. An open type of structure, frequently made of carbonised nickel, is therefore adopted. The grid is usually of a cylindrical form surrounding the cathode, or may simply be a metal disk with a central aperture. It also is usually made of carbonised nickel (see fig. 45*a*).

Circuits Employing Gas-filled Relays. These gas-triodes find considerable application in remote control radio apparatus in which a transmitted voltage pulse can be received at the valve grid, and be made to switch on a considerable anode current. They can also be readily adapted to the construction of a cathode-ray tube time-base (see p. 267); to electronic switching circuits such as those used to control the time of actuation of a number of independent circuits, such as amplifiers; again they find much application in electrical engineering in the control of electric motors, rectifiers, inverters, and in the measurement of peak voltage, etc.

In the above introductory account of the thyratron it has been shown that, in a simple circuit, a priming voltage at the grid can cause the anode current to be switched on, but that to switch such current off again, it was necessary to reduce the anode potential to below the gas ionisation potential. This may be highly undesirable, particularly since switching at the grid circuit is easily accomplished because of the small currents necessary, whereas switching at the anode circuit usually requires the interruption of a large current, demanding heavier switch gear and slow operating time.

Two circuits are discussed below which avoid this difficulty. In the first of these, the grid and anode voltages are derived from an A.C. source, whereas in the second a two-valve circuit is employed.

In the circuit, fig. 46*a*, transformer secondary windings are so arranged as to provide the cathode heater current, the anode potential and the grid input for a gas-filled triode. The resistance R_G in the grid circuit is simply for the purpose of avoiding excessive grid current when the grid is driven positive. If the switch is at position 1 then the grid potential varies in accordance with the end *A* of the transformer secondary, whereas the anode potential varies with end *B*. Since the cathode is connected to this transformer winding at point *C* between *A* and *B*, so a positive increase of anode potential is inevitably accompanied by a negative increase of grid potential. If the point *C* is such that this negative grid potential is, at all times, greater than that necessary to bias the grid beyond cut-off as the anode potential increases, then the thyratron will not strike. Suppose now that the switch is put to position 2. Then the grid potential varies in phase with the anode variations, and the valve will "fire". The anode load will be actuated by the flow of anode current during the positive cycle of the anode potential alternations, but will cut off during the negative cycles. It will therefore receive a series of D.C. pulses of duration just less than half the A.C. period, and separated by time intervals approximately equal to this same time. Such a circuit is therefore switched on at position 2, and switched off by returning to position 1.

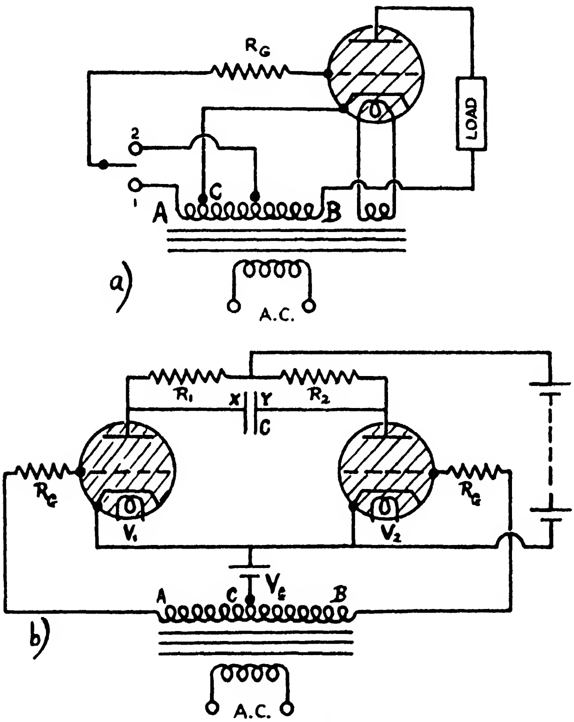


FIG. 46. *a*, Gas-filled Triode Circuit using A.C. Supply. *b*, A Gas-Triode Switching Circuit.

Two gas-filled triodes can be used in a circuit to provide a switching circuit capable of converting the steady D.C. power from an H.T. supply into an output of rectangular wave-form of which the period is controlled by an auxiliary A.C. supply (fig. 46*b*).

Initially, suppose that the fixed grid bias from battery V_G biases the grids of both valves to beyond cut-off. The H.T. supply is switched on, and then the A.C. supply to the transformer primary is applied. This transformer has a secondary which is centre-tapped at the point C, so that point A is negative when

B is positive with respect to C, and vice versa. When A is positive, the grid potential of tube V_1 rises to a maximum of $(V_0 - V_G)$, where $2V_0$ is the total peak potential developed across the transformer secondary AB. This relaxes the cut-off bias of V_1, so that V_1 passes almost instantly a large anode current. There is consequently a steady flow of current through R_1 for the whole of one half-cycle of the A.C. input. The corresponding P.D. built up across R_1 charges the condenser C through the resistance R_2 so that the plate X of C is negative, and the plate Y is positive.

During the subsequent half-cycle, the grid of V_1 is driven more negative than $-V_G$, whilst the grid of V_2 reaches a maximum potential of $V_0 - V_G$, relaxing its cut-off bias. As a result, the gas-triode V_2 fires, causing a pulse of current to pass through R_2. At the same time it allows that the plate Y of the condenser C can be connected to the common cathodes via the low resistance path of the switched-on gas triode V_2. Consequently, the potential across C is simultaneously put negatively on to the anode of V_1, reducing it to below the ionisation potential of the valve gas, so extinguishing the anode current in V_1. The nett result is that V_2 produces a steady current flow in R_2 during the period of the second half-cycle, whilst V_1 is shut off.

During the third half-cycle, the state of affairs pertaining during the first half-cycle is repeated. V_1 strikes, and this time V_2 is extinguished by the negative potential developed on its anode by the condenser C. It follows that for the whole of the A.C. input time there will be alternately developed across R_1 and R_2 rectangular-shaped pulses of current. Such D.C. voltages can be used to control the operation of auxiliary apparatus, e.g. two amplifiers feeding a cathode-ray tube, so that the inputs to the amplifiers can alternately be recorded on the tube screen.

An extension of the above methods can lead to the firing in sequence of a number of gas-filled triodes connected in a circuit, giving a number of switched voltages to independent positions.

Design Data for Vacuum Triode Valves. The relationship between the anode current I_A, the anode voltage V_A and the grid voltage V_G can be represented by the equation

$$I_A = K(V_A + \mu V_G)^{3/2} \qquad . \quad . \quad . \quad . \quad (163)$$

derived from the diode formula (146), where V_A is increased by an amount μV_G, since the grid voltage is μ times as effective as the anode voltage in changing the anode current, where μ is the amplification factor.

Put $V_A+\mu V_G=0$ in this formula, then $I_A=0$. Hence the anode current is reduced to zero at a cut-off bias given by $V_G=-V_A/\mu$. This evaluation of the cut-off bias is only approximately correct because of the "tail" effect evinced when the grid is near the cut-off value: the $I_A:V_G$ characteristic is curved near cut-off because the negative retarding electric field set up by the grid at the cathode is not uniform over the entire cathode surface.

Formulae can be quoted for the value of μ based on the geometry of the electrodes. Two important cases are (*a*) plane-parallel geometry of cathode, grid and anode, and (*b*) a concentric cylindrical arrangement (fig. 47).

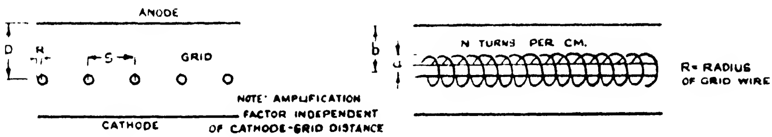


FIG. 47. *a*, Plane-Parallel Triode Geometry. *b*, Concentric-cylinder Geometry.

(*a*) Plane-parallel geometry. Using the dimensions indicated in fig. 47*a*, then

$$\mu=\frac{2\pi D/S-\log_e \cosh (2\pi R/S)}{\log_e \coth (2\pi R/S)} \qquad . \quad . \quad (164)$$

or

$$\mu=\frac{2\pi nD}{\log_e (1/2\pi nS)} \qquad . \quad . \quad . \quad (165)$$

is a simpler formula, where R, the grid wire diameter, is ignored compared with S, the space between grid wires, and n is the number of grid turns per cm.

(*b*) Concentric-cylinder geometry. Using the dimensions shown in fig. 47*b*, then

$$\mu=\frac{2\pi aN \log b/a}{\log 1/2\pi NR}.$$

CHAPTER 6

Valve Amplifiers

D.C. Amplifier. Suppose some such device as a photocell, thermocouple or glass-electrode for pH measurements produces a small direct current of the order of 1 μA., which it is required to amplify so that it can be registered on a milliammeter, or made to actuate a circuit requiring greater power than the device itself yields. A basic circuit arrangement whereby this can be arranged is indicated in fig. 48, where a triode valve is used as a D.C. amplifier.

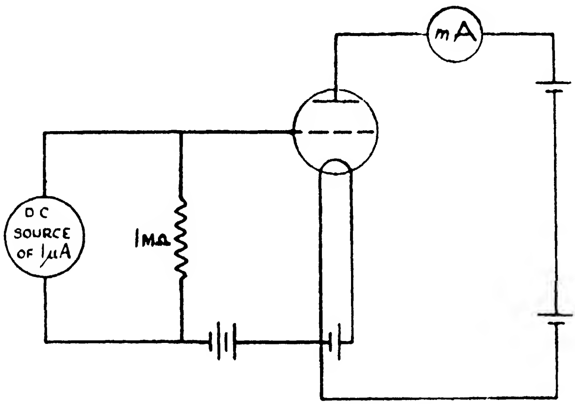


FIG. 48. Basic D.C. Amplifier Circuit.

The small current available is caused to pass through a resistance of, say, 1 MΩ. Thus a current of 1 μA. passed through 1 MΩ. will produce a P.D. of 1 V. This P.D. is placed across the grid-cathode of a triode valve in series with a negative grid bias. The negative grid bias is large enough to prevent the grid going positive, and so drawing grid current on application of the D.C. input. The input resistance of the triode is therefore high, of the order of 5 to 8 MΩ. for an ordinary triode valve. Positive ion current and electrical leakage prevents the input resistance of a normal valve from being higher, though these disadvantages can be overcome using a special type of electrometer triode.

The great virtue of the triode valve as a D.C. amplifier lies in its high input resistance, since it enables an E.M.F. input of high internal resistance and low current to be applied without damping, i.e. the high resistance is not reduced by a lower resistance due to the valve being put in parallel with it. Even so an input E.M.F. source of resistance greater than 2 MΩ. cannot be effectively employed using a standard triode in a normal circuit.

If the triode of the circuit has a mutual conductance of, say, 2 mA./V. at the anode potential used, then the anode current is changed, by the 1 V. P.D. applied to the grid, by 2 mA. Hence a device giving a current of only 1 μA. has given rise to an anode current change of 2 mA. This is a D.C. current amplification of 2 mA./1 μA., or 2000 times. Likewise, if the H.T. battery potential is raised so as to accommodate in the valve anode circuit a series resistance of 50 kΩ., and yet achieve the same anode potential as before, despite the D.C. volts drop in this anode load, then the change of 2 mA. across 50 kΩ. produces a voltage change across the load of $2/1000 \times 50{,}000 = 100$ V. A grid potential change of 1 V. has then given rise to an anode potential change of 100 V.; a D.C. voltage amplification effect of 100 times is achieved.

Alternating Voltage Amplifier. Class A. A common requirement, especially in radio receivers, is that a small alternating potential should be amplified, without distortion of wave-form, to give rise to a larger alternating potential. This is achieved using a triode Class A amplifier circuit.

The circuit of fig. 49*a* is used. An anode load resistance R is essential in order that a voltage variation of the anode is produced by the variation of anode current due to the grid potential changes. The valve mutual characteristic indicated in fig. 49*b* hence has to be the so-called dynamic characteristic, which is obtained exactly as the static characteristic, but with the necessary anode load resistance in the circuit. This load resistance will reduce the mutual conductance g_m of the valve to a lower value g_m', and, moreover, the dynamic mutual characteristic will be more linear than the static curve. If the grid potential is increased positively, the corresponding increase of anode current causes an increased volts drop across the anode load resistance. Since the actual anode potential is the H.T. potential minus the voltage drop across the anode load, so the anode potential falls. As the grid volts are increased so the anode volts are decreased, and, vice versa, a decrease of grid volts causes an increase of anode volts.

If the grid voltage is alternating, then the anode voltage, with an anode load resistance in circuit, alternates in anti-phase. Obviously the change of 1 V. on the grid will produce an anode voltage change which affects the anode current in the reverse direction to the effect of the grid volts change. In other words the static mutual conductance is reduced to the dynamic mutual conductance.

In operating the amplifier without distortion, Class A conditions are observed:

(1) The straight portion only of the dynamic mutual characteristic is utilised, i.e. the input alternating potential to the grid must produce grid voltage changes which are such that the anode current variations are in direct proportion.

(2) The grid must not be driven positive, otherwise grid current will flow, and power be drawn from the input source, causing damping.

Thus if the input voltage has a R.M.S. value of 3 V., then the positive peak input volts is $3\sqrt{2}=4{\cdot}2$ V. The negative steady grid bias used must therefore be at least $-4{\cdot}2$ V., so that the alternating plus direct inputs in series do not give rise to a positive grid.

In accordance with requirements (1) and (2), the sinusoidal wave-form representing the input to the circuit of fig. 49*a* is accommodated under the negative straight portion of the dynamic characteristic of the chart, fig. 49*b*.

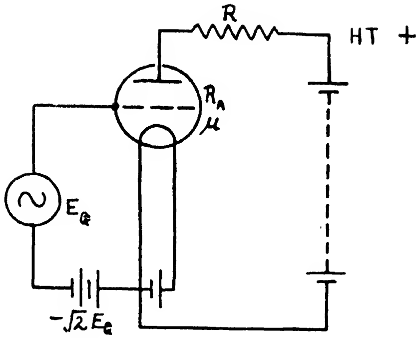
If the R.M.S. input voltage $=E_G$, then the R.M.S. anode current variation produced in series with the steady anode current is $g_m' \,.\, E_G$, where g_m' is the dynamic mutual conductance of the valve. Consequently, the R.M.S. value of the alternating voltage across the anode load R, is $g_m' \,.\, E_G R$.

The voltage amplification factor (V.A.F.) or stage-gain (m) of the circuit is defined as

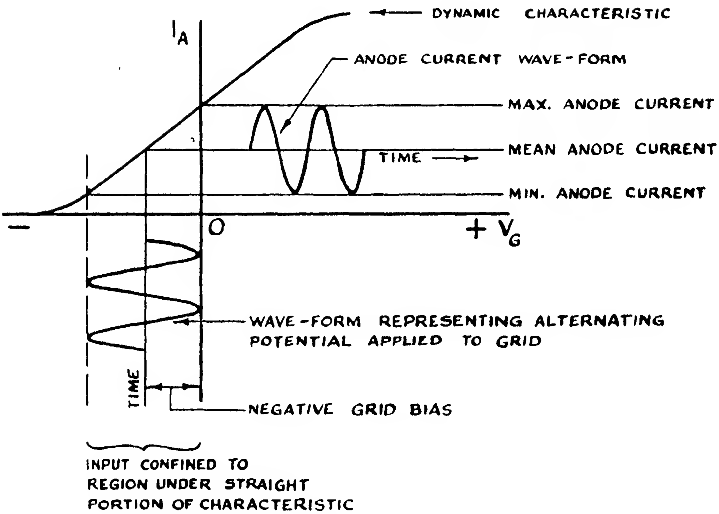
$$\frac{\text{Output R.M.S. voltage across anode load}}{\text{Input R.M.S. voltage to grid}}.$$

$$\therefore \quad m=g_m'{}_R.$$

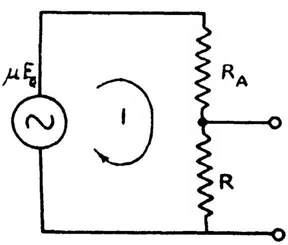
This is not the most convenient stage-gain formula to use in practice, since g_m' is not known unless it is obtained by plotting



(*a*) CLASS A VOLTAGE AMPLIFIER



(b) CHART ILLUSTRATING CLASS A AMPLIFICATION



(c) CIRCUIT ELECTRICALLY EQUIVALENT TO CLASS A AMPLIFIER

FIG. 49. *a*, Class A Voltage Amplifier. *b*, Chart illustrating Class A Amplification. *c*, Circuit Electrically Equivalent to Class A Amplifier.

the mutual characteristic for the valve with the anode load to be used.

Stage-Gain Formula. A circuit which is electrically equivalent to that of fig. 49*a* is shown in fig. 49*c*.

The valve is dispensed with in this circuit on the assumption that the change of potential E_G on the grid is equivalent to an anode voltage change of μE_G, by definition of the amplification factor, μ. Across μE_G there is effectively the anode load R, and the A.C. resistance of the valve R_A, in series. Since linear operation of the circuit is pre-arranged, so Ohm's law is applicable, and the R.M.S. value of the series current in the circuit is

$$I=\mu E_G/(R+R_A).$$

Correspondingly, there appears across the anode load R, a voltage variation of R.M.S. value given by IR, equal to $\mu E_G \,.\, R/(R+R_A)$.

The gain m is therefore

$$\frac{\mu E_G \,.\, R}{R+R_A} \div E_G = \frac{\mu R}{R+R_A} \quad . \quad . \quad . \quad (166)$$

Selection of Anode Load Resistance Value. In fig. 50 a curve is plotted of gain vs. anode load resistance R for a particular valve.

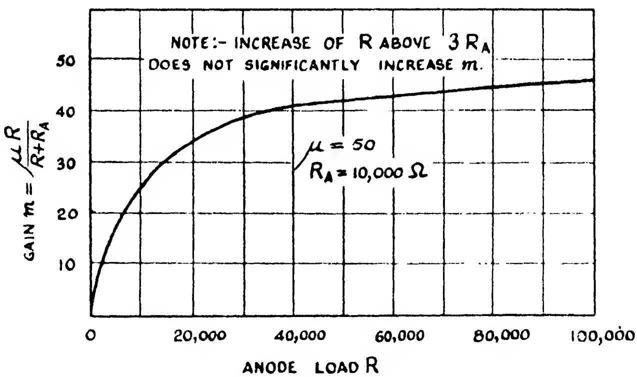


FIG. 50. Graph of Gain vs. Anode Load R, for Class A Amplifier.

As R is increased, so the gain increases to a maximum value of μ, the valve amplification factor. It would therefore seem desirable to use as high a value for R as possible. Against this consideration, however, it must be realised that the presence of a great value for R will introduce an excessively large D.C. volts drop between the H.T. supply and the valve anode. This voltage drop is inevitable using a resistance load because of the necessary direct anode current through the valve. In practice, therefore, a compromise is adopted whereby to obtain the maximum output alternating voltage without using an excessive H.T. supply, the anode load resistance is chosen to be three to four times the valve A.C. resistance. From the curve of fig. 50 it can be readily seen that this does not produce any serious loss of gain. In such circumstances the valve steady anode potential will be about one-third of the full H.T. supply volts.

Multi-stage D.C. Amplifiers. The circuit of fig. 48 can be extended to obtain another stage of amplification by using a second valve to amplify the output of the first. Thus a direct current or voltage amplification of, say, 100 times can be obtained from each stage, giving a total amplification of 100^2, or 10,000 times. Indeed, more stages can be added so that a very small D.C. voltage input to the first valve can be magnified by as much as a million times: an input of 1 μV. can give rise to an output of 1 V. Theoretically, there is no limit to the total amplification obtainable in this manner, but a practical limit is set by "shot effect", thermal agitation "noise", and fluctuations of H.T. supply voltages (see pp. 164 and 166).

A circuit showing D.C. amplifying stages connected in cascade is shown in fig. 51*a*.

A small input D.C. voltage applied to valve V_1 grid gives rise to a larger voltage across the anode load R_1, between the points A and B. The point A is connected to the cathode of valve V_2 via the H.T. battery of voltage E_B; the point B is connected to the grid of V_2 via a grid bias battery of voltage V_{G2}. The total D.C. potential appearing across the grid to cathode of V_2 is therefore: $E_B - I_{A1}R_1 - V_{G2}$, where I_{A1} is the anode current of valve V_1.

If the valve V_2 is to operate without grid current flow, then this voltage must be zero, or somewhat negative, so the grid bias battery supplying V_{G2} must have a potential of at least $-(E_B - I_{A1}R_1)$.

The chief disadvantage of such a method of connection is, therefore, the large grid bias battery required for the second stage. Again, this amplifier exhibits a tendency to "drift", which is more pronounced the greater the amplification attempted. This drift

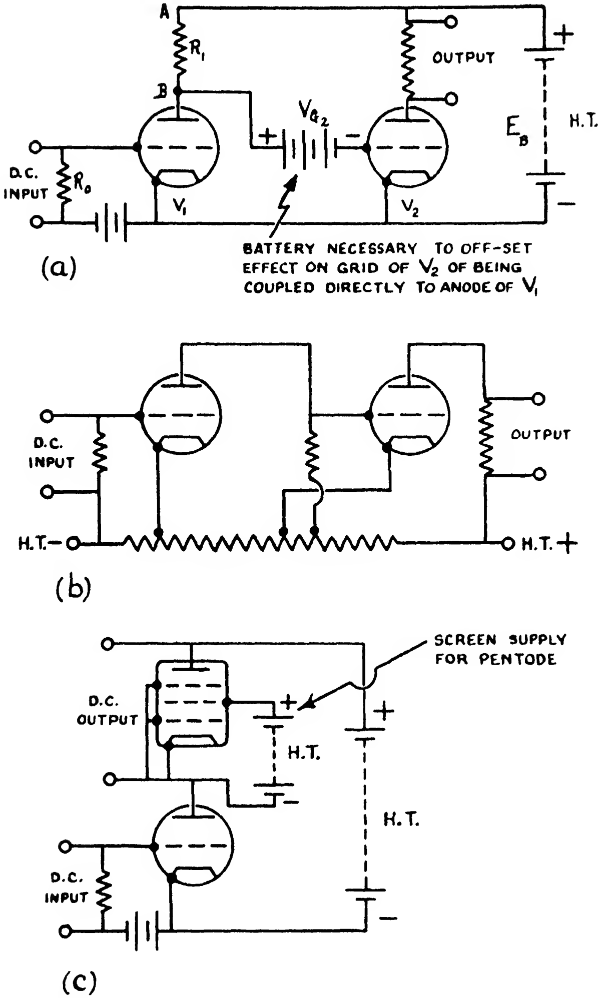


FIG. 51. *a*, Two-stage D.C. Amplifier. *b*, D.C. Amplifier of Loftin and White. *c*, D.C. Amplifier due to Horton.

is particularly prevalent since any D.C. supply potential change, anywhere in the circuit, causes the potentials in all succeeding stages to vary.

As the input across R_0 to valve V_1 increases positively, so the anode current I_{A1} increases, and the potential $(E_B - I_{A1}R_1)$

decreases, causing a drop of the anode current I_{A2} of valve V_2. If more stages are added, each succeeding valve works alternately; if the first, third and other odd stages undergo an anode current increase, then the second, fourth and other even stages suffer an anode current decrease.

Loftin and White* have developed a D.C. amplifier circuit operated by single H.T. supply of E.M.F. equal to the sum of all the grid and anode voltages. Such amplifiers can also be used for A.C. operation (fig. 51*b*).

Horton† introduced an ingenious D.C. amplifier circuit in which the high A.C. resistance of a pentode valve (see p. 139) is used as the anode load instead of the usual resistance. This provides an effective anode load resistance of one to two megohms, yet without introducing an excessive D.C. volts drop between the main H.T. supply and the anode of the amplifier valve proper (fig. 51*c*).

Multi-stage Low-frequency A.C. Amplifiers. The most widely used amplifier for the magnification of alternating voltages or currents at frequencies less than 20 kc./s. is the resistance-capacity coupled amplifier. Alternative circuits which give a greater overall amplification, but with more distortion, are the choke-capacity coupled arrangement, and the transformer coupled amplifiers of the series-fed and parallel-fed types.

The Resistance-Capacity Coupled Amplifier (R.C.C.). The discussion on p. 120 shows that the alternating voltage input to the grid of valve V_1 appears in amplified form across the anode load resistance R_1. The circuit of fig. 52 enables this voltage to be amplified again by making it the input to valve V_2. The method of connection, and a calculation of the amplification obtainable, can be appreciated by considering the equivalent electrical circuit, fig. 52*b*.

Let E_G be the R.M.S. alternating voltage input to V_1 grid, and m be the gain furnished by the first stage. The alternating voltage component across the load resistance R_1 will then be mE_G.

This alternating voltage mE_G is, in effect, supplying a circuit consisting of condenser C_G and resistance R_G in series. The fraction of this voltage appearing across R_G is applied to valve V_2 grid. The condenser C_G is essential to avoid the positive anode voltage of valve V_1 being superimposed on the grid of valve V_2.

* E. H. Loftin and S. White, *Proc. Inst. Rad. Eng.*, **16**, 281, 1928.
† J. W. Horton, *Jour. Frank. Inst.*, **216**, Dec. 1933.

A decision as to the component values C_G and R_G that should be adopted demands consideration of the following factors:

(*a*) As large a fraction as possible of the alternating voltage across R_1 needs to appear across R_G if the utmost possible voltage amplification is to be achieved. Since R_G and C_G in series form an

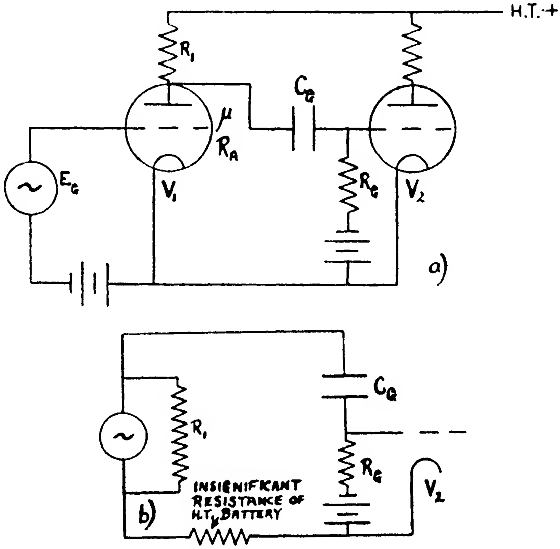


FIG. 52. *a*, Resistance-Capacity Coupled Amplifier. *b*, Electrically Equivalent Circuit.

A.C. potential divider across the supply of voltage mE_G from R_1, so the reactance of the condenser C_G needs to be small compared with the value of the resistance R_G. From the formula on p. 51, the voltage across R_G is

$$\frac{m \,.\, E_G R_G}{\sqrt{[R_G{}^2 + (1/\omega^2 C_G{}^2)]}},$$

where $\omega = 2\pi f$, f being the frequency of the alternating voltage input.

If this expression is to be as nearly as possible equal to mE_G, then $1/\omega C_G$ must be small compared with R_G, necessitating a large value of the coupling condenser C_G.

(*b*) The effective anode load resistance of valve V_1 is not R_1 alone, but less than R_1 because of the presence of the parallel circuit formed by R_G and C_G in series. If C_G is made sufficiently large in accordance with (*a*), then, to a first approximation, the anode load of V_1 may be considered as R_1 and R_G in parallel, equivalent to a resistance of $R_1R_G/(R_1+R_G)$. So m, the gain of the first stage will not be $\frac{\mu R_1}{R_1+R_A}$ but $\frac{\mu[R_1R_G/(R_1+R_G)]}{[R_1R_G/(R_1+R_G)]+R_A}$. In order that this actual value of gain should be as nearly as possible equal to the full gain, R_G must be large compared with R_1. A value of R_G of five to ten times R_1 is usually chosen.

(*c*) It would seem that the ideal alternating voltage amplifier is obtainable by using values of C_G and R_G as large as possible. However, there are practical limitations to the maximum values of these components which can be used. If R_G is made excessively large, say 10 MΩ., then any positive ion current,* say 1 μA., collected by the negative grid of valve V_2 will pass through the resistance R_G, producing a D.C. potential difference across it. In the case of the figures quoted, this P.D. will be 10 MΩ. × 1 μA. = 10 V., making the grid positive to this extent. Such a potential will usually overcome the effect of the negative potential brought about by the grid bias battery, so the resultant grid potential will be positive. The grid will then collect electrons; the effect will be cumulative, and excessive grid current will flow. This grid current will form a resistance in parallel with the high value of R_G selected, reducing its effective value considerably. In certain cases, the positive grid current may be so great as to overheat the grid inordinately. So values of R_G used with normal valve types rarely exceed 2 MΩ.

The value of C_G can be made large, provided that a large capacity condenser of good insulation resistance is available. If, however, as in radio communication receiver amplifiers, the audio-frequency voltage input to the amplifier varies over a frequency-range of 50 c./s. to 10,000 c./s., and also varies greatly in magnitude, then a transient powerful signal may temporarily drive the grid of valve V_2 positive, causing grid current to charge the condenser C_G. To prevent the condenser charge blocking the grid potential, the time constant C_GR_G must be sufficiently short to enable the condenser to leak away rapidly: a value of 0·005 is recommended,

* Such positive ion current is due to the ionisation of the residual gas in the valve by the electrons accelerated away from the cathode. If this ion current, as collected by a negative grid, exceeds 1 μA., then the vacuum in the valve is not good enough.

so that with $R_G = 0{\cdot}5$ MΩ., $C_G = 0{\cdot}01$ μF. An R.C. coupled amplifier operating with such component values will exhibit a reduced amplification at the lower frequencies because of the increased value of $X_{C_G} = 1/2\pi f C_G$ as f is reduced.

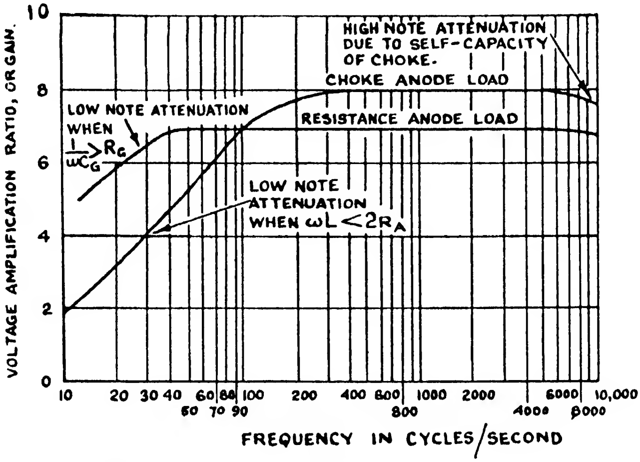


FIG. 53. Voltage-amplification vs. Signal-frequency Characteristic for R.C.C. Amplifier, and C.C.C. Amplifier.

The magnification of the first stage of a resistance-capacity coupled amplifier may be conveniently expressed by the formula:

$$\frac{E_{G2}}{E_{G1}} = \frac{\mu[R_1R_G/(R_1+R_G)]}{[R_1R_G/(R_1+R_G)]+R_A} \cdot \frac{R_G}{\sqrt{[R_G{}^2+(1/\omega^2C_G{}^2)]}}, \qquad (167)$$

where E_{G2} is the input alternating voltage to the grid of valve V_2, and E_{G1} is the input to valve V_1. In this formula, the effect of condenser C_G is ignored in calculating the effective anode load of valve V_1.

Choke-capacity Coupling. A choke can be used as the anode-load of valve V_1, in place of the resistance; otherwise the circuit is the same as the R.C.C. amplifier. The advantage is that the choke can have a high reactance to A.C. but a negligible resistance to D.C., and can therefore act as a high value load to produce a high stage-gain, yet not introduce any significant D.C. voltage-drop between the H.T. supply and the valve anode. If the low D.C. resistance of the choke is ignored, then the gain for the first stage will be $\mu\omega L/\sqrt{(\omega^2L^2+R_A{}^2)}$ (see fig. 54*a*).

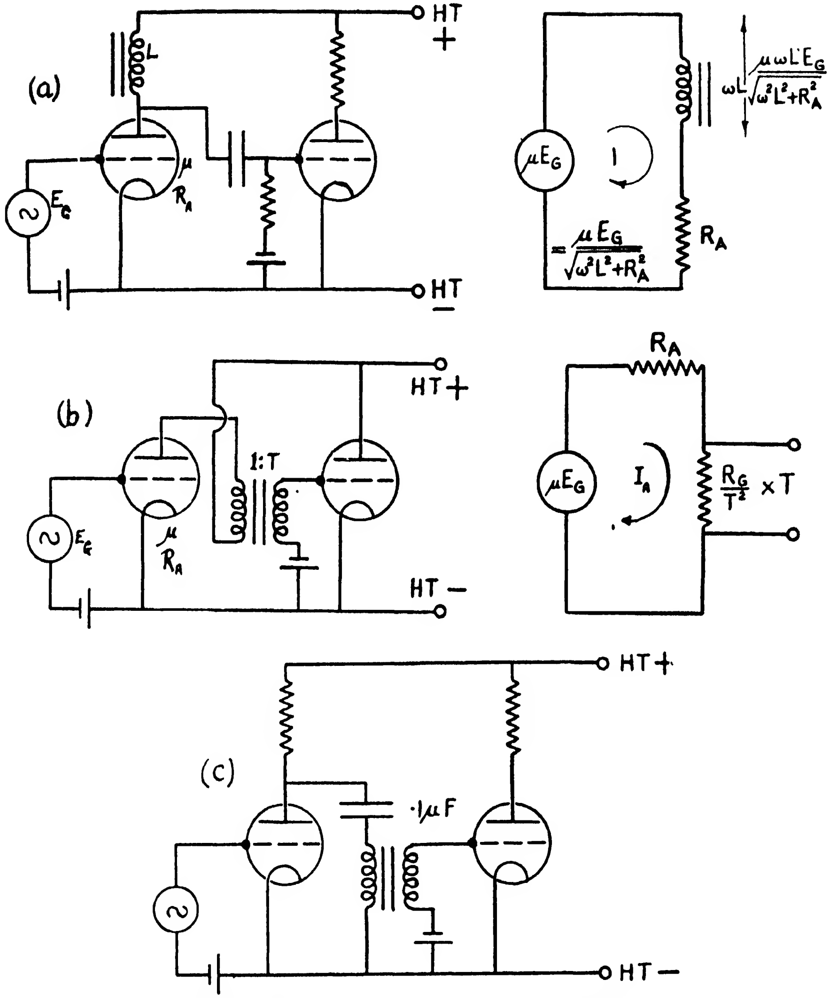


FIG. 54. *a*, Choke-capacity Amplifier, and Equivalent Circuit. *b*, Series-fed Transformer Coupled Amplifier, and Equivalent Circuit. *c*, Parallel-fed Transformer Coupled Amplifier.

The disadvantages attending the use of a choke as an anode load are:

(*a*) From the formula for the stage-gain it is apparent that the gain will be less for the lower values of the frequency $f=\omega/2\pi$.

The attenuation of low frequency signals, compared with higher values, will be reduced by the use of the maximum practicable value for the inductance L, so that ωL, even if ω is small, is still at least three times greater than R_A. In audio-frequency amplifiers, an iron-cored choke of value greater than 20 H. is used. The use of an iron core, necessary to ensure sufficiently high inductance in a component of moderate physical size, leads to a second difficulty in that the D.C. component of the valve anode current flowing through the choke winding will tend to cause magnetic saturation of the core, causing a reduction of effective inductance. The choice of an efficient component is made with both these points in mind.

(*b*) Since a large inductance necessarily has an associated self-capacity, a reduction of gain is also experienced at the higher frequencies, since then the choke impedance is considerably reduced by the effect of a parallel capacitance of value $1/\omega C$, where C is the choke self-capacity.

Transformer Coupling. The advantage offered by the use of a transformer as the coupling component between two valves in an amplifier is that an additional gain is furnished by the use of a step-up transformer with a ratio of as much as 1 : 8 times. Series-fed and parallel-fed connections are used (fig. 54): the latter circuit method has the advantage that the D.C. component of the valve anode current does not pass through the transformer primary, so its effective inductance is not thereby reduced.

At medium frequencies (200 to 3000 c./s.), the effects of the capacitances and inductances in the circuit shown in fig. 54*b* can be neglected in order to calculate, as correctly as is usually necessary, the gain of such an amplifier. A much simplified equivalent circuit can then be considered, as in fig. 54*b*. In this circuit, the primary inductance reactance and the effects of circuit capacitances are negligible compared with the effect of the reflected resistance R_G/T^2 due to the secondary load R_G, which is the input resistance of the second valve.

$$I_A = \frac{\mu E_G}{R_A + (R_G/T^2)}.$$

∴ the R.M.S. voltage across the transformer primary is

$$I_A \cdot \frac{R_G}{T^2} = \frac{\mu E_G (R_G/T^2)}{R_A + (R_G/T^2)}.$$

The output voltage across the transformer secondary is T times this amount, and thus will be $\frac{\mu E_G(R_G/T)}{R_A+(R_G/T^2)}$.

$$\therefore \text{ gain} = \frac{\text{voltage at grid of valve } V_2}{\text{voltage at grid of valve } V_1} = \frac{\mu R_G/T}{R_A+(R_G/T^2)}. \qquad (168)$$

This gain will be a maximum for a value of the transformer ratio T given by differentiating (168) with respect to T, and equating to zero.

$$\therefore \quad \frac{-\left(R_A+\frac{R_G}{T^2}\right)\frac{\mu R_G}{T^2}-\frac{\mu R_G}{T}\left(-\frac{2R_G}{T^3}\right)}{\left(R_A+\frac{R_G}{T}\right)^2}=0.$$

$$\therefore \quad \frac{\mu R_G}{T^2}\left(R_A+\frac{R_G}{T^2}\right)=\frac{2\mu R_G{}^2}{T^4}.$$

$$\therefore \quad R_A+\frac{R_G}{T^2}=\frac{2R_G}{T^2}.$$

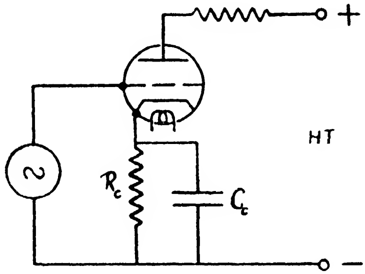
$$\therefore \quad \frac{R_G}{T^2}=R_A.$$

$$\therefore \quad T=\sqrt{\frac{R_G}{R_A}}. \qquad . \quad . \quad . \quad (169)$$

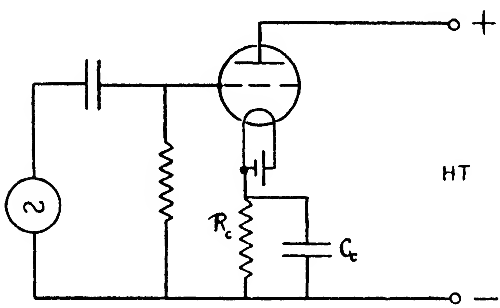
Since R_A, for an average triode valve, is about 50 kΩ., whilst R_G, the input resistance to a valve (grid negatively biased) is, about 5 MΩ. so the maximum values of transformer ratio T encountered in practical audio-frequency amplifiers are given by $T=\sqrt{(5\times10^6)/(5\times10^4)}=10$.

✓**Automatic Bias.** In amplifier practice the use of auxiliary grid-bias batteries is deprecated; a preferable method is to use an automatic bias arrangement. A resistance is arranged in the anode current circuit of the valve so that its cathode, or filament, depending on whether the valve is indirectly or directly heated, is maintained positive to the extent of the necessary bias relative to H.T.—, whereas the steady grid potential is zero (i.e. at H.T.—), (figs. 55). This is equivalent to arranging the grid potential negative with respect to the cathode.

The positive cathode bias is then equal to $I_A R_C$, where I_A is the steady component of the anode current, and R_C is the bias resistance used. Since I_A is known from valve data at the bias required, so R_C can be readily calculated, using Ohm's law, to obtain $I_A R_C$=bias required.



(a) CATHODE BIAS



(b) AUTOMATIC GRID BIAS, DIRECTLY-HEATED VALVE

FIG. 55. *a*, Cathode Bias Applied to an Indirectly heated Valve. *b*, Automatic Grid Bias used with a Directly heated Valve.

The anode current contains an A.C. component superimposed on the D.C. when it is amplifying an alternating signal input. The bias developed across the resistance R_C will, therefore, fluctuate at the signal frequency. To avoid this usually undesirable effect a by-pass condenser C_C is put in parallel with R_C of such capacity that its reactance $1/2\pi f C_C$ is less than 10% of R_C at the lowest signal frequency encountered.

For example, a valve has an anode current of 32 mA. when the bias is -10 V. What values of bias-resistance and condenser are necessary if the lowest signal frequency at the valve input is 100 c./s.?

The bias voltage $= 10 = I_A R_C = \frac{32}{1000} R_C$.

$$\therefore \quad R_C = \frac{10{,}000}{32} = 313 \ \Omega.$$

$$\frac{1}{\omega C_C} = \frac{1}{2\pi \,.\, 100 \,.\, C_C} \text{ should equal } 31{\cdot}3 \ \Omega. \text{ max.}$$

$$\therefore \quad \text{minimum } C_C = \frac{10^6}{200\pi \times 31{\cdot}3} \mu\text{F.} = \frac{320}{2\pi} = 51\mu\text{F.}$$

Grid Bias developed by Alternating Input. By the insertion of a condenser-resistance combination *CR* in the grid input circuit to a valve as shown in fig. 56, the negative bias developed on the grid relative to the cathode can be made to depend on the amplitude of the input alternating voltage. The condenser *C* is chosen to have such a capacity that it acts as a low reactance at the input frequency concerned. Such a circuit, therefore, is used only when the input is at high frequency (usually above 100 kc./s). A large fraction of the alternating input voltage will appear across the grid-cathode of the valve. Grid current will flow only when the valve grid is at a positive potential because of the input. Such grid current pulses will charge the condenser *C*, the plate *B* of this condenser acquiring a negative charge due to the accumulation of the electrons which reach the grid. Since plate *A* of the condenser is connected, via the resistance of the input generator circuit, to the valve cathode, so the D.C. potential developed across *C* will put a negative bias on the valve grid. If there is no leak resistance *R* across the condenser, then the bias developed will very soon be as great as the positive peak input voltage from the generator. The circuit will then cease to amplify the input and there will be no corresponding anode current alternations. On the other hand, if an appropriate value of resistance *R* is connected across the condenser, or connects the plate *B* to the cathode, then the steady P.D. across the condenser, and so the negative grid bias, will reach an equilibrium value which is some fraction of the positive peak input from the generator; this fraction depending on the magnitudes of *C* and *R*, and increasing towards unity as *R* is increased.

Such a method of producing a negative grid bias on a valve is widely used in oscillator circuits (Chapter 7). Moreover, if the input alternating potential has an amplitude which varies in accordance with a lower frequency modulation voltage (Chapter 8), then a detector circuit can be developed.

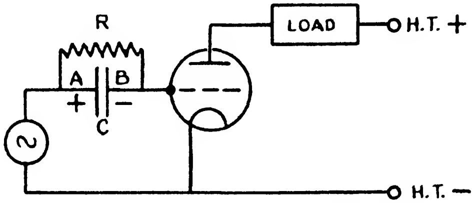


FIG. 56. Self-bias depending on Alternating Input.

High-frequency* Amplifiers. When amplifying signals at frequencies in excess of 50 kc./s. the effect of capacity in the circuit becomes important. In particular the capacitances between the electrodes of the valves used have to receive serious consideration. The effect of inter-electrode capacity in a valve is considered by Miller.

The Miller Effect.† The input admittance of a triode valve is controlled mainly by the effect of the inter-electrode capacitative link between the input circuit to the grid, and the output circuit at the anode.

In fig. 57 let the anode-grid capacity be C_{GA}, the anode-cathode capacity be C_{AC}, and the grid cathode capacity be C_{GC}. In the analysis which follows, the effect of C_{AC}, which is the smallest of these capacities, is neglected. In any case, it can be considered as simply adding slightly to the other two capacities. Using the notation indicated in fig. 57, the following equations are developed, on applying Kirchhoff's laws:

$$I_G = I_G' + I_G'', \qquad (170)$$

$$I_G' = E_G[(1/R_G) + j\omega C_{GC}]. \qquad (171)$$

$$I_G'' = (E_G + mE_G/\underline{\theta})\, j\omega C_{GA}, \qquad (172)$$

where m is the stage-gain, and θ is the phase-angle between $-E_G$, the grid input voltage, and mE_G the voltage across the anode load impedance Z. The magnitude of θ will depend upon the relative values of the resistance, inductance and capacitance comprising Z.

* The terms high frequency (H.F.), and radio frequency (R.F.), are usually considered to be synonymous.

† J. M. Miller, *Bureau of Standards Scientific Paper*, 351.

I_G is the total effective grid current from the supply at pulsatance ω, whilst I_G' and I_G'' are the components of this current via C_{GC} and C_{GA} respectively. R_G is the input resistance to the valve.

Equation (172) is conveniently rewritten in the form:

$$I_G''=j\omega C_{GA}[E_G+mE_G(\cos\theta+j\sin\theta)]. \qquad . \quad (173)$$

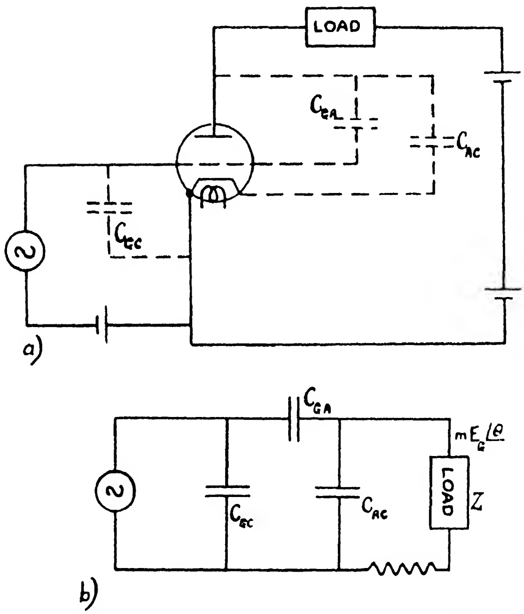


FIG. 57.—The Input Admittance of a Triode Valve (Miller Effect).

Substituting for I_G' and I_G'' from (171) and (173) in (170) gives

$$I_G=E_G[(1/R_G)+j\omega C_{GC}]+j\omega C_{GA}[E_G+mE_G(\cos\theta+j\sin\theta)].$$

$$\therefore \quad \text{Input admittance } Y_G=\frac{I_G}{E_G}$$

$$=\frac{1}{R_G}-m\omega C_{GA}\sin\theta+j\omega(C_{GC}+C_{GA}+mC_{GA}\cos\theta). \qquad (174)$$

Equation (174) clearly indicates that both the in-phase and 90° out-of-phase components of the input admittance are affected by m, and hence by the anode load. The input admittance will have a negative in-phase, or resistive component if $m\omega C_{GA}\sin\theta$ is positive, and exceeds $1/R_G$ in value. This implies that energy is being fed back via the anode-grid capacity to the input circuit in such a manner as to overcome the effects of positive resistance energy-loss at the input. With suitable input and output LC circuits, the input LC circuit can have its resistance loss reduced to zero, or made negative, so that continuous oscillations may be sustained (cf. the T.A.T.G. oscillator p. 182).

It is seen from (174) that the effective input capacity of a triode valve is considerably greater than its nominal value, which depends on electrode geometry. For example, suppose the anode load is resistive, so that $\theta=0°$, and let C_{GC} be 5 μμF., C_{GA} be 8 μμF. and the stage-gain be twenty times. Substitution of these values in equation (174) shows that the effective input capacity C_i is given by:

$$C_i = C_{GC} + (m+1)C_{GA} = 5 + (21 \times 8) = 173 \text{ μμF.}$$

At a frequency of 10,000 c./s., at the upper limit of the audio-frequency range, a nominal grid-cathode capacity of 5 μμF. will be a reactance of $\dfrac{1}{\omega C_{GC}} = \dfrac{1}{2\pi \times 10^4 \times 5 \times 10^{-12}} = 3{\cdot}18\ \text{M}\Omega.$, which is negligible in its effect. But the true effective input capacitative reactance is $\dfrac{1}{2\pi \times 10^4 \times 173 \times 10^{-12}} = 92{,}000\ \Omega.$, which is certainly not negligible. At radio-frequencies of the order of 1 Mc./s. such an input capacitative reactance will be only 920 Ω., so that an R.F. voltage input of only 0·92 V. will produce a reactive current component of 1 mA.

An amplifier intended for signals at radio-frequency cannot then be efficient if resistance-capacity coupled triode valves are used. The use of a tuned rejector-circuit as an anode load in place of a resistance will partly overcome the deleterious effects of the inter-electrode capacities discussed above, since the capacities are then simply in parallel with the necessary tuned circuit capacity for resonance, the gain per stage being $\dfrac{\mu(L/CR)}{(L/CR)+R_A}$, where L/CR is the dynamic resistance of the rejector-circuit at resonance (see p. 57). However, the use of input- and output-tuned circuits associated with a triode valve, tend to make the circuit prone to oscillate (see p. 172). Again, the effective input capacity of $C_{GC}+(m+1)C_{GA}$ will limit the minimum capacity which can be associated with the tuned circuit used. The remedy is to use a valve of much lower anode-grid capacity of the screen-grid or pentode type. Such valves give an added advantage for use as R.F. voltage amplifiers, in that their A.C. resistance and amplification factor are much higher than those of triodes.

The Screen-grid Tetrode. A second grid is introduced between the grid and anode of a triode, reducing the capacity between these electrodes, giving a tetrode, or four-electrode valve. This screen grid is usually maintained at a positive D.C. potential of value about two-thirds the anode D.C. potential. As regards R.F. voltages, however, this screen is earthed by a by-pass condenser, usually 0·1 μF., connected between the screen and earth (see fig. 58).

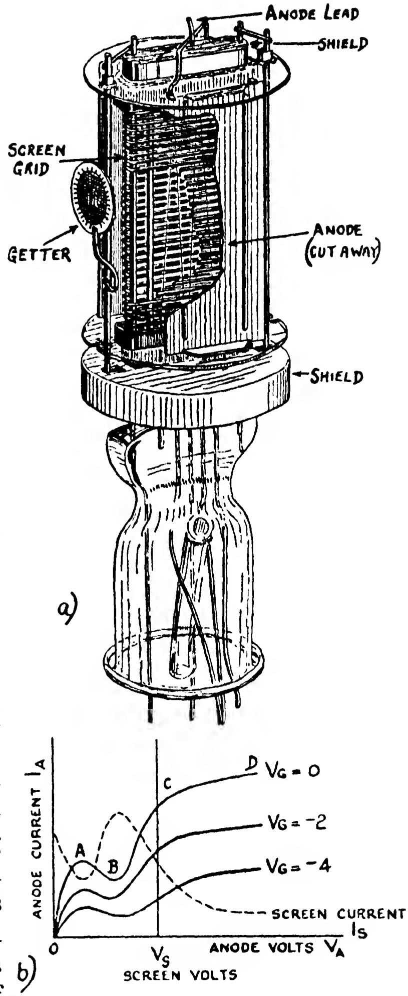


FIG. 58. *a*, The Screen-grid Valve.
b, Anode Characteristic of Screen-grid Valve.

In the construction of this valve particular attention is paid to the problem of ensuring an anode-grid capacity which is as low as possible. To this end the anode area is reduced in size; the grid and anode leads are taken to separate ends of the valve, and skirts at screen potential are arranged at the top and bottom of the electrode assembly to reduce the electrostatic field which exists between the control grid ends and the anode. In this fashion, anode-grid capacities as low as 0·01 μμF. are achieved compared with values of the order of 5 μμF. obtained in the usual receiver triodes.

The anode-current vs. anode-voltage ($I_A : V_A$) characteristic of this tetrode is peculiarly kinked, as is shown in fig. 58*b*. This effect is due to secondary emission from the anode.

The useful portion of this characteristic in R.F. amplifier practice is between points *C* and *D*; it is essential to confine the anode load voltage variations obtained in the amplifier to within these limits, since departure from this restricted linear region will result in distortion. The line *CD* is inclined by only a very small angle to the anode volts axis, and this inclination represents a high value of A.C. resistance. This high resistance is directly due to the introduction of the positive screen. A change of anode voltage ∂V_A will not be very effective in altering the anode current, since the electric field at the valve cathode, due to the positive anode, is much reduced by the shielding effect of the screen grid. For a change ∂V_A the corresponding anode current change ∂I_A will be very small; consequently the ratio $\partial V_A/\partial I_A = R_A$ is large. Likewise, the amplification factor $\partial V_A/\partial V_G$ will be considerably greater than in a triode valve, since the anode voltage will have to be changed by a large amount to be as effective as a grid voltage change in altering the anode current. In screen-grid valves, the A.C. resistance is between 0·5 and 2 MΩ., whilst the amplification factor is between 500 and 2000. The mutual conductance will be much the same as in the corresponding triode, since the electric field of the control grid at the cathode is little affected by the presence of the screen.

If the anode voltage of this type of tetrode is reduced to below the screen voltage, then the anode current decreases rapidly, as is indicated by the region of the curve *CB*. The anode will emit secondary electrons. In the normal operation of the valve, with the anode voltage greater than the screen volts, these secondaries will simply return to the anode, as they do in a triode valve. If, however, the screen is more positive than the anode, then many of the secondary electrons will go to the screen-grid instead; the anode current will then be reduced because electrons are leaving it in numbers almost as great as those arriving, whereas the screen-current will increase. If the anode voltage is much reduced—to only 20 to 30 V.—secondary electrons are no longer produced since the primaries will have insufficient energy. A subsidiary rise of anode current to a minor peak value at *A* is therefore experienced, indicated by the region *BA* of the characteristic curve.

Then as the anode voltage is decreased to zero, so the anode current falls to zero.

Over the region AB of the curve, it is remarkable that the anode current decreases with increase of anode voltage, so $R_A = \partial V_A / \partial I_A$ is negative. The dynatron oscillator (p. 183) utilises this effect, since the negative resistance of a screen-grid valve, operated at the correct potentials, is balanced against the positive dynamic resistance of a tuned circuit in the anode. This negative resistance is due to the fact that, within the range of anode voltages experienced between points A and B of the curve, an increase of anode voltage brings about a greater increase of secondary emission than it does of primary electrons arriving at the anode.

The Pentode and Beam Tetrode Valves. In modern electronic practice the screen-grid valve is rarely used: the pentode or beam tetrode, which developed from the screen-grid type, are preferable. In both these valves, the kink in the characteristic due to secondary emission is eliminated: in the pentode by the use of a third grid, the suppressor, situated between the screen and anode; in the beam tetrode by the use of a more critical anode spacing in conjunction with the use of "beam" plates. Both radio-frequency and audio-frequency patterns of these valves are made: in the former, care is taken in the electrode construction and connections to ensure low anode-grid capacity; in the low-frequency model this low capacity is not an essential requirement, the chief attention being paid to ensuring high efficiency working concomitant with low distortion.

In the pentode valve, the suppressor grid, which is usually at cathode potential, has fewer turns per cm. than the control or screen grids, and is placed just inside the anode. As a result, the suppressor does not significantly impede the fast primary electrons accelerated through the screen to the anode, whereas the comparatively slow secondaries, travelling in all directions away from the anode plane, are readily decelerated to zero velocity before this third grid, and so return to the anode instead of penetrating to the screen grid.

When the anode potential is lower than the screen potential, this suppression of secondaries is still effective, but then the slower primary electrons become somewhat impeded by the suppressor grid, giving rise to the rounded "knee" typical of the pentode characteristic (fig. 59).

The beam tetrode valve dispenses with a suppressor grid, but instead is furnished with a pair of plates disposed about the screen-grid supports, and maintained at cathode potential, so that primary electrons which reach the anode are confined to a beam (see fig. 59). Electrons which leave the sides of the cathode, and which travel through the distorted electrostatic fields around the grid support wires, do not reach the anode, since they impinge on the beam plates. Consequently, a more rigid relationship is established between electrode shape and disposition, and anode current. In some types of beam tetrode, particularly the 6L6 (KT 66) valves, the screen and control grids have the same number of turns per cm., and are aligned so that the screen wires are exactly behind the control grid wires. This leads to further confinement of the effective electrons to beams, and also reduces the screen current.

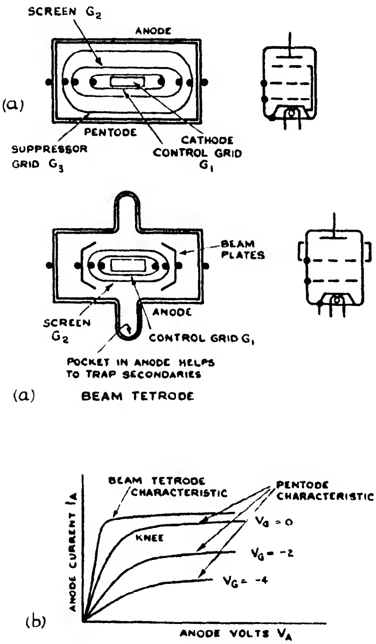


FIG. 59. *a*, Pentode and Beam Tetrode Valves. *b*, Pentode and Beam Tetrode Characteristics.

To prevent secondary electrons leaving the anode from reaching the screen of this valve, the anode is placed at such a distance that when its potential is less than the screen potential, then the primary electrons build up a negative space-charge between screen and anode. A potential minimum is thereby introduced which does not prevent the majority of the fast-moving primaries from reaching the anode but, on the other hand, slows down the slower secondaries to such an extent that they return to the anode instead of reaching the more positive screen. This potential minimum is further enhanced by the presence of the beam plates. The result is an anode characteristic similar to that of the pentode valve, but with a rectangular-shaped rather than rounded knee, occurring at a lower anode potential. Such a valve can therefore be used with a bigger variation of anode load voltage without distortion occurring than is the case in the equivalent pentode. As a power-amplifier valve it is therefore capable of greater working efficiency.

The R.F. pentodes and beam tetrodes are furnished with an anode of small area, a top-cap grid connection, and skirts at the upper and lower ends of the electrode assembly, as in the case of the screen-grid valve, to ensure low anode-grid capacity. Likewise the grids have many turns of wire per cm., and the anode spacing is great to obtain a valve of high A.C. resistance and amplification factor. These requirements are assisted by the introduction of a suppressor grid which, shielding the anode, enables even higher μ and R_A to be obtained than in the screen-grid valve.

In the case of the low-frequency beam power valve, used in particular as the output valve of an amplifier supplying considerable load current, the designer concentrates on ensuring a high mutual conductance, sufficient anode dissipation and with the knee of the characteristic at as low an anode potential as possible.

Radio-frequency Tuned Voltage Amplifiers. To amplify a signal voltage alternating at high frequency (100 kc./s. to 50 Mc./s.), R.F. pentode valves are usually employed, with parallel tuned circuits as anode loads. The commonest types of coupling are known as (*a*) tuned anode, (*b*) tuned grid, (*c*) R.F. transformer, and (*d*) tuned R.F. transformer coupling (see fig. 60).

(*a*) *Tuned-anode Coupling.* In this case a tuned rejector circuit (see p. 57) is used as the anode load for the first valve. Since such a circuit behaves as a high resistance of value L/CR at resonance, so the gain exhibited by the first stage is given by

$$m=\frac{\mu(L/CR)}{(L/CR)+R_A}. \quad . \quad . \quad . \quad (175)$$

Since $L/CR=Q\omega L$, from equation (78), so a high value of m is obtained in this circuit by using a coil of high "Q". This necessarily implies a sharply selective tuned circuit, which may involve the cutting of the higher audio-frequencies in the sidebands (see p. 197) if such an R.F. amplifier is used in a domestic receiver. It is noteworthy that the A.C. resistance, R_A, of an H.F. pentode is of the order of 1 MΩ. The dynamic resistance of a tuned rejector circuit cannot usually be more than 200 k Ω. Substitution of these figures in equation (175) evinces that the gain m obtainable is much smaller than the maximum possible value, μ. Thus if μ is 1000 for a particular H.F. pentode, then

$$m = \frac{1000 \times 100{,}000}{100{,}000 + 1{,}000{,}000} = \frac{1000}{11},$$

only $\frac{1}{11}$th of the maximum possible amount.

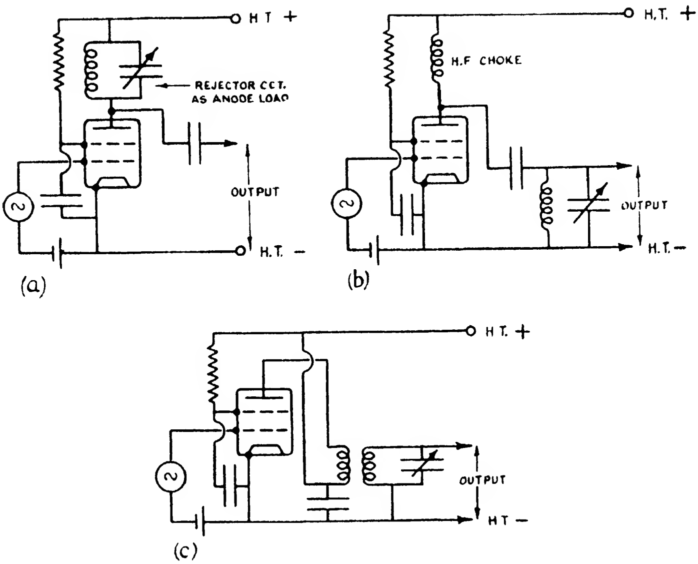


FIG. 60. R.F. Voltage Amplifiers. *a*, Tuned-anode Coupling. *b*, Tuned-grid Coupling. *c*, R.F. Transformer Coupling.

The gain m can be conveniently expressed, in such circumstances, as

$$m = \frac{\mu (L/CR)}{R_A} = g_m \frac{L}{CR} = g_m \omega L Q, \quad . \quad . \quad (176)$$

where L/CR is neglected compared with R_A in the denominator. Thus the effective Q of the amplification curve, deciding the selectivity of the circuit, is approximately the same as the Q of the rejector circuit alone.

(*b*) *Tuned-grid Coupling.* This coupling method is illustrated in fig. 60*b*. An air-cored inductance, or choke, which has a high impedance to A.C. at the radio frequency used, is employed as the anode load. The voltage variations produced across it are coupled by a condenser to a tuned circuit which forms the input to the second stage of the amplifier. This tuned circuit improves the selectivity of the amplifier, but reduces the effective anode load presented to the valve, because it is effectively in parallel with the choke. This type of coupling is not frequently used, methods (*c*) and (*d*) being preferable.

(*c*) *R.F. Transformer Coupling.* This method of connection has the practical advantage that the grid input circuit of the second valve is isolated from the H.T. supply. The primary or secondary or both can be tuned circuits. In the third case the transformer can have band-pass characteristics (see pp. 68 and 197) to enable it to select a required frequency band width. Considering the case where the secondary only is tuned, then from formula, p. 66,

$$R = R_1 + \frac{\omega^2 M^2 R_2}{Z_2^2} \quad \text{and} \quad X = X_1 - \frac{\omega^2 M^2 X_2}{Z_2^2},$$

where ω is the pulsatance of the signal concerned, M is the mutual inductance of the R.F. transformer, R_1 and X_1 are the primary resistance and reactance respectively and R_2, X_2 and Z_2 are the secondary resistance, reactance and impedance respectively. R is the effective primary resistance, increased by the tuned circuit coupled to it; X is the effective primary reactance, decreased by the coupled tuned circuit.

Since the secondary is tuned, $Z_2 = R_2$ and $X_2 = 0$, and the reflected resistance in the primary due to the secondary becomes $\frac{\omega^2 M^2 R_2}{X_2^2 + R_2^2} = \frac{\omega^2 M^2}{R_2}$. Since R_2 is small, and ω is a large value at radio frequency, so this reflected resistance is great compared with R_1 and X_1, and these latter can be neglected.

$\therefore \quad Z_1 = \frac{\omega^2 M^2}{R_2}$ approximately gives the effective primary impedance.

$$\therefore \quad \text{Primary current } I_1 = \frac{\mu E_G}{R_A + (\omega^2 M^2 / R_2)},$$

where E_G = alternating voltage applied to grid of the first valve, and μ = amplification factor of first valve.

The voltage induced in the secondary $= \omega M I_1$.

$\therefore$ Secondary current $=\dfrac{\omega M I_1}{R_2}=I_2.$

$\therefore$ Voltage across secondary$=E_G'=\omega L_2 I_2$, where L_2 is the inductance in the tuned secondary circuit.

$$=\omega L_2 \cdot \frac{\omega M}{R_2} \cdot \frac{\mu E_G}{[R_A+(\omega^2 M^2/R_2)]}.$$

$$\therefore \quad \text{Gain} \quad m=\frac{E_G'}{E_G}=\frac{\omega^2 \mu M L_2}{R_2 R_A+\omega^2 M^2}.$$

This gain will be a maximum for a value of coupling given by putting $dm/dM=0$.

$$\therefore \quad \frac{d}{dM}\left[\frac{\omega^2 \mu M L_2}{R_2 R_A+\omega^2 M^2}\right]=0.$$

$$\therefore \quad (R_2 R_A+\omega^2 M^2)\,\mu\omega^2 L_2=\mu\omega^2 M L_2 \,.\, 2\omega^2 M.$$

$$\therefore \quad R_2 R_A+\omega^2 M^2-2\omega^2 M^2=0.$$

$$\therefore \quad \omega^2 M^2=R_2 R_A.$$

$$\text{Then} \quad m=\frac{\mu\omega^2 M L_2}{2\omega^2 M^2}=\frac{\mu L_2}{2M}.$$

But $M=K\sqrt{(L_1 L_2)}$, where K is the coefficient of coupling between the two circuits, where L_1 is the primary inductance.

$$\therefore \quad m=\frac{\mu L_2}{2K\sqrt{(L_1 L_2)}}=\frac{\mu}{2K}\sqrt{\frac{L_2}{L_1}}. \quad . \quad . \quad . \quad (177)$$

Again, since $\omega^2 M^2=R_2 R_A$

$$\therefore \quad \omega^2 K^2 L_1 L_2=R_2 R_A$$

and putting $\omega^2=\dfrac{1}{L_2 C_2}$, therefore $\dfrac{K^2 L_1 L_2}{L_2 C_2}=R_2 R_A$, where C_2 is the tuned secondary capacitance.

$$\therefore \quad K^2=\frac{R_2 R_A C_2}{L_1} \quad . \quad . \quad . \quad . \quad . \quad (178)$$

gives the required optimum coupling value, which on inserting practical values for R_2, R_A, C_2 and L_1, where the frequency is 1 Mc./s., makes K about 0·8.

(*d*) *Tuned R.F. Transformer Coupling.* If both primary and secondary circuits are tuned, it can be shown by a complete analysis of the circuit that optimum coupling occurs when

$$\omega^2M^2=\left(R_1+\frac{L_1}{C_1R_A}\right)R_2 \quad . \quad . \quad . \quad (179)$$

and the gain $$m=\frac{\mu(L_1/C_1R_A)}{2\sqrt{\{[R_1+(L_1/C_1R_A)]R_2\}}}. \quad . \quad . \quad (180)$$

Optimum coupling at 1 Mc./s. occurs for values of K of the order of 0·01.

Variation of Gain with Frequency. In all these methods of coupling together the stages of an R.F. amplifier, it is important to note that the gain and condition for optimum coupling depend on the signal frequency, $\omega/2\pi$. Hence in normal broadcast receiver practice, where it is required to accept and amplify a widely varying range of radio-frequencies depending on the station being received, it is apparent that the receiver performance must needs vary considerably from one end of the wave-band to the other. This variation is largely overcome by employing the super-heterodyne principle whereby all receiver aerial signals are changed to a constant frequency value, called the intermediate frequency, I.F., and I.F. amplifier stages employ coupling transformers with both primary and secondary tuned, and designed to work at the optimum for the constant frequency value chosen.

The Variable-mu Valve. Equation (176) indicates that the gain of an H.F. pentode, employing a tuned circuit directly or indirectly coupled as an anode load, depends directly on the mutual conductance of the valve. In the variable-mu valve, the control grid of a screen-grid tetrode or H.F. pentode is more open at the centre than at the ends (see fig. 61), or is wound so that the spaces between adjacent grid turns increase exponentially from one end of the grid to the other. As a result the mutual characteristics ($I_A : V_G$) of such a valve exhibit pronounced "tail", since the cathode surface behind the more open spaces of the grid continues to furnish electrons to the valve anode current when the grid bias is sufficiently negative to prohibit by space-charge the flow of electrons from the rest of the cathode. Obviously such a valve has a large cut-off bias, and a value of mutual conductance which increases continuously as the negative grid bias is reduced from cut-off value to zero.

If the grid bias on a variable-mu valve is altered by means of a manually operated potentiometer, or is automatically controlled depending on the strength of the alternating input to the valve grid, so either manual or automatic control of the gain of an R.F. amplifier is achieved. This leads to the common forms of manual and automatic volume control methods used in broadcast receivers.

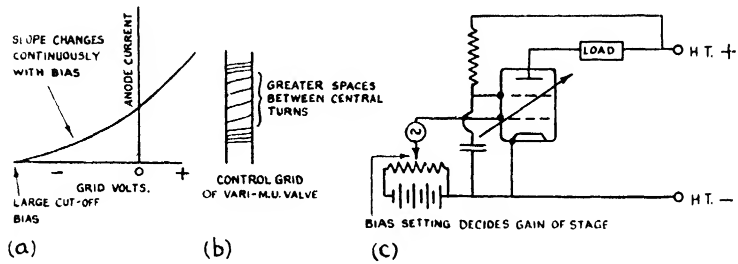


FIG. 61. The Action of the Variable-mu Valve.

Power Amplification. In using a valve as a power amplifier, the aim is to supply maximum power to the anode load with as little distortion as possible, instead of maximum voltage variation across the load. This demands large current variations through the load, so that if the use of excessive voltage variations across the valve is to be avoided, the valve must necessarily have a low A.C. resistance. Hence the power valves used as output valves in receivers, transmitters and other electronic apparatus where electrical power is converted into mechanical power or electromagnetic radiation, have much lower A.C. resistances than the voltage amplifying valves in electronic equipment. The latter usually serve to supply the power valve with a sufficiently large grid swing voltage, demanding little or no power consumption.

Let a valve amplifier be operated so that

E_G = input voltage to grid, R.M.S. value.
μ = amplification factor of valve.
R_A = A.C. resistance of valve.
R_L = anode load resistance.

Assuming the grid is not driven positive, so that grid current is avoided, then the effective anode voltage variation due to an input of E_G volts is μE_G. The corresponding current I is $\mu E_G/(R_A+R_L)$. The power developed across the anode load is therefore $I^2R_L=\frac{(\mu E_G)^2 \cdot R_L}{(R_A+R_L)^2}=P$.

This will be a maximum for a value of R_L given by putting $dP/dR_L=0$.

$$\therefore \quad (\mu E_G)^2\left[\frac{(R_A+R_L)^2-R_L \cdot 2(R_A+R_L)}{(R_A+R_L)^4}\right]=0.$$

$$\therefore \quad R_A+R_L-2R_L=0.$$

$$\therefore \quad R_A=R_L. \qquad . \quad . \quad . \quad . \quad . \quad (181)$$

As in the case of any other form of electric generator, maximum power is supplied to the load when the load resistance equals the internal resistance of the source of supply. However, the requirement of low distortion must also be fulfilled, requiring extra consideration of the optimum load resistance value to be used.

Class A Power Amplification. The anode load of a power amplifier must necessarily be a resistive device; power cannot be dissipated in the wattless reactive components, condensers and inductances, in which the current differs in phase by 90° from the voltage. The usual loads are therefore ordinary resistances, or resistive elements such as loudspeakers, headphones, electric meters, telephone meters or aerials radiating energy or such resistances coupled to the power valve by a matching transformer or tuned circuits of the rejector circuit type with a make-up current in phase with the applied A.C. voltage: simulating a resistance in their effect. If a resistance is inserted directly in the anode circuit of the valve, then class A amplification conditions must be observed using an ordinary amplifier if distortion is to be avoided. This brings about the limitation that the grid swing voltage is restricted, and also that an increase of anode current brought about by a reduction of negative grid voltage is accompanied by an increase of anode load voltage, with a consequent reduction of anode volts, and hence power output.

Suppose a power valve has constants μ, g_m and R_A, and it operates with a mean, steady anode voltage V_A. Let $-V_G$ be the cut-off grid bias with this anode voltage. An alternating grid voltage is applied which, at its positive peak value, just brings the grid potential to zero, whereas at the negative peak value the dynamic characteristic ($I_A - V_G$) of the valve just begins to bend, so that class A conditions are observed (cf. p. 119). Let $2I_A$ be the total peak-to-peak anode current excursion brought about by the maximum permissible grid voltage change. Then the anode potential will be caused to vary from $V_A + I_A R$ to $V_A - I_A R$, if R is the resistive anode load. A change of anode potential of $I_A R$ is compensated by a change of grid voltage of $I_A R/\mu$, so that when the anode potential is a maximum at ($V_A + I_A R$), the necessary cut-off bias will be $[-V_G - (I_A R/\mu)]$. Assuming the valve $I_A - V_G$ characteristic is linear down to the cut-off bias value, then the maximum permissible peak grid voltage input is $\frac{1}{2}[V_G + (I_A R/\mu)]$.

But $I_A = \dfrac{\mu E_G}{R + R_A}$, and substituting for E_G

$$I_A = \frac{\mu}{R+R_A} \cdot \frac{1}{2} \cdot \left(V_G + \frac{I_A R}{\mu}\right).$$

$$\therefore \quad I_A \left\{ 1 - \frac{R}{2(R+R_A)} \right\} = \frac{\mu V_G}{2(R+R_A)}.$$

$$\therefore \quad I_A(2R_A + R) = \mu V_G.$$

$$\therefore \quad I_A = \frac{\mu V_G}{R + 2R_A}.$$

The power output P will be $I^2_{R.M.S.}R$, where $I_{R.M.S.} = I_A/\sqrt{2}$.

$$\therefore \quad P = \frac{\mu^2 V_G^2 R}{2(R+2R_A)^2}.$$

To find the value of the anode load R at which this power output is a maximum, put $dP/dR = 0$.

$$\therefore \quad \frac{\mu^2 V_G^2}{2} \cdot \frac{d}{dR} \left\{ \frac{R}{(R+2R_A)^2} \right\} = 0.$$

$$\therefore \quad (R+2R_A)^2 - R \cdot 2(R+2R_A) = 0.$$

$$\therefore \quad R^2 + 4RR_A + 4R_A^2 - 2R^2 - 4RR_A = 0.$$

$$\therefore \quad 4R_A^2 = R^2.$$

$$\therefore \quad R = \pm 2R_A. \qquad (182)$$

So the maximum power output is obtained when the anode load is twice the A.C. resistance of the valve.

The power output at this optimum load value will be

$$P=\frac{\mu^2 V_G{}^2 R}{2(R+2R_A)^2}=\frac{V_A{}^2 \cdot 2R_A}{32R_A{}^2}=\frac{V_A{}^2}{16R_A}.$$

The Load Line. The selection of an anode load of value $R=2R_A$ leads theoretically to the attainment of maximum undistorted power output from a power valve. Though serving as a useful guide in practice to the choice of load values in the case of triodes, yet a more accurate assessment of the load is required if distortion is to be kept to a minimum, since allowance must be made for the departure from linearity of the valve characteristics when operating near anode current cut-off. Moreover, using pentode power valves, the selection of a load $R=2R_A$ gives far from satisfactory results. The approach adopted is to draw a load line, which is a line representing the voltage-current characteristic of the anode load, superimposed on the anode voltage-anode current characteristics for the valve in question. If the anode load is a resistance then, by Ohm's law, the load line is straight. For an impedance as anode load which is sensibly reactive in its effect, the load line will be an ellipse.

In general, two methods of approach are used. Either the load resistance value is known, and it is required to estimate the percentage harmonic distortion if the valve operating conditions are also specified, or alternatively, a certain power valve is to be used, and the graphical method outlined below is adopted as a means of estimating the best operating potentials for the valve, and the best load resistance value to use in order to keep the distortion satisfactorily low.

Suppose a triode power valve with a total H.T. supply of 400 V. is used. A limitation to the total power which can be dissipated is set by the maximum temperature at which the anode can be safely operated. Let this correspond to 16 W. maximum anode dissipation. The static anode voltage : anode current curves are known, as in fig. 62. The curve representing 16 W. dissipation is drawn on these characteristics by plotting $V_A I_A$=constant, 16. When the anode current is zero (obtained in practice by setting the grid bias at cut-off), the voltage drop in the anode load is zero so the anode will be at the full H.T. voltage, 400. Hence the point A where $V_A=400$, $I_A=0$ is one point on the required load line. If class A conditions are to be observed, the peak positive grid input voltage just makes the grid potential zero or, at the most, slightly positive. Let B be the point where the zero grid voltage characteristic intersects the maximum dissipation curve. Then AB is the load line for maximum allowable power output. The slope of this load line obtained from $\frac{\text{AC in volts}}{\text{MC in amp.}}$ gives the corresponding resistance value to be used.

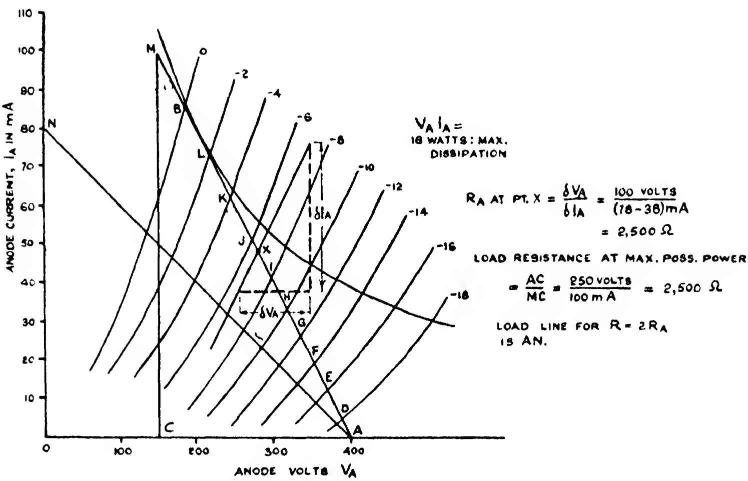


FIG. 62. The Load Line.

What is to be the criterion as regards distortion? Manifestly a small change of grid voltage anywhere in the region from $V_G=0$ to $V_G=$cut-off should bring about the same change in current through the anode load irrespective as to whether this grid voltage change takes place near cut-off bias, near the operating bias or near zero. Interpreted graphically, this means that the intercepts cut off along the load line by neighbouring $V_A : I_A$ characteristics should be equal for equal change of grid voltage in moving from one characteristic to the next, i.e. the intercepts DE, EF, FG, GH, HI, IJ, JK, KL, LB (fig. 62) should all be equal. If the load line as drawn does not produce such equal intercepts then it must be varied. Graphically this implies swinging the load line about the fixed point A to arrange it below the line AB so that these intercepts are as nearly equal as possible. At the same time the slope of the line does not want to be too small (i.e. the resistance too high) otherwise the power dissipated in it will depart too seriously from the maximum possible. A useful, simple rule is to realise that the total percentage harmonic distortion will be less than 10% if the maximum departure from equality of the intercepts referred to does not exceed 10 : 11.

Having determined the best load resistance value, the operating bias can be determined as being at the point X, where X is the mid-point of the load line such that $DX = XB$. If a fuller investigation is then required, the dynamic mutual characteristic $I_A - V_G$, with the appropriate anode load, should be drawn to establish that this operating grid bias is midway between $V_G = 0$ and the value of V_G at which this $I_A - V_G$ characteristic commences to bend.

An examination of the most suitable anode loads for power pentodes and beam tetrodes in accordance with these conceptions indicates that, for maximum undistorted power output, the anode load needs to be $\frac{1}{8}$th to $\frac{1}{16}$th of the valve A.C. resistance.

Anode Efficiency. Defined as the ratio

$$\frac{\text{alternating power output}}{\text{D.C. power input}}.$$

For example, consider the case of a class A amplifier operated for maximum power output, disregarding distortion. It has been shown (p. 147) that then $R_L = R_A$. Obviously the anode efficiency will be 50%.

This can be proved in an alternative manner. Suppose a class A amplifier operates with steady anode voltage V_A and anode current I_A. Then the D.C. power input is $V_A \cdot I_A$. The maximum positive to negative peak excursion that the anode voltage can execute by virtue of an alternating input to the grid will be from V_A to zero, and to a positive maximum of $2V_A$. Correspondingly the maximum possible anode current excursion will be $2I_A$. But the R.M.S. value of the alternating anode voltage will then be $V_A/\sqrt{2}$, and for the current $I_A/\sqrt{2}$. Consequently the alternating power output is $\frac{V_A \cdot I_A}{\sqrt{2} \cdot \sqrt{2}} = \frac{1}{2} V_A I_A$, and the anode efficiency is 50%.

Class B Amplification. A maximum possible efficiency of 50% is a limitation to the power handling capabilities of a valve, especially in the case of radio transmitters where, with apparatus of restricted size, it is required to supply the aerial with as much energy as possible. For this reason class B and class C amplification practices have been developed. These involve using the valve with a steady, operating grid bias at, or beyond, cut-off. Such practice is inadmissible in cases where a single valve is used with a resistance as anode load, since the negative half-cycle of the input does not then produce any corresponding anode current change: rectification is involved. However, if the anode load is an oscillatory rejector circuit, then the introduction of such distortion is not of importance. Fortunately such anode loads are those most useful in R.F. power amplifier practice.

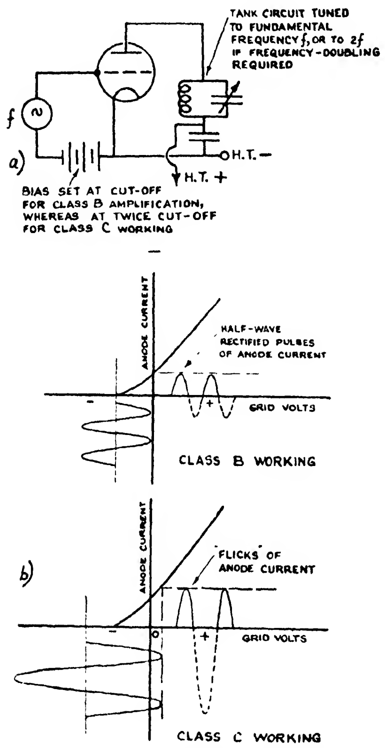


FIG. 63. Class B and Class C Amplification.

Though the anode current variations of such amplifiers have wave-forms which are rectified and distorted replicas of the input wave-form to the grid, yet such distortion necessarily corresponds to the introduction of second and higher order harmonics (see p. 69), whereas the rejector circuit anode load will only respond to the fundamental. The deliberately introduced distortion is, therefore, of no consequence as regards the ultimate output wave-form. Again, if the distortion introduced is predominantly second harmonic distortion, as in the case of a class B operated triode, then a class B push-pull amplifier, which eliminates even harmonic distortion, is valid (see p. 154), though the anode load is a resistance.

From the analysis given on p. 71, it is seen that the wave-form in the case of half-wave rectification possesses a first harmonic component of which the peak value is approximately half the maximum voltage attained. Hence, if the grid alternating input is approximately doubled, in the class B case (fixed bias at cut-off) compared with the class A case, then the fundamental component of the rectified output is capable of supplying the same power to a rejector circuit as would be supplied by a class A amplifier.

The anode efficiency of the class B amplifier is, however, greater because there is a smaller demand on the H.T. supply than when class A working operates. This is because anode current only flows when the alternating grid input is positive, and not when it is negative.

From the analysis given on p. 71, the average value of anode current during one-half cycle is seen to be $2I/\pi$, where I is the peak current. But during every alternate half-cycle the current is zero. Therefore the average current drain on the H.T. supply in class B working is I/π. In class A working, the mean D.C. anode current is $I/2$. Hence class B working demands an average current supply which is only $2/\pi$ of that in class A practice. Consequently a class B amplifier is $\pi/2$ times as efficient, giving a maximum possible efficiency of $\pi/2 \times 50\% = 78\%$.

Class C Amplification. The operating fixed negative bias can be made greater than cut-off, still further restricting the fraction of the input cycle time during which anode current flows. Thus in class C working, the grid bias is as much as twice cut-off value, giving anode current pulses for less than one-third of the operating time, with a consequent anode efficiency of as much as 85%. The considerable second and third harmonic distortion introduced restricts the use of such amplifier practice to solely those cases in which a rejector circuit is used as anode load. Push-pull working is now inadmissible, owing to the odd harmonics involved.

R.F. Power Amplification. To amplify the output of an oscillator, as in the usual radio transmitter master-oscillator, power amplifier system, a class C amplifier is commonly used. Apparatus with a total power output in excess of 250 W. commonly makes use of triode valves, but pentodes are frequently encountered in smaller gear.

To ensure the maximum anode efficiency, the alternating voltage applied to the grid (grid drive) is made great enough to produce a positive grid at the positive peak grid input volts. Thus the input circuit to the amplifier needs to be able to supply a moderate amount of power. When the grid is most positive, the valve anode current will be large, and the alternating voltage across the tuned circuit anode load will be at its maximum. Since the anode voltage alternates in anti-phase to the grid drive volts (see p. 119), the maximum positive grid voltage will occur simultaneously with the minimum possible anode potential, so that overheating of the anode due to the large current flicks brought about by driving the grid positive is not excessive.

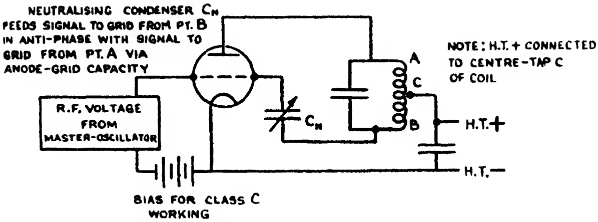


FIG. 64. R.F. Power Amplifier Circuit.

Such R.F. power amplifier circuits need neutralising, as is indicated in fig. 64, to prevent the power amplifier from going into oscillation due to feedback via the anode-grid capacity from the anode tuned circuit to the grid input tuned circuit.

Push-pull Amplification. A circuit arrangement which eliminates distortion at the second and other even harmonics which may be introduced by the valve characteristic, in which a pair of valves is used, generally as a power amplifier, where the grid inputs to the two valves are equal in magnitude, but differ in phase by 180°.

Suppose an A.C. input of sinusoidal wave-form is applied at the primary of the input transformer (fig. 65). Since the centre-tap of the secondary of this transformer is connected to fixed bias battery, voltage V_G, so the grid of valve V_1 will have potentials decided by $-V_G$ plus the alternating voltage across half-secondary A_1C_1, whereas V_2 grid will vary in accordance with $-V_G$ plus the half-secondary voltage across B_1C_1. But the potential at point A_1 will be positive with respect to C_1 when the potential of B_1 is negative to the same extent. Therefore the grids of V_1 and V_2 will have potentials varying in anti-phase.

A decrease of the negative potential on V_1 grid will cause the anode current of V_1 to rise, giving a rise of potential across the half-primary of the output transformer C_2A_2, which is the anode-load presented to V_1. The potential of point A_2 must therefore decrease, since the potential of the centre point C_2 is constant at H.T. potential. This action will necessarily be accompanied by a

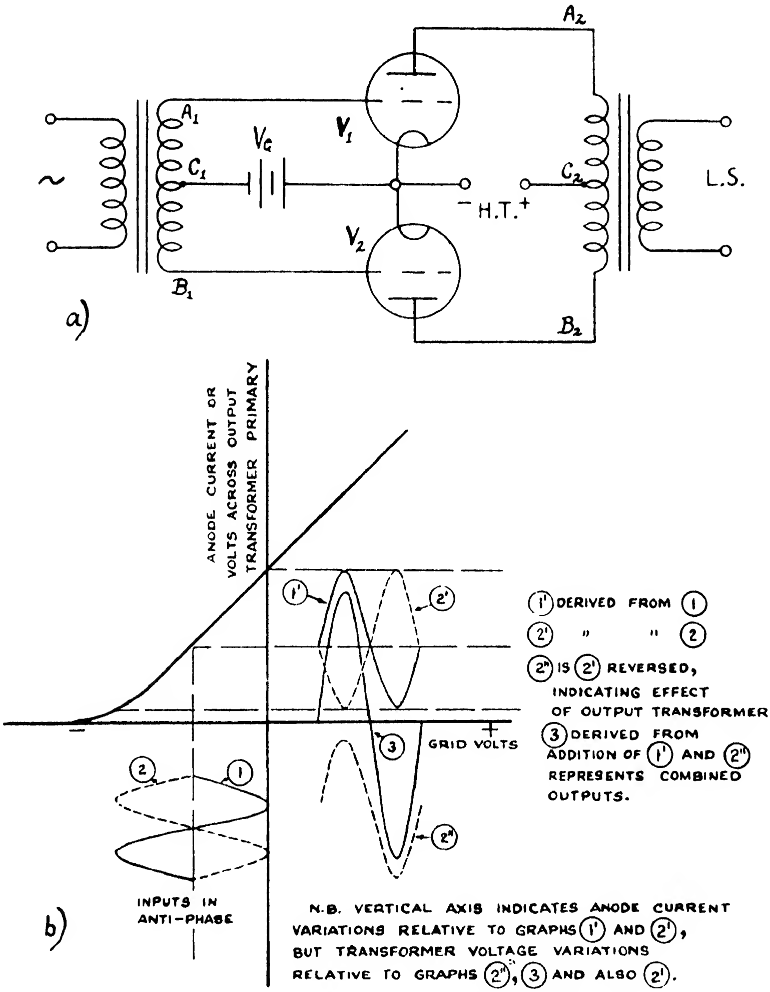


FIG. 65. *a*, Push-pull Circuit. *b*, Graphical Illustration of Push-pull Action.

corresponding increase of the negative potential of V_2 grid, producing a fall of the anode current of V_2. The voltage across the half-primary C_2B_2 of the output transformer must consequently decrease, implying an increase of the potential at point B_2, since the potential of centre-tap C_2 is fixed. The total potential variation across the whole of the output transformer primary A_2B_2 is then double the potential variation across either half, A_2C_2 or B_2C_2, since A_2 potential rises with respect to C_2 to the same extent to which B_2 falls. Hence the two valve outputs are effectively combined across the primary, giving a double output across the secondary, which is connected to a resistive load, such as a loudspeaker.

The D.C. components of the valve anode currents through the output transformer primary are in opposite directions, so the nett D.C. magnetisation of the transformer core is practically zero. The effect of D.C. magnetic saturation of the transformer core causing distortion is therefore eliminated.

A second, and more important, advantage of the push-pull technique is that second and other even harmonic distortion, due to lack of linearity of the valve characteristic over the operating region, is eliminated. Since triode valves produce distortion due to characteristic curvature which is almost exclusively at the second harmonic, so the grid bias may be set at the bend, or even at the valve cut-off region, giving class B working conditions with improved efficiency, and little distortion (see p. 151).

That even harmonic distortion is reduced to zero may be realised on considering a simple mathematical analysis of push-pull action. Suppose the relation between the valve anode-current I_A and the grid potential E_G is given by the general relationship

$$I_A = A + BE_G + CE_G^2 + DE_G^3, \text{ etc.} \qquad (183)$$

The alternating grid potential on one valve will be of the form $V_0 \sin \omega t$, whilst that on the other valve will be represented by the anti-phase voltage, $-V_0 \sin \omega t$, V_0 being the peak potential across half the input transformer secondary, $\omega/2\pi$ being the frequency of the input voltage.

Hence for valve V_1

$$I_A = A + B(V_0 \sin \omega t) + C(V_0 \sin \omega t)^2 + D(V_0 \sin \omega t)^3 + \text{etc.} \quad (184)$$

and for valve V_2, *which must have exactly the same shape characteristic*,

$$-I_A = A + B(-V_0 \sin \omega t) + C(-V_0 \sin \omega t)^2 + D(-V_0 \sin \omega t)^3 + \text{etc.} \quad (185)$$

Note I_A is written with negative sign here, since it is varying in opposition to the current from valve V_1.

The combined output across the output transformer primary is $2I_AZ$, where Z is the primary impedance, and $2I_A$ is given by subtracting equation (185) from equation (184).

$$\therefore \quad 2I_A = 2BV_0 \sin \omega t + 2C(V_0 \sin \omega t)^3 + 2E(V_0 \sin \omega t)^5 + \text{etc.}$$

Note that the squared terms, and other terms raised to an even power, vanish in this expression which determines the effective output, since $(-V_0 \sin \omega t)^2 = V_0^2 \sin^2 \omega t$. Moreover,

$$V_0^2 \sin^2 \omega t = V_0^2\left(\frac{1+\cos 2\omega t}{2}\right),$$

corresponding to a current component at the second harmonic of the input signal frequency. It is seen that such second, and other even harmonic distortion introduced by the curvature of the valve $I_A - V_G$ characteristics, are therefore eliminated.

It is noteworthy, however, that the percentage of third and other odd harmonic distortions remains the same as if a single valve were used. Since such third harmonic distortion is not noticeably introduced by triode valves, this push-pull method lends itself admirably to the design of receiver and other amplifier push-pull output stages where freedom from distortion is essential. Pentode valves, however, are prone to third harmonic distortion, and should therefore be used with caution in a low-frequency power amplifier push-pull arrangement.

Application of Feed-back to an Amplifier. If a fraction of the output voltage from a valve amplifier is fed back to be placed in series with the input circuit, then desirable or undesirable effects can be produced, depending on the phase of the feed-back voltage with respect to the input signal, and the purpose of the amplifier. If the feed-back is so arranged as to be in anti-phase with the initial input to the amplifier, then *negative* feed-back is achieved. A feed-back that is in the same phase as the input gives *positive* feed-back. Negative feed-back is also called "reverse" or "degenerative"; positive feed-back is also known as "regenerative", or "reaction".

If E=input voltage to the amplifier which is combined with a series feed-back voltage βE_0 derived from the output circuit (see fig. 66*a*), then the actual voltage to the amplifier input is

$$E_G = E \pm \beta E_0. \quad . \quad . \quad . \quad . \quad (186)$$

A plus or minus sign is attached to β depending on whether the feed-back is positive or negative.

The nominal gain, m, of the amplifier, is given by

$$m=\frac{\text{output voltage}}{\text{input voltage}}=\frac{E_0}{-E_G}\left(=\frac{\mu R}{R+R_A}\text{ if resistive anode load used}\right). \quad . \quad . \quad . \quad . \quad (187)$$

A negative sign precedes E_G, since the anode voltage of an amplifier varies in anti-phase with the input grid voltage (see p. 120).

$$\therefore \quad E_G=\frac{-E_0}{m}. \quad . \quad . \quad . \quad . \quad . \quad (188)$$

Substituting for E_G in (186) from (188)

$$\frac{-E_0}{m}=E\pm\beta E_0.$$

$$\therefore \quad E_0(1\pm m\beta)=-mE.$$

$$\therefore \quad E_0=\frac{-mE}{1\pm m\beta}. \quad . \quad . \quad . \quad . \quad . \quad (189)$$

The nett gain, including the effects of feed-back, m_f, is given by

$$m_f=\frac{\text{output voltage}}{\text{signal input voltage}}=\frac{E_0}{-E}.$$

Therefore from (189) $\quad m_f=\dfrac{m}{1\pm m\beta}. \quad . \quad . \quad . \quad . \quad . \quad (190)$

If $m\beta$ is made considerably greater than unity, then (190) becomes

$$m_f=\pm\frac{1}{\beta}\text{ approx.} \quad . \quad . \quad . \quad . \quad (191)$$

indicating that the overall gain of the amplifier depends only on the factor β, which is independent of the characteristic of the valve, the effects of valve "noise" and variations in supply voltage, but depends only on the feed-back network. The gain of the amplifier is, however, much reduced by such feed-back.

If $m_f=+1/\beta$, in the case of positive feed-back, then the feed-back voltage is in phase with the input voltage. Such an amplifier is generally unstable, since any increase of the output voltage

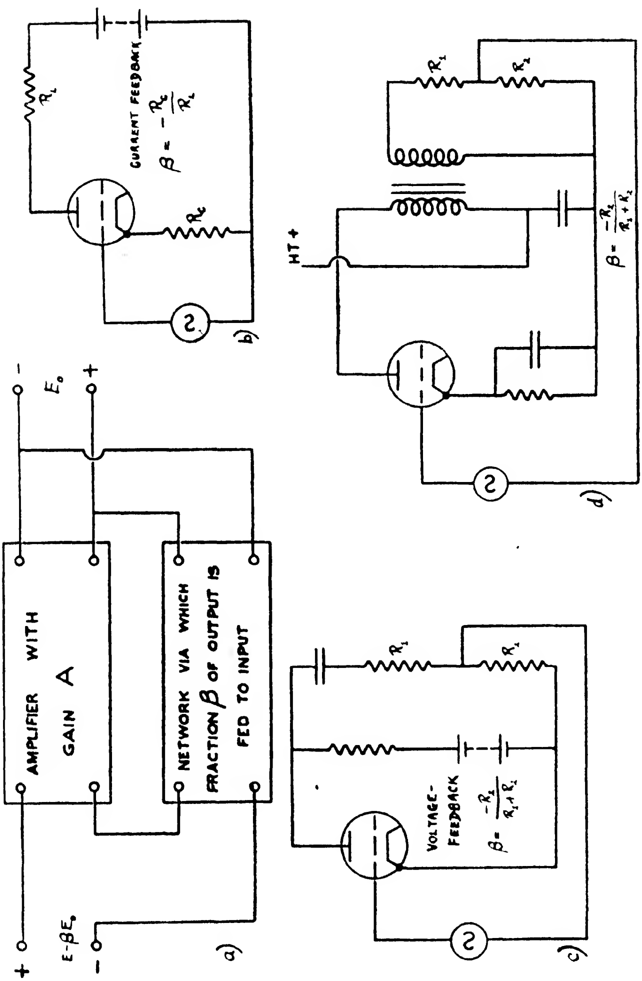


FIG. 66. Negative Feed-back Amplifier Circuits.

causes an increase of the total input signal, causing the output voltage to increase still further, the effect being cumulative.

If $m_f = -1/\beta$, in the case of negative feed-back, then the amplifier is much more stable than is the case of the amplifier without feed-back. This is particularly so the larger the value of $m\beta$, since then the actual input voltage to the amplifier is a small difference between comparatively large signal and feed-back voltages. If the gain m, due to the valve, changes, then the difference between signal and feed-back voltages changes, increasing if m decreases, and vice versa. The actual input voltage therefore changes in such a manner as to compensate for change of the gain m.

From equation (189) $\dfrac{E_0}{E} = \dfrac{-m}{1-m\beta}$.

The change in gain (E_0/E) caused by a change in amplifier nominal gain m is given by evaluating $\dfrac{d(E_0/E)}{dm}$.

$$\frac{d(E_0/E)}{dm} = -\frac{d[m/(1-m\beta)]}{dm}$$

$$= -\left[\frac{(1-m\beta)+m\beta}{(1-m\beta)^2}\right] = \frac{-1}{(1-m\beta)^2}. \quad . \quad . \quad (192)$$

Hence the change in gain with feed-back for a given change in gain without feed-back is decreased as β is increased negatively, i.e. as the negative feed-back is increased. Vice versa, if β is increased positively, in the case of positive feed-back, then the amplifier stability becomes poorer. This equation (192) indicates also how negative feed-back reduces amplitude distortion, since such distortion results generally from variations of m during the operating time.

The circuits involved in negative feed-back are of (*a*) the current feed-back type in which the voltage inserted at the input is proportional to the current in the load, (*b*) the voltage feed-back class where the voltage inserted at the input is proportional to the voltage across the load and (*c*) current-voltage feed-back in which a combination of (*a*) and (*b*) is used. These circuits are illustrated in fig. 66.

Since the gain of these amplifiers is decided by $1/\beta$, in the case where β is a considerable fraction ($m\beta > 10$), and the network whereby β is achieved may be other than a resistive arrangement, such as a combination of inductance and resistance or capacitance and resistance, so an amplifier of desired frequency-response can be designed, whereas if a resistive feed-back circuit is used, the gain is largely independent of frequency, except for the influence of stray inductance and capacity.

Noise reduction (see p. 164), is brought about in a negative feed-back amplifier. However, this reduction is only for that arising within the valve concerned. Any noise present at the input to the amplifier will be present to the same extent in the output. The reduction of noise as regards that introduced by the amplifier valve itself may be expressed as

$$\frac{\text{Signal to noise ratio with feed-back}}{\text{Signal to noise ratio without feed-back}}=\frac{m_f}{m(1-m\beta)}.$$

The Cathode Follower. If a negative feed-back amplifier is used in which the feed-back is obtained by the use of a resistance in the cathode circuit, as in fig. 67*a*, but the anode load is omitted, the anode being connected directly to the H.T.+ supply, the equation (190) $m_f=m/(1\pm m\beta)$ is suitably modified by putting $m=\mu R_C/(R_C+R_A)$, where μ and R_A are the constants of the valve used, R_C is the cathode resistance (cf. equation 166) and $\beta=1$.

$$\therefore \quad m_f=\frac{\mu R_C/(R_A+R_C)}{1+\mu R_C/(R_A+R_C)}=\frac{\mu R_C}{R_A+(1+\mu)R_C}. \quad . \quad (193)$$

The gain of such an amplifier is thus necessarily less than unity. It is therefore used as a current amplifier and not as a voltage amplifier, of which the purpose is to feed the grid with a high impedance input but in which the cathode load can be a low impedance output, yet without serious loss of voltage.

From equation (193),

$$I_A=\frac{\mu E_G}{R_A+R_C(\mu+1)}, \quad . \quad . \quad . \quad (194)$$

where E_G is the alternating grid input, and I_A the alternating anode current, R.M.S. values.

Dividing numerator and denominator of this expression by $(\mu+1)$ gives

$$I_A=\frac{[\mu/(\mu+1)]E_G}{R_A/(\mu+1)+R_C}. \quad . \quad . \quad . \quad (195)$$

6

By comparison with the circuit of fig. 49c, it can be seen that this cathode-coupled stage corresponds to a normal amplifier in which the amplification factor of the valve μ is reduced to $\mu/(\mu+1)$, and the A.C. resistance R_A is lowered to $R_A/(\mu+1)$, the equivalent circuit becoming that shown in fig. 67*b*.

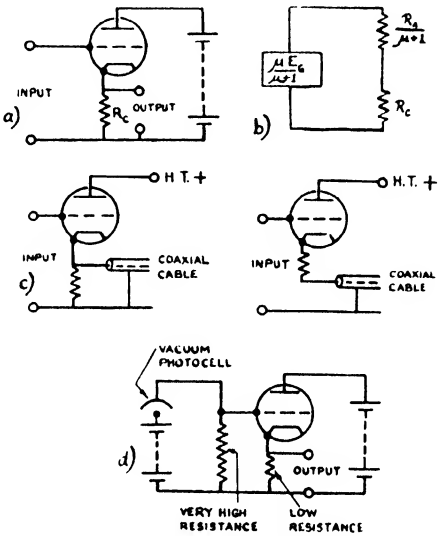


FIG. 67. *a*, The Cathode Follower. *b*, Equivalent Circuit for Cathode Follower. *c*, Cathode Follower used to couple to Transmission Line, *d*, Cathode Follower as D.C. amplifier.

The output resistance R_0 is effectively the valve resistance and R_C in parallel, and equals

$$\frac{R_C[R_A/(\mu+1)]}{R_C+R_A/(\mu+1)}=\frac{R_C R_A}{R_A+R_C(\mu+1)},$$

which in the case of a pentode,* where $\mu \gg 1$, becomes

$$R_0=\frac{R_C}{1+R_C g_m}, \quad . \quad . \quad . \quad . \quad (196)$$

on employing the relationship $g_m=\mu/R_A$. For example, in the

* In using a pentode in a cathode-follower circuit, the screen must be kept at constant potential with respect to the cathode.

case of a pentode, suppose $\mu=1000$, $R_A=250$ kΩ., $R_C=1000$ Ω. and $g_m=4$ mA./V.$=0{\cdot}004$ ohms, then

$$R_0=\frac{R_C}{1+1000\times 0{\cdot}004}=\frac{1000}{5}=200\,\Omega.$$

$$\text{and the gain}=\frac{\mu R_C}{R_A+R_C(\mu+1)}=\frac{1000\times 1000}{250{,}000+1000(1001)}=0{\cdot}8.$$

This low impedance output, one side of which is at earth potential, makes such circuits eminently suitable for coupling the output of an oscillator to a transmission line or aerial circuit in a transmitter, where a careful match is required between the output circuit and the load (see fig. 67*c*). Moreover, because of the low output impedance, the output voltage has good regulation where the input to the cathode-follower stage may have poor regulation. Another advantage of this circuit is that there is no polarity inversion: the output voltage variation is in phase with the input alternating voltage.

The cathode follower circuit also exhibits reduced Miller effect (cf. p. 134), the effective input capacity being reduced in accordance with the relationship

$$\text{Equivalent input capacity} = \text{actual input capacity } (C_{GC})\times\left(1-\frac{E_0}{E_G}\right).$$

A cathode follower may be used with advantage in photoelectric cell amplifiers, since this circuit technique enables the grid resistance across the input to be very high, yet with a low output cathode resistance. Thus using ordinary valve types, the presence of positive ion current and grid emission prohibits the effective use of an input resistance greater than 2 MΩ. (see p. 119). Suppose the photocell current is $0{\cdot}05\,\mu$A., then a D.C. voltage input to the amplifier of $0{\cdot}05\times 2=0{\cdot}10$ V. is achieved on switching on the cell illumination. If the D.C. amplifier is normal (fig. 48) then a voltage gain of 100 times, and so an output of 10 V. is easily obtained. Compare the use of a cathode-follower amplifier. Even using normal triode and pentode valve types, the input resistance can now be as much as 100 MΩ., giving an effective input voltage of $0{\cdot}05\times 100=5$ V. If a gain of 0·8 is realised, then the output voltage across, say, a 5 kΩ. cathode load will be 4 V. It would seem as though the cathode technique was thus only 40% as effective as the provision of the ordinary D.C. amplifier. In practice, however, great benefit is obtained from the use of the cathode-follower because it is virtually insensible to the supply voltage fluctuations and circuit variations which make normal D.C. amplifier practice so tedious and unreliable.

"Noise" in Amplifiers. It would seem that a consideration of the methods of coupling both A.C. and D.C. amplifiers in cascade would lead to the belief that there was no limit to the total amplification possible: it was simply a matter of ensuring constant voltage supplies and using a sufficient number of valve stages coupled together to obtain amplifications of several million. In practice this is not so: a limit is set by the "noise" which arises in the amplifier, due to minute random fluctuations of the current. These are particularly obnoxious in the first stage, since it is there that the true signal is small, and the "noise" becomes of comparable magnitude, giving a signal to "noise" ratio which is sufficiently small to make the eventual output unintelligible.

This "noise" arises in two places:

(*a*) Thermal agitation noise, or Johnson noise, brought about by fluctuations of the current in ordinary electrical conductors due to the thermal agitation of the conducting particles. In this connection the input resistance to the first amplifier is usually the only one which need be of concern, because it is only there that the signal current is sufficiently small for the "noise" current to be of comparable magnitude.

An electric current is due to the motion of electrons through conducting material, and the current due to an applied signal voltage is combined with the small currents due to the thermal agitation of the electrons. During any long period of time the nett effect of these thermal currents is zero. At any given instant, however, this is not the case—there may be a slight preponderance of random electron motion in one direction over that in the other. Such transient currents are of very complex wave-form, and produce voltages across the resistance of which the harmonic components extend over a very wide frequency band.

An expression for thermal "noise" is

$$E_N{}^2 = 4kT\int_{f_1}^{f_2} R\,df. \qquad (197)$$

Due to Nyquist,* this equation gives E_N, the R.M.S. value of the E.M.F. produced in series with the resistance of the conductor R, where k is Boltzmann's constant, T is the absolute temperature, f is the frequency and R is the actual or equivalent shunt resistance of the circuit, being L/CR in the case of a rejector circuit.

Integrating this equation over the frequency range (f_1-f_2) gives

$$E_N=\sqrt{[4kTR(f_1-f_2)]}. \quad . \quad . \quad . \quad (198)$$

Substituting for k, then at normal room temperatures

$$E_N=1{\cdot}25\times10^{-10}\sqrt{[R(f_1-f_2)]}. \quad . \quad . \quad (199)$$

An immediate partial remedy to the elimination of this noise voltage is to restrict the band width (f_1-f_2) which the amplifier passes. If $R=100$ kΩ., and $(f_1-f_2)=10$ kc./s., then substitution in (199) gives

$$E_N=1{\cdot}25\times10^{-10}\sqrt{(10^5\times10^4)}=4\ \mu V.\text{ approx.}$$

(*b*) Shot or "schrot" noise. First investigated by Schottky,† is due to a comparable fluctuation of the current in valves due to the random emission of individual electrons. Thus the anode current of a valve is constant considered over any finite time interval, but at a particular instant the number of electrons arriving may be slightly different from the number at another instant, though the average number is constant. In other words, "shot noise" is due to the finite magnitude of the electron charge motions which constitute an electric current, where the number of electrons flowing per second will be constant, but where the number per microsecond will fluctuate slightly about the mean value.

Added to this inevitable cause of current variation, there is a second source of fluctuation which is decided by the physical nature of the emitting cathode surface. This is called "flicker effect", and is particularly prevalent if the cathode surface is not smooth, but pitted with minute holes which occasionally release bursts of electrons. Added to this there are noise effects due to ionisation of the residual gas in the valve and due to secondary emission from the valve electrodes.

* H. Nyquist, *Phy. Rev.*, **32**, 110, 1928.
† W. Schottky, *Ann. der Physik*, **57**, 541, 1918.

Shot noise, considered as separate from flicker effect, etc., can be expressed by the equation

$$I_N^2 = 2I_A e(f_1 - f_2) \qquad . \quad . \quad . \quad (200)$$

in the case of a saturated diode, where I_A is the anode current, and I_N is the R.M.S. value of the variation of I_A.

In the case of non-saturated triode and multi-electrode valves, the effect is reduced by the presence of space-charge, which acts as a cushion, or reservoir of electrons. Thus valves in which the space-charge effect is high are preferable for reduction of shot noise. In these cases, the effect of total noise is stated as the equivalent resistance which would give the same noise voltage due to thermal effects, and the magnitude of this resistance is proportional to I_A/g_m^2. Hence a valve of large mutual conductance g_m for a given anode current I_A, is necessary for low noise levels.

The triode valves are most free from this noise defect. The pentode has an equivalent shot noise resistance about five times that for the corresponding beam tetrode. This is because the beam tetrode has low screen current, high space-charge effect between beam plates and low secondary emission. Frequency-changer valves (see p. 210) are bad because their conversion conductances are low, and they also produce noise by virtue of their oscillator section. Thus it is preferable to use an R.F. beam tetrode amplifier before the frequency-changer in a superheterodyne receiver if a particularly high signal/noise ratio is required.

A minimum signal/noise ratio of 5 to 6, or some 15 db.* is desirable in amplifier and receiver practice, where the noise concerned is the total noise brought about by the above effects.

Supply Voltage Regulation. If electronic circuits, such as amplifiers, are used in any kind of measuring or indicating device where a constant output reading is required from day to day for the same input conditions, then it becomes essential to pay considerable attention to the question of the supply of constant valve filament and anode voltages. This is particularly the case where the H.T. and L.T. supplies depend ultimately on the A.C. or D.C. mains, as in power-pack practice.

* Decibels = db. = 10 $\log_{10}(P_1/P_2)$, where P_1 and P_2 are two powers,
= 20 $\log_{10}(V_1/V_2)$, where V_1 and V_2 are two voltages.
∴ 15 db. = 20 $\log_{10}(V_1/V_2)$. ∴ V_1/V_2 = antilog (0·75) = 5·63.

Apart from manual control of a potentiometer, or variac across the mains, bringing an indicating voltmeter to a constant reading, there are four ways in which stabilisation of a voltage (or current) supply can be arranged. They are (*a*) by means of a constant-voltage transformer; (*b*) using a barretter; (*c*) by gas-filled discharge tube, or stabilovolt; (*d*) by use of thermionic vacuum tubes. In all these devices an unavoidable difficulty arises in that the voltage (or current) change must occur before it can be compensated; it is a desirable feature of a stabiliser, therefore, that such compensation takes place with as little time-lag as possible. In this connection, and in other ways, the use of thermionic valves as regulating means are the most successful.

(*a*) Constant voltage transformers employ magnetic saturation to achieve a constant A.C. voltage output. Thus transformers are available with ratings up to 25 kW. which give a secondary voltage constant to within $\pm 1\%$ for variations of primary voltage up to $\pm 15\%$. In such transformers the usual arrangement is for two transformers to be used with their primary windings in series additively, and their secondary windings in series with the load, but so wound that the A.C. voltage outputs are in anti-phase with each other. One of these transformers has an iron core which is partially magnetically saturated over the range of primary voltage required, and gives the larger secondary voltage. When the A.C. voltage applied increases in magnitude the fraction of the total applied to the saturated transformer decreases. Thus the secondary voltages become more nearly equal, it being a matter for the designer to arrange that such change of output compensates exactly for the primary supply change. Frequently the saturated transformer has a condenser connected across its secondary to improve the regulation, and to correct the wave-form. The chief difficulty with such arrangements is that the output wave-form is no longer sinusoidal, and in some circuits the change of wave-shape with change of supply voltage may be as disturbing as a change of voltage would be. Again, if the frequency of the A.C. mains supply varies, then the voltage variations from such transformers are worse than from an ordinary transformer. Transformers are available, however, which are guaranteed to be unconscious of both primary voltage and frequency changes within certain limits.

A constant voltage transformer feeding the primary of an ordinary transformer power pack can regulate the ultimate H.T. supply from such a pack. Again, these transformers can be used for maintaining constant A.C. heater voltages, or lamp filament supplies.

(*b*) The barretter is somewhat like an ordinary electric lamp of which the filament is iron wire operating in a hydrogen atmosphere. The filament and hydrogen pressure are so adjusted that the current through the lamp is largely independent of the supply voltage. The commonest use of these barretters is in universal A.C.-D.C. mains receivers where a barretter is placed in series with all the valve heaters across the A.C. mains supply, fig. 68, barretters being available which have a current capacity equal to the current rating of the 13 V. heaters commonly employed in A.C.-D.C. valves.

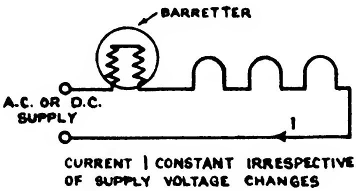


FIG. 68. The Barretter.

The iron filament presents a resistance to the supply E.M.F. which varies in such a manner that the series current is constant. Thus if the supply E.M.F. increases the current will tend to rise. However, such current increase causes greater heating of the filament, and if the filament resistance temperature coefficient is correct, then the increased filament resistance can compensate for the increased supply voltage. This coefficient is arranged to be correctly compensative over a wide range of voltages by the use of iron as the resistance element, in conjunction with the rate of conduction of heat away from the filament depending on the pressure of the surrounding hydrogen atmosphere.

The time lag for compensation to occur is a snag in using this method of stabilising. Again, the glass bulb of the barretter takes more than half an hour to reach an equilibrium temperature. Apart from its great use in providing valve heater currents in universal receivers, the author has found that barretters are of little use in trying to arrange constant current supplies for any purpose where laboratory measurements are to be made.

(*c*) Gas-discharge tube regulators, such as the neon lamp and the specially prepared stabilovolt, are of considerable value in achieving constant H.T. supplies. This is due to the peculiar characteristic of the two-electrode tube filled with an inert gas at low pressure whereby, once a glow discharge begins, the voltage across the tube remains nearly constant irrespective of the current through the gas.

The current through the tube increases with voltage until the striking potential is reached. Thereafter the voltage drop across the tube falls to a lower value, and further current increase is accompanied by practically no voltage change. The cathode glows, and as the current is raised, this glow extends to cover the whole cathode area.

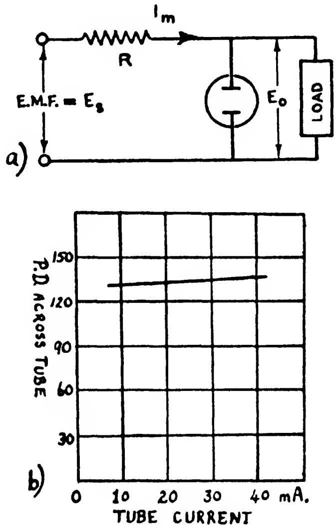


FIG. 69. *a*, Gas-filled Voltage Regulator. *b*, Voltage-current Characteristic.

The power supply voltage needs to be more than the striking potential. The circuit used is shown in fig. 69*a*, constant voltage being obtained across the load shown. The series resistance R is necessary to prevent exceeding the tube current rating. $R=(E_S-E_O)/I_m$, where E_S is the total E.M.F., E_O is the load voltage, and I_m the maximum tube current. Any increase of current through the load, due to fall of load resistance, is accompanied by a tendency to greater voltage drop across R. This tends to reduce the voltage across the tube and load, but less current is then taken by the tube, compensating for the increase of load current, so that the tube and load voltage remain constant, because the voltage drop across R remains the same. Again, if the supply voltage increases, the additional current through the regulator tube will be sufficient to make the voltage drop across R compensate for the change, so the load voltage remains the same.

These tubes are especially useful in the form of the stabilovolt in which the cathode is in the form of a number of metal cylindrical electrodes, one inside the other and insulated from one another. With appropriate external resistances, such an arrangement can be used to provide a total constant voltage of some 300 V. or more which can be subdivided into three or four parts as a potential divider. The alternative is to use a number of neon tubes, or preferably stabilisers like the Cossor S130, in series with one another.

Electronic Regulators. The best types of voltage and current regulators are those employing valves. Whereas a gas-discharge tube cannot readjust itself in less than 1/10th msec., an electron tube is practically instantaneous in its action. The chief disadvantage associated with the use of valves is in obtaining large current values. Thus a regulated current of 0·5 amp. would either require the use of a large type of power valve, or a number of normal valve types in parallel.

A saturated diode, with its constant anode current above saturation anode potential can be used. However, valves with oxide-coated cathodes do not operate for long under saturation current conditions. The bright emitter tungsten filament could be used, but then the provision of a carefully regulated filament current is a problem.

The basic circuit employed using a triode valve regulator is shown in fig. 70*a*. If the H.T. supply voltage rises, the voltage across the load tends to rise, and so does the cathode potential of the triode valve in series with the load. With a given bias, obtained as shown, an increase of H.T. is accompanied by a tendency to a positive increase of cathode potential, and as a result an increase of negative grid bias relative to cathode. So the valve current is reduced, and the voltage drop across it rises, which can be made to compensate for the increased supply volts, the load voltage remaining constant if correct working conditions are arranged. The opposite action takes place if a fall of supply voltage occurs. Similarly, any undesirable changes of the load current are regulated. A resistance is inserted in series with the valve grid to prevent excessive grid current flow when the H.T. supply is switched off, and the grid goes positive because of the continued supply from the bias battery. Alternatively, a gas-discharge tube can be used to provide the grid bias, as in fig. 70*b*.

A mutual-conductance, or g_m-regulator can be arranged as in fig. 70*c*. Here the H.T. supply E.M.F. to be regulated is placed across a potential divider, R_1 and R_2 in series. Let the total H.T.$=E$ volts, and let $E_L=$voltage across the load supplied. Suppose E changes by an amount ΔE, where ΔE is positive or negative. Then the change of grid voltage is $R_1 \Delta E/(R_1+R_2)$.

This bias change will produce a change of anode current through R_3 of $g_m R_1 \Delta E/(R_1+R_2)$, where g_m=valve dynamic mutual conductance. If this is equal to the change ΔE of the supply E.M.F.,

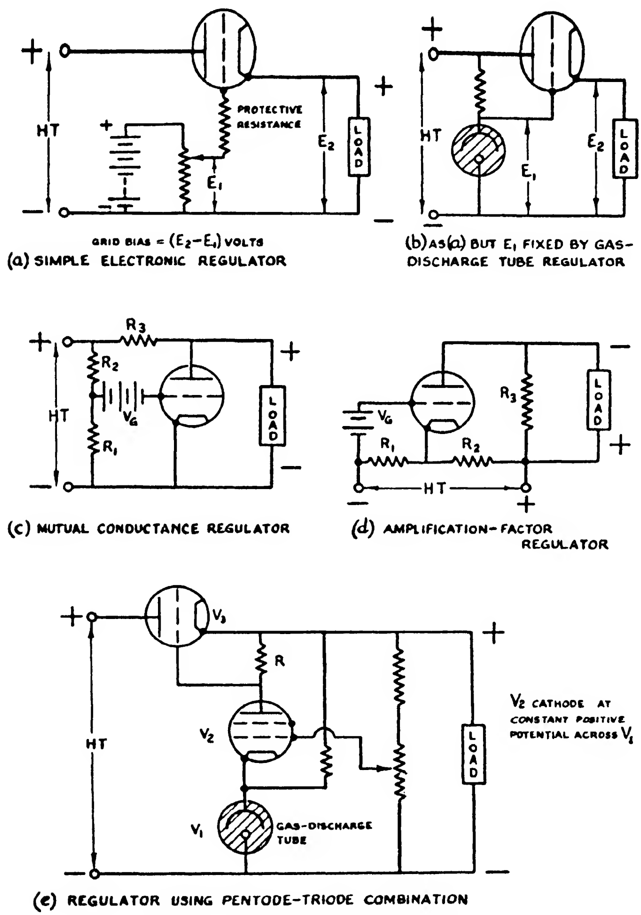


FIG. 70. Electronic Regulator Circuits.

then the anode voltage, and so load voltage, will remain constant, because then an increase of E is accompanied by an equal increase of the P.D. across R_3, and correspondingly for a decrease of E there is an equal decrease across R_3, whilst the anode voltage is equal to the difference between E and the P.D. across R_3. Hence for effective stabilisation of the load voltage

$$\Delta E = g_m \frac{R_1 R_3}{R_1 + R_2} . \Delta E.$$

$$\therefore \quad g_m = \frac{R_1 + R_2}{R_1 R_3}.$$

The actual values of R_1, R_2 and R_3 chosen will depend on the valve characteristics considered in conjunction with the load E.M.F. and current required. The grid bias relative to the cathode will have a mean value of $[R_1/(R_1+R_2)E - V_G]$, which decides the value of the bias battery voltage necessary to achieve a suitable negative bias for class A operation of the valve.

An amplification-factor, or μ-regulator, is shown in fig. 70*d*. Here a change of E by an amount ΔE causes a grid voltage change of $R_1 \Delta E/(R_1+R_2)$ and an anode voltage change of $R_2 \Delta E/(R_1+R_2)$. Since the change of grid volts is here necessarily of opposite polarity to the change of anode volts, so the anode current through R_3, and hence the P.D. across R_3 and the load will remain constant, if the effect on the anode current of the grid voltage change is equal and opposite to the effect on the anode current of the anode voltage change. But in accordance with the definition of amplification-factor, this will be the case if the change of anode voltage divided by the change of grid voltage equals μ. Therefore the circuit is an effective regulator when

$$\frac{R_2 . \Delta E}{R_1 + R_2} \div \frac{R_1 \Delta E}{R_1 + R_2} = \mu.$$

$$\therefore \quad \frac{R_2}{R_1} = \mu.$$

The ratio of R_1 to R_2 is therefore decided. The values of R_1, R_2 and R_3 will depend on the valve chosen, and the demands of the load. To operate the valve at its usual grid bias for class A working, it is to be noted that the grid potential relative to the cathode is $[V_G - R_1 E/(R_1+R_2)]$.

Regulators using pentode valves with their constant-current characteristic and high amplification factor are capable of very great control ratio. A typical circuit is shown in fig. 70*e*. The bias on V_3 depends on the P.D. across *R*, which rises negatively if the current through *R* rises, and decreases if the current through *R* falls. But the rise and fall of the current through *R* depends on the grid bias on V_2, which will change in direction in the same way as any change of load voltage. So a tendency for the load voltage to increase, sensitively controls the increase of negative bias on V_1, which makes the A.C. resistance of V_3 greater. The P.D. across V_3 therefore rises, which can be made to compensate exactly for any tendency for the load voltage to increase. Vice versa, a tendency for a decrease of load voltage to occur is offset.

CHAPTER 7

Valve Oscillators

The Maintenance of Oscillations in the Freely Oscillating Circuit by the Use of a Valve. In Chapter 4, p. 61, it was considered that if a charged condenser were allowed to discharge through an inductance L, then provided that $R^2/4L$ was less than $1/LC$, where R was the total effective resistance introducing energy loss by heat and radiation in the circuit, then an oscillatory current was obtained in the circuit, which, however, decayed in amplitude because of heat and radiation losses at a rate governed by the magnitude of $R/2L$.

To sustain oscillations in such a circuit use can be made of the amplifying property of a triode valve, so that the energy loss per cycle during the oscillations is restored. This energy fed to the circuit must be arranged in accordance with two conditions:

(1) That it is done at the correct rate; i.e. at a frequency decided by the oscillatory circuit itself.

(2) That the current impulses fed to the LC circuit have a component of sufficient amplitude in phase with the existing current oscillations in the circuit.

The Tuned Anode Reaction Oscillator. A commonly used circuit in which these provisos are arranged is shown in fig. 71*a*. Here LCR forms the basic freely oscillating circuit. This is placed in the anode circuit of a triode valve of amplification factor μ, and A.C. resistance R_A. On switching H.T. supply on to the valve the condenser C becomes charged, and then commences to discharge in an oscillatory fashion through the inductance L. Some of the alternating energy associated with the magnetic field around L is coupled by mutual induction to the grid coil L_1. The consequent alternating potentials across L_1 render the valve grid oscillatory in potential, and, by virtue of the mutual conductance action of the valve, the anode current pulses fed back into the LC circuit are likewise oscillatory. It is readily seen that in this manner requirement (1) is satisfied, that the feed-back energy should occur at the correct rate; it remains to be seen under what conditions the requirement (2) as regards phase is also fulfilled.

To work out the theory of this oscillator, two assumptions make the analytical work involved far less complex. Firstly, suppose

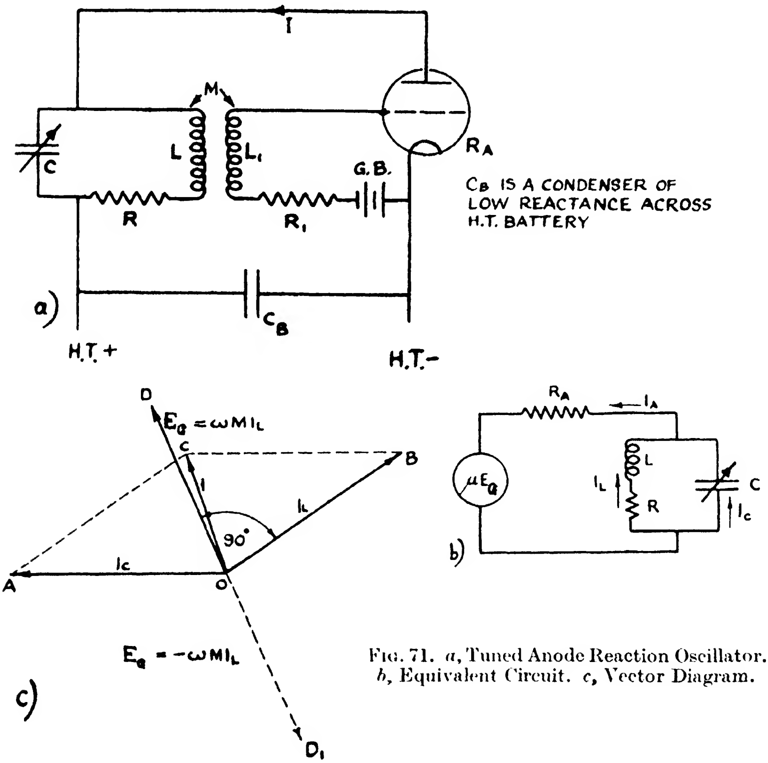


FIG. 71. *a*, Tuned Anode Reaction Oscillator. *b*, Equivalent Circuit. *c*, Vector Diagram.

the valve is biased in such a way, and accommodates such an alternating input that it is operating as a linear amplifier; secondly, the grid bias is sufficiently negative to prohibit the flow of grid current, and that the input impedance of the valve is infinite.

An equivalent electrical circuit to that of fig. 71*a* can then be considered in fig. 71*b*.

Here the amplifying action of the valve is considered by realising that an A.C. voltage of E_G on the grid is equivalent to one of μE_G on the anode provided the valve resistance of R_A is included (see p. 122). Let I_A be the R.M.S. value of the alternating component of the anode current; whilst I_L and I_C are correspondingly the currents through L and C. Equating the effective A.C. supply voltage to the sum of the voltages across the valve resistance and the coil,

$$\mu E_G = I_A R_A + I_L(R + j\omega L). \qquad (201)$$

But E_G is derived by mutual induction M from current I_L in the coil L.

$$\therefore \quad -E_G = j\omega M I_L, \qquad (202)$$

where ω is the current pulsatance.

Note that E_G is here prefixed by a negative sign, because variations of grid voltage E_G are in antiphase with variations in anode voltage μE_G (see p. 119).

Substituting for E_G from (201) in (202)

$$-\mu j\omega M I_L = I_A R_A + I_L(R + j\omega L). \qquad (203)$$

To eliminate I_A and I_L from this equation, consider that

$$I_A = I_L + I_C. \qquad (204)$$

Moreover, the A.C. potential across L and R must equal that across C in the oscillatory circuit.

$$\therefore \quad I_L(R + j\omega L) = I_C \cdot \frac{1}{j\omega C}. \qquad (205)$$

$$\therefore \quad I_A = I_L + I_L \cdot j\omega C \cdot (R + j\omega L) \qquad (206)$$

on replacing I_C in (204) from (205).

Substituting from equation (206) for I_A in equation (203)

$$-\mu j\omega M I_L = I_L\{1 + j\omega C \cdot (R + j\omega L)\} R_A + I_L(R + j\omega L).$$

Now divide through by I_L, and the relationship between component values, and valve constants which must be satisfied in the oscillator circuit is

$$-\mu j\omega M = R_A + j\omega C \cdot RR_A - \omega^2 LCR_A + R + j\omega L. \qquad (207)$$

Equating real, or resistive terms in this equation

$$R_A - \omega^2 LCR_A + R = 0.$$

$$\therefore \quad \omega^2 = \frac{R_A + R}{LCR_A}.$$

$$\therefore \quad \omega = \frac{1}{\sqrt{(LC)}} \sqrt{\left(\frac{R_A + R}{R_A}\right)}. \qquad . \quad . \quad (208)$$

This equation gives the pulsatance ω, and hence the frequency $\omega/2\pi$ at which this valve circuit oscillates.

Note that this frequency is affected by the ratio R/R_A. If the valve A.C. resistance changes due to a change of grid or anode voltage, a change in valve emission, or substitution of another valve, then the circuit oscillatory frequency will alter. Likewise, a change of L, C or R due to mechanical or thermal variations will alter this frequency. These factors need to be borne in mind in designing such an oscillator to have a constant frequency. In particular, the ratio R/R_A should be as small as possible. Thus resistance loss in the circuit needs to be small, whereas the valve resistance should be large.

Equating imaginary or reactive terms in equation (207)

$$-\mu\omega M = \omega CRR_A + \omega L;$$

divide through by ω,

$$\therefore \quad -\mu M = CRR_A + L.$$

$$\therefore \quad M = -\left(\frac{CRR_A + L}{\mu}\right). \qquad . \quad . \quad . \quad . \quad (209)$$

This equation gives the condition that must be fulfilled if the feed-back energy is sufficient to overcome the losses represented by R, so that oscillations are sustained. If this equation is exactly fulfilled, then the feed-back energy will just be sufficient, so oscillations of constant amplitude and good wave-form will obtain. Note that if excessive mutual inductance M is to be avoided, then $R_A/\mu = 1/g_m$ needs to be small, and also μ needs to be large. Generally, for efficient and stable operation of such an oscillator, a valve of high g_m, and high A.C. resistance is necessary. Since in valve design technique (p. 108), these two requirements are somewhat in conflict with one another, so a compromise must needs be effected.

If the amount of mutual induction used is smaller than that necessary to maintain oscillations, then the circuit will not oscillate continuously, but will merely undergo a reduced decrement, $R/2L$. In this case, the effective resistance of the LCR circuit is reduced, but not to zero. This principle is applied in the use of reaction in radio receivers, where the resistance of a tuned circuit is reduced by feed-back from the anode circuit of a valve (usually the detector). This reduction of resistance brings about an increased "Q" factor of the tuning coil, with improved voltage magnification and selectivity (p. 55).

On the other hand, if the feed-back in the oscillator circuit is in excess of that stipulated by equation (209), then the effective resistance of the LCR circuit is made negative. Now the oscillatory current amplitude will grow until it is finally limited by the valve used. This effect is likely to lead to excessive harmonic distortion in the oscillator output, since the current wave-form is distorted due to gross departure from linear operation of the valve under these limiting conditions.

It remains to consider whether the requirement as to correct phase of the feed-back energy has been observed. Considering the vector diagram, fig. 71c, OA represents the condenser current I_C, whilst OB represents the inductance current I_L; these are not quite in anti-phase because of the resistance of the inductance. The resultant vector OC* obtained by completing the parallelogram $OACB$ gives the phase of the anode current I_A, which "makes-up" the current required to sustain oscillations (compare the theory of the rejector circuit, p. 57). The grid voltage variations, $E_G = \omega M I_L$, will lead or lag by 90° on the inductance current I_L depending on whether the mutual induction M is positive or negative. In the one case the E_G vector will be OD, practically in phase with the anode current variations I_A. This is the necessary condition for correct operation of the valve as an amplifier, since grid voltage increase is accompanied by an in-phase anode current rise. On the other hand, if the direction of the winding of either coil is reversed, then the mutual inductance effect will be reversed in sign. Now the grid voltage variations will be represented by a vector OD_1, in anti-phase approximately with the current I_A. This will lead to negative feed-back, or reverse reaction, resulting in an increased decrement of the damped oscillations occurring in the LCR circuit. The resistance of the oscillatory circuit is then effectively increased, so that oscillations die away more rapidly

* Erratum: The vector OC in Fig. 71c should be marked I_A not I.

than in the freely oscillating case. This procedure is sometimes of advantage, as in the case of the neutralising of a radio-frequency power amplifier to prevent it oscillating.

Self-bias of a Valve Oscillator. There are several alternative valve circuits used for sustaining oscillations in a valve oscillator. In most of these self-bias is used. This will be understood by considering it as an addition to the tuned-anode oscillator as in fig. 72.

On p. 133 it was considered that the presence of a condenser with parallel resistance in the grid circuit of a triode valve, provided with an A.C. input, brought about a negative bias on the

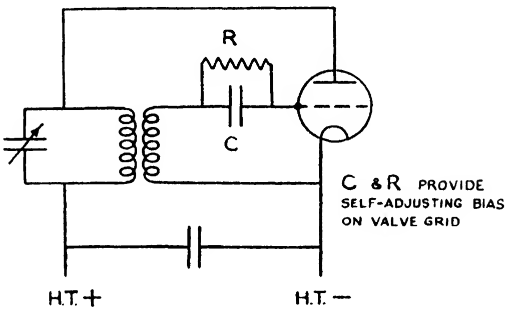


FIG. 72. Use of Self-bias in an Oscillator.

grid which was a fraction of the input peak A.C. voltage depending on the values of C and R. On p. 151 it was shown that a power amplifier operating with a tuned circuit as anode load was most efficient if the negative D.C. grid bias was beyond cut-off or class C operated. It would be desirable, therefore, in order to arrange an efficient oscillator, that its mean negative bias be beyond cut-off. If such bias were provided by a battery then, on applying H.T. to the circuit, there would be no anode current so that oscillations would not begin. If automatic grid bias is provided by a CR arrangement in the grid circuit, on the other hand, then the bias is zero until oscillations commence. Therefore, on switching on the H.T. supply, anode current will flow through the valve, and oscillations can begin. Such oscillations will build up, producing negative grid bias via CR, until an equilibrium is attained which, provided that the C and R values are aptly chosen, can be for class C operation leading to a highly efficient oscillator.

This common method of arranging bias in an oscillator means that the mathematical analysis of the tuned-anode oscillator already considered is greatly over-simplified. Nevertheless, the results obtained for the oscillation frequency, and for the feed-back necessary, though approximate, are useful, and indicate the effect of the alteration of component values.

The values of C and R used for automatic grid bias in an oscillator are important. The product CR must be sufficiently large to maintain the grid-bias reasonably constant throughout the cycle, i.e. CR must be long compared with the oscillation period. If the time-constant CR is too long, then an effect known as "squegging" or "grid blocking" occurs. This is usually avoided, though such a "squegging" or "blocking" oscillator has been applied in pulse producing circuits, particularly in radar practice.

The manner in which such a block effect occurs can be understood on considering the following sequence of events. Suppose the time constant CR is too long. Then the negative grid bias developed is so great that the power supplied by the valve to the oscillatory circuit will be temporarily insufficient to maintain oscillations. Consequently, oscillations cease, so does the alternating potential input to the grid, and therefore the grid bias falls towards zero at a rate decided by CR. Very soon the bias will be low enough to allow oscillations to recommence, but then the cut-off action results again. Obviously the result will be an output from the oscillator in the form of pulses of alternating energy, each pulse being separated by a short period of no action.

The use of automatic bias improves the oscillator stability. Consider that for some reason, say a temporary drop of H.T. supply, the oscillation amplitude tends to fall. This effect will be cumulative, since the reduced feed-back will cause a loss of grid excitation, and so still smaller amplitude. The oscillator is then in danger of ceasing altogether. If the bias is obtained by CR, however, then an initial tendency for the oscillation amplitude to fall will automatically produce less grid bias, and so greater amplification: the sequence of events leading to failure to oscillate is avoided.

Other Reaction Oscillators. Instead of using a tuned LC circuit at the anode of an oscillator employing mutual induction to obtain feed-back, the tuned circuit can be associated with the grid, giving the tuned-grid oscillator. This circuit gives less power output, other conditions being equal, than does the tuned-anode

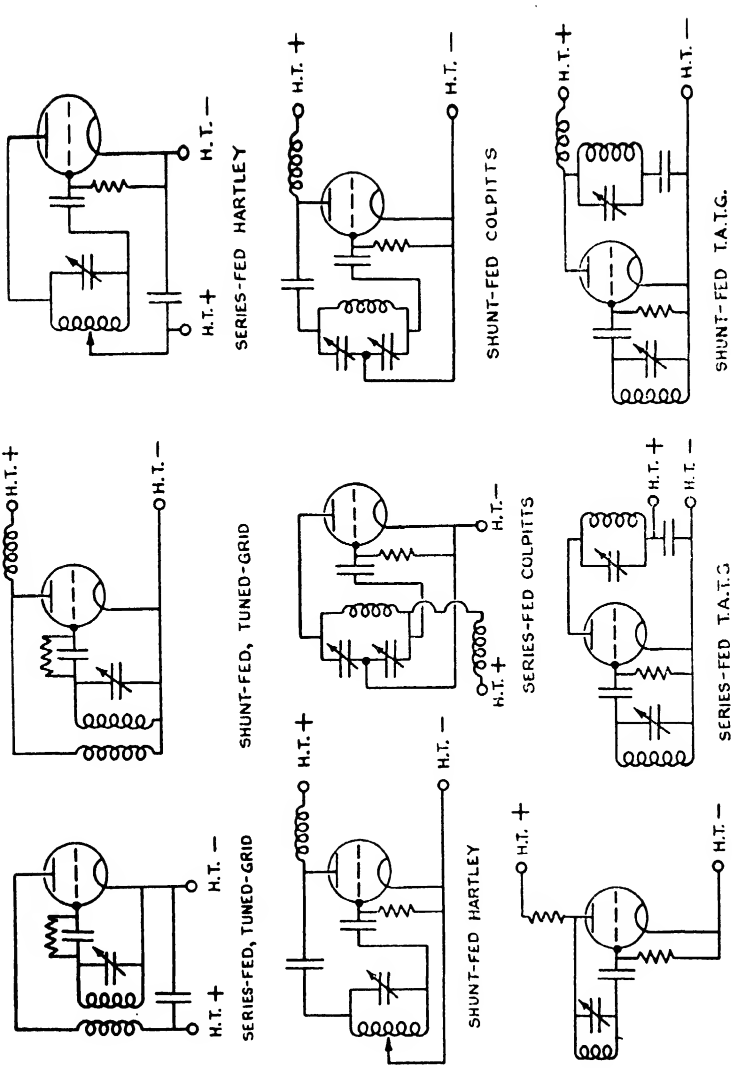


Fig. 73. Some Valve Oscillator Circuits.

type, but its frequency stability is greater, and it is therefore especially favoured as an oscillator for the frequency-changer stage of a superheterodyne receiver.

An efficient and popular type of oscillator for audio and medium-high frequency work is the Hartley. In this circuit the grid alternating potential is obtained in direct fashion by tapping off about one-third of the total alternating potential developed across the tuned-anode circuit inductance.

The Colpitts oscillator is a somewhat similar arrangement. In this case the grid alternations are arranged by tapping the capacitative arm of the tuned-anode circuit. The necessary division of potential across this arm is achieved by using two condensers in series to form the total capacity deciding the frequency, whilst the grid potential is derived from the potential variations across one of these condensers. The ultra-audion short-wave oscillator is essentially a Colpitts circuit in which the valve inter-electrode capacities, C_{GC} and C_{GA} form the voltage dividing network.

A tuned-anode, tuned-grid oscillator is especially favoured by workers in the short-wave field. This oscillator, containing an *LC* circuit associated with both grid and anode, is capable of efficient working up to 60 Mc./s.

Both the grid and anode circuits are tuned to a frequency slightly higher than the operating frequency, so that both tuned circuits offer inductive reactance to the flow of current. The necessary feed-back from the anode to grid circuit is obtained via the link provided between these circuits by the inter-electrode capacities of the valve. The conditions which need to be observed for oscillations to occur are therefore indicated by equation (174) for the Miller effect (see p. 134).

As was asserted on p. 135, the input admittance to the valve will have a negative resistance component if $m\omega C_{GA} \sin \theta$ is positive. This will be so if θ is positive, implying that the voltage across the anode load impedance leads the anode current, and so make up current variations. This will be the case if the reactance of the inductive arm of the *LC* circuit in the anode is slightly greater than the reactance of the capacitative arm. The negative resistance effect thereby brought about in the grid input circuit must be greater than the total positive losses due to the resistance and grid current associated with the circuit if oscillations are to be obtained.

The Dynatron and Transitron Oscillators. Two types of oscillator circuit which are capable of producing particularly pure sinusoidal wave-form outputs, with less frequency variation than those already described, depend on using the negative resistance effects obtainable from screen-grid and pentode valves to make up for the positive resistance losses in an LC circuit, and so sustain oscillations in it. These circuits are much favoured for laboratory measurement work at medium radio frequencies, but do not generate sufficient power to make them valuable for transmitter technique.

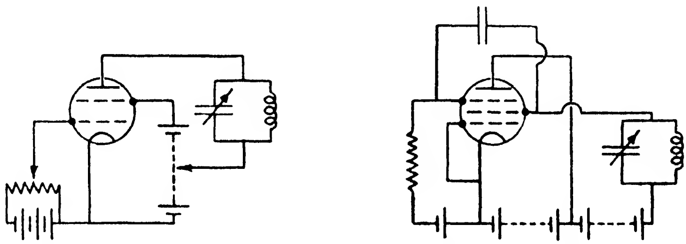


FIG. 74. The Dynatron and Transitron Oscillators.

The dynatron oscillator makes use of the negative resistance effect obtainable over a limited range of anode potentials in a screen-grid valve. The frequency of oscillation of this circuit is not particularly susceptible to variation in H.T. supply voltage. As will be understood from the data on p. 139, the screen-grid valve must be operated with its mean anode potential less than the screen potential for the negative resistance effect due to secondary emission to be realised. Included between the H.T. plus terminal and the valve anode is a parallel rejector circuit. This will have a dynamic resistance at a frequency $f=(1/2\pi)\sqrt{(1/LC-R^2/L^2)}$, of magnitude L/CR (see p. 58). If this is less than the maximum negative resistance of the valve, decided by the slope of the portion AB of the curve, fig. 58b, then oscillations will occur. The amplitude of the oscillations produced will then increase to the point where the dynamic negative anode resistance is equal to the dynamic resistance of the anode load. The circuit will then oscillate at a frequency given by $f=1/[2\pi\sqrt{(LC)}]$, since the effect of R is reduced to zero.

By reducing the value of the negative resistance of the valve by making the grid bias more and more negative until oscillations just cease, the magnitude of L/CR for the tuned circuit can be found, since it is then exactly equal to the negative resistance of the valve at the critical bias point. Hence if L and C are known, the H.F. resistance R of the inductance used can be evaluated at the operating frequency.

The transitron oscillator, which is more stable and preferred in practice, makes use of the negative resistance effect obtainable between the screen grid and suppressor grid of a pentode valve. Variation of the voltage on the suppressor grid of a pentode has a negligible effect on the total number of electrons received by the anode and screen together, deciding (I_A+I_S), where I_A is the anode current, and I_S the screen current, but will affect the division of this space-current between screen and anode. Thus an increase positively of the suppressor voltage causes I_A to increase and I_S to decrease. Vice versa, if the suppressor voltage is reduced, then I_A falls and I_S rises. Therefore the conductance effect obtained between the suppressor-grid potential and the screen grid is negative, i.e. dV_{G3}/dI_S is negative, where V_{G3} is the suppressor potential.

In the transitron circuit, the suppressor and screen are directly connected by a condenser of sufficiently large capacity that its reactance is low at the resonant frequency of the LC circuit used. Hence, when oscillating, the R.F. variations of the screen- and suppressor-grid potentials will be practically in phase. Now when the screen potential rises, the suppressor potential must rise, and under suitable conditions, the screen current decrease due to rise of suppressor potential is greater than the screen current increase due to rise of screen potential. It follows that, when so connected, a rise of screen potential will cause a nett reduction of screen current. There is now effectively a negative resistance between the screen and cathode. This is balanced against the positive dynamic resistance of the LC circuit which is placed in series with the H.T. supply to the screen. The losses in this tuned circuit are thereby reduced to zero, so that continuous oscillations are performed.

Crystal-controlled Oscillators. To avoid the frequency variation associated with valve oscillators due to thermal, mechanical and electrical changes, use is made of the piezo-electric effect exhibited by certain crystals, and in particular quartz, tourmaline and Rochelle salt. This is of special importance in making accurate frequency measurements in the laboratory, and is, moreover, required by law in the maintenance of stable frequency in any radio-telephony transmitting apparatus.

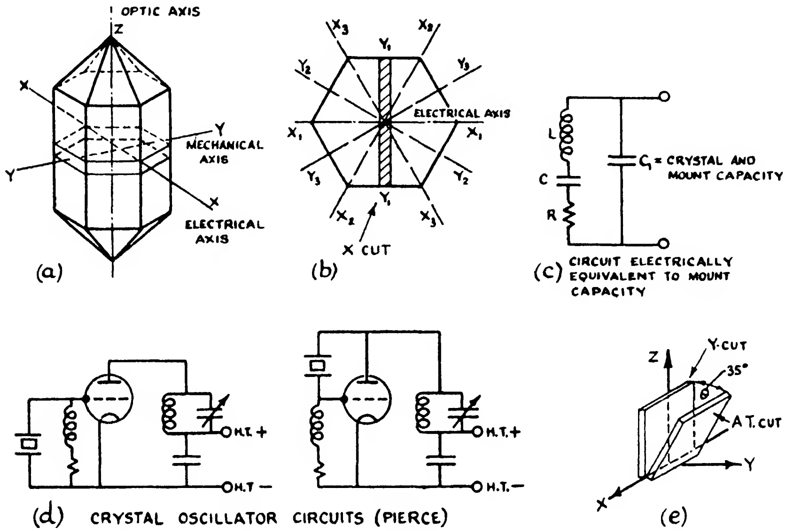


FIG. 75. Quartz Crystal Cuts and Oscillators.

Quartz crystals occur in nature in the form of hexagonal prisms terminated at either end by pointed pyramids. Twinned crystals are common, so that the selection of a crystal suitable for piezo-electric work demands considerable technique, usually assisted nowadays by X-ray crystallographic analysis methods.

The axis joining the pointed ends of the crystal is known as the Z- or optical axis (see fig. 75*a*).

Consider a cross-section of hexagonal shape cut from the crystal, perpendicularly to the Z-axis (fig. 75*b*). The three axes joining opposite points of this hexagonal are the X-axes, X_1, X_2 and X_3. The Y-axes, Y_1, Y_2 and Y_3, are drawn perpendicular to the crystal sides, and meet these sides at their centres.

The X-axes are called electrical axes; the Y-axes are the mechanical axes.

An X-cut (or Curie-cut) crystal is a flat section cut from the crystal with its flat sides *perpendicular* to an X-axis. Likewise, a Y-cut crystal is perpendicular to a Y-axis.

If a mechanical stress is set up across the mechanical Y-axis of an X-cut crystal then a P.D. occurs between the flat sides, along the X- or electric-axis. The polarity of this P.D. is reversed if the mechanical stress direction is changed from compression to tension.

Vice versa, if a P.D. is set up along the X-axis, then mechanical strain is produced in the Y-direction, and the direction of this strain (compression or tension) is decided by the polarity of the applied P.D.

These phenomena are known as the piezo-electric effects. A dynamic effect is also producible. If the P.D. across the flat sides, and along the electrical X-axis is alternating, then alternate compression and tension in the Y-direction is brought about. When the frequency of this applied alternating P.D. is close to a frequency at which mechanical resonance can exist in the crystal, depending on the crystal dimensions, then the amplitude of vibration will exhibit a pronounced maximum. The crystal under such conditions is then electrically equivalent to a series LCR circuit, with a small condenser across this circuit due to the effect of the crystal and holder capacity (fig. 75*c*).*

In this circuit the inductance L is electrically equivalent to the crystal mass, and is usually of the order of 100 H. C, the series capacity, is equivalent to the mechanical compliance of the crystal, and is very small, being usually less than 0·1 μμF. The resistance R is the equivalent of friction in the crystal, and has values ranging from 1000 to 10,000 Ω. Hence, as a resonant circuit, a quartz crystal exhibits very much higher values of LC and $\omega L/R$ than are obtainable with the conventional acceptor circuit.

In practice the crystal is mounted between electrodes and connected in a valve oscillator as shown in fig. 75*d* (Pierce).† Such circuits are similar to that of the T.A.T.G. oscillator, the feed-back from an inductive LC circuit in the anode being capable, by Miller effect, of supplying energy to overcome the losses in the crystal. The coupling between the valve and the LCR circuit equivalent to the crystal is via C_1, the crystal and mounting capacity, and so depends on the ratio C/C_1, which is small. As a result of this small coupling, the high "Q" of the crystal, and the

* Erratum: The caption to Fig. 75c should read "Circuit equivalent to mounted quartz crystal".

† G. W. Pierce, *Proc. Amer. Acad.*, **59**, 81, 1923.

permanent and small nature of the quartz plate, easily subjected to control of temperature by a thermostat, the frequency stability of the circuit can be made very high. A frequency constant to within one part in 10^6 is obtainable using a thermostated *AT* cut crystal (see below).

For an *X*-cut plate

$$\text{Frequency} = \frac{2{\cdot}86}{t}\ \text{Mc./s.} \quad \text{or} \quad \frac{2{\cdot}86}{w}\ \text{Mc./s.},$$

where t and w are the thickness and width respectively of the crystal in mm. The formula involving t is for thickness vibrations of the crystal, whereas that involving w is for vibrations of the crystal across its width. Both oscillatory conditions can be attained at will by suitable adjustment of the *LC* circuit in the valve anode. The *X*-cut crystal, for either direction of vibration, has a frequency variation with temperature which is negative, being -20 per 10^6 per °C.

For a *Y*-cut plate

$$\text{Frequency} = \frac{2{\cdot}86}{t}\ \text{Mc./s.} \quad \text{or} \quad \frac{1{\cdot}96}{w}\ \text{Mc./s.}$$

and for thickness vibrations the effect of temperature change is the same as for *X*-cut crystals, whereas for width vibrations, the frequency change with temperature can be negative or positive, depending on the operating temperature and the ratio of width to thickness of the section, being between -20 and 100 per 10^6 per °C.

Crystals can also perform flexural and torsional vibrations, use being made of the latter in the case of piezo-electric gramophone pick-ups, using Rochelle salt. Unfortunately, coupling within the crystal itself can occur between its various modes of vibration, especially in the case of *Y*-cut crystals, so that, when oscillating, the crystal may change from a thickness vibration to any one of the other width, flexural or torsional modes, resulting in instability of the oscillator. Such changes can be avoided by suitable crystal dimensioning, but in modern practice the choice of plates cut out of the crystal in a plane rotated by an angle of 31° about the *X*-axis results in a cut in which coupling between the various modes of vibration is practically impossible, whilst at an angle of 35°, such coupling is small, and moreover the change of frequency with temperature over a considerable range is reduced to zero. See fig. 75*e* for such an *AT* cut.

The power which is produced by a quartz crystal oscillator is usually kept below 10 W. in practice. If excessive power is developed at high frequency, the crystal will overheat, whereas at lower frequencies the large amplitudes of vibration to which it is subjected may cause it to crack.

Quartz is the commonest material employed for frequency-stabilised oscillators. It can be cut in slabs sufficiently thin to allow it to resonate at frequencies up to 7 Mc./s. Beyond this frequency, and up to 20 Mc./s., tourmaline, which has greater mechanical robustness, but smaller piezo-electric effect, is usually employed. Rochelle salt exhibits a very great piezo-electric effect, E.M.F.s of as much as 100 V. being producible by suitable mechanical stresses. However, its fragility and hygroscopic nature render it suitable only for gramophone pick-up and crystal microphone constructions, where it is usually mounted in a hermetically sealed housing.

Resistance-Capacity Oscillator. To obtain oscillatory conditions in a valve circuit it is necessary to provide power amplification and feed-back with a 180° phase-change between anode and grid, so that any decrease of anode potential is accompanied by an increase of grid potential, and vice versa. In the phase-shift oscillator this is done by using *RC* circuits instead of the usual *LC* circuit. A circuit of pure wave-form and very stable frequency, especially for audio-frequency work, is obtained.

Suppose an alternating E.M.F. represented by $E \sin \omega t$ is applied across a condenser and resistance in series (fig. 76*b*). The current through the condenser leads on the voltage across it by 90°, but the voltage and current are in phase at the resistance, so the voltage IR across the resistance leads on the voltage $I/\omega C$ across the condenser by 90°. From the vector diagram (fig. 71*c*), it is seen that the voltage across the resistance therefore leads on the applied voltage by an angle θ, where $\tan \theta = \dfrac{1/\omega C}{\sqrt{(R^2+1/\omega^2 C^2)}}$. This angle cannot be as much as 90° without $1/\omega C$ being infinitely large, and C therefore very small. However, suppose $\theta = 60°$, then

$$\tan \theta = \frac{\sqrt{3}}{2} = \frac{1/\omega C}{\sqrt{(R^2+1/\omega^2 C^2)}}.$$

$$\therefore \quad \frac{1}{\omega^2C^2R^2+1}=\frac{3}{4}.$$

$$\therefore \quad 3\omega^2C^2R^2+3=4.$$

$$\therefore \quad \omega^2C^2=\frac{1}{3R^2} \quad \text{and} \quad \omega C=\frac{1}{\sqrt{(3)}R}. \qquad (210)$$

If the values of C and R are selected to satisfy equation (210) then the alternating voltage across R will lead the applied voltage

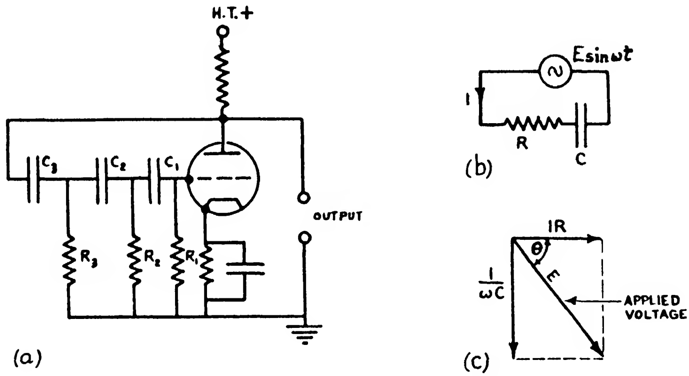


FIG. 76. The Phase-shift Oscillator.

by 60°. It is now possible to place a second CR combination across the resistance of the first combination and obtain a total phase-shift relative to the initial supply of 120°, whilst a third CR combination gives a total phase-shift of 180°, as in C_1R_1, C_2R_2 and C_3R_3 of fig. 76*a*. In the phase-shift oscillator it is therefore possible to obtain the necessary anti-phase relationship between anode and grid voltages by placing the three CR combinations in this manner, where the original alternating voltage input to the first CR circuit is derived from the anode to earth supply of an amplifier valve operating with a resistive anode load. Such a circuit will begin to oscillate as a result of any current change due to H.T. supply variation, or "shot" effect in the valve. It will only oscillate at the specific frequency decided by equation (210), depending on the values of C and R used. Any such initial random current change will be fed back to the grid with a 180° phase-shift: a drop of anode potential being accompanied by an increase of grid potential causing a still further drop of anode potential, the effect being cumulative, resulting in oscillations.

The Kipp Relay. Also known as a "flip-flop" circuit, or Eccles-Jordan relay, this device is really a rapid-acting electronic switch, rather than an oscillator. However, its study is a useful preliminary to a consideration of the valuable multivibrator. In the circuit two valves of the same type are used. They have equal anode load resistances, and a common H.T. supply. The anode of the one valve is connected to the grid of the other, the grid bias batteries shown being necessary to avoid excessively positive grid potentials.*

On switching on the circuit, one valve will always begin to conduct an instant before, or slightly more heavily than the other. Assume valve V_1 initially predominates over V_2 (see fig. 77*a*). The current through R_1, the anode load of V_1, will then be greater than that through R_2. Therefore the P.D. across R_1 exceeds the P.D. across R_2 if $R_1 = R_2$. The anode voltage of V_1 will then be less than that of V_2, and consequently the potential of the directly coupled grid of V_2 will decrease. The anode current of V_2 will therefore fall, so will the P.D. across R_2, whilst V_2's anode potential will rise. But this anode is connected to V_1's grid, so this grid potential will increase, and the anode current of V_1 grow still further. Consideration of this rapid sequence of events indicates that the effect is cumulative. The grid potential of V_2 will be driven to cut off, and anode current through V_2 will stop. This action is practically instantaneous owing to the high speed of the electrons' motions through the valves. The circuit will therefore stay in the condition where V_1 is conducting heavily, and V_2 is cut off. If, now, a positive pulse is applied to both valve grids simultaneously, or to V_2's grid alone, the effect on V_1 will be negligible, since it is already carrying considerable anode current, but at V_2 this positive pulse will relieve this valve of its cut-off bias. It will pass current, then the action detailed above will take place in the reverse direction, resulting in V_2 conducting heavily, and V_1 cut off. Alternatively, this switch-over action can be initiated by a negative pulse applied to the grid of the conducting valve.

This circuit is valuable in that a voltage pulse, without power, applied to one valve grid can cause the switching on of a large

* **Erratum: The polarity of the grid bias batteries shown in Fig. 77*a* should be reversed.**

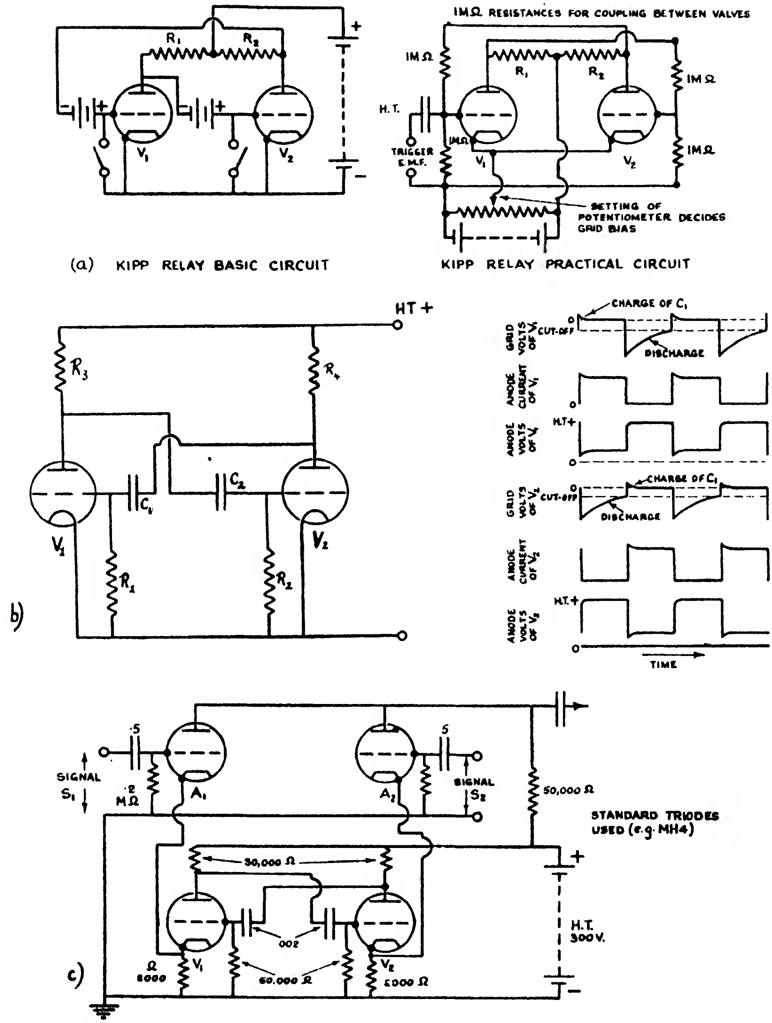


FIG. 77. *a*, The Kipp Relay. *b*, The Multivibrator.
c, Multivibrator as an Electronic Switch.

current through the anode circuit of the untriggered valve. The circuit is therefore directly comparable in purpose with the gas-filled relay (p. 110), but its action is much more rapid, occurring in a fraction of a microsecond.

The Multivibrator. This is an oscillator producing a series of square-topped pulses. The circuit consists of two valves in which, in effect, valve V_1 is resistance-capacity coupled to valve V_2, and V_2 is coupled back by an identical RC arrangement to V_1. Since any signal which is applied to the grid of either valve is amplified by the other valve, reversed in phase at the output and coupled back to the grid of the first valve, so the necessary conditions of feed-back and phase relationship for oscillations to be maintained are achieved.

Suppose that due to "shot" effect, or thermal agitation, a slight increase in the current drawn by valve V_1 occurs. This causes an increase in the P.D. across its anode load resistance, R_3, and a fall of its anode potential. But V_1 anode is capacitatively coupled by C_2 to the grid of V_2. The consequent decrease of V_2's grid voltage causes a reduction of the anode current through V_2. The anode potential of V_2 will therefore rise, and being capacitatively coupled by C_1 to V_1's grid, will cause an increase of this grid voltage, and a further rise of the anode current through V_1. The action is again a cumulative one, resulting in the anode current of V_1 rising to a high value, whilst the current through V_2 is cut off. However, the cut-off grid potential developed on V_2 cannot remain there for long, since it is due to the charge on condenser C_2, which is leaking through the resistance R_2. In a time depending on the time-constant C_2R_2 the condenser potential will fall sufficiently to relieve V_2 of its cut-off grid bias, anode current will therefore begin to flow and the circuit action will be reversed, resulting in V_1 being cut off and V_2 passing maximum current.

As shown in the wave-form diagrams of fig. 77*b*, the currents of V_1 and V_2 will therefore be switched on, alternately, at a constant level, the total period corresponding to one pulse and one gap being T, where $T=K(R_1C_1+R_2C_2)$ sec., where R and C are in ohms and farads respectively, K being a constant depending on the variation of potential which occurs across C_1 and C_2.

Multivibrators have poor frequency stability, and are therefore usually synchronised with another source, usually of sinusoidal or pulse form, applied to the grid circuit of one valve. Such positive pulses, if occurring at a period approximately equal to the multivibrator period, and of sufficient amplitude, predetermine the time at which the cut-off bias on the non-conducting valve is relieved, and so rigidly control the period of operation.

The multivibrator is used as a course of square waves, and as an electronic switch. Since a square wave-form is analysable into a fundamental sinusoidal wave-form plus waves of harmonics up to orders of several hundred times the fundamental, so, with appropriate filter-circuits, a multivibrator can be used as a source of alternating currents all related to the fundamental by integral multiples up to values of a hundred or more. Again, the type of distortion introduced by a valve amplifier when operating with such a square wave-form input can be related to the frequency-response characteristics of the amplifier. Thus a square wave-form with a fundamental frequency of 50 c./s. will contain harmonics up to at least 10,000 c./s., so that an amplifier which will magnify such an input without distortion must have a flat response over at least this frequency range.

As an example of the use of a multivibrator as an electronic switch, consider it as supplying gate voltages to a pair of amplifiers so as to enable two wave-forms to be recorded simultaneously on a single-beam cathode ray tube.

Suppose two signals S_1 and S_2, say at 50 c./s. and 400 c./s. respectively, are to be recorded together on the screen of the tube. Signal S_1 is applied to amplifier A_1, and S_2 to A_2 (fig. 77*c*). The cathode potentials of these two amplifiers depend on the voltages developed across resistances in the cathodes of the two valves V_1 and V_2 of the multivibrator. Thus the P.D. across V_1's cathode resistor decides the positive bias on the cathode of amplifier valve A_1, and likewise for V_1 and A_2. In accordance with the operation of the multivibrator, the P.D. on the cathode of V_1 will be switched on whilst that on V_2 is off, and vice versa. If the potentials developed across the multivibrator cathode resistors exceed the cut-off bias values of the two amplifier valves A_1 and A_2, then these amplifiers will be permitted to work alternately, the period of alternation depending on the selection of the CR values used in the multivibrator. It follows that when A_1 is working, the amplified signal S_1 will be recorded on the cathode ray tube, whilst A_2 is shut off. Immediately afterwards, A_2 will operate, S_2 will be recorded, S_1 will be off. If this gate-voltage occurs sufficiently rapidly, say 2000 times per second, then the two traces will be successively present at the cathode-ray tube, but with such rapid alternation that they will both appear to be recorded continuously, and can be readily compared.

CHAPTER 8

Modulation, Detection and Heterodyne Techniques

The Modulation Principle. The vast field of radio-communication can be said to come within the province of applied electronics, but it is considered here that the discussion of radio technique is outside the scope of electronic principles. Nevertheless, the fundamentals of modulation and detection are of such importance to communication methods, and find such extensive application to electronic devices outside the field of radio-telephony, that a survey of ideas and methods is inevitable.

In conveying energy from a transmitter to a receiver by setting up in the transmitter aerial an alternating current developed by a valve oscillator, it is evident from a study of electromagnetic radiation theory that the field strength set up at any distance from the transmitter greater than five times the aerial length is too small to be detected unless the alternating current frequency is high. Moreover, the length of an efficient aerial for transmitting purposes needs to be about a quarter of the wave-length. Since wave-length × frequency = velocity of radiation (3×10^8 m./sec.), simple calculation readily evinces that transmission demands excessively large aerial currents through a very long aerial unless the frequency employed is in excess of 100 kc./s. Most of the radio-communication with which we are familiar in domestic broadcasting operates at frequencies of the order of 0·3 to 50 Mc./s. (from 1000 to 6 m.).

In radio-telephony the purpose is to convey audible sounds of frequency range from 30 to 20,000 c./s. High fidelity reception achieves a range of 50 to 10,000 c./s., omitting the extremely low and high audio frequencies. If the intelligence to be conveyed is of such comparatively low frequency, whereas frequencies of the order of Mc./s. are necessary to convey the energy efficiently, then manifestly the audio-frequency currents produced by the microphone and A.F. amplifier must, in some manner, be impressed on a *carrier* wave produced by a valve oscillator operating at radio frequency. The technique whereby this effect is achieved is known as *modulation*. There are three main modulation methods:

(*a*) Amplitude modulation; the peaks of the R.F. carrier are made to trace out the A.F. wave-form.
(*b*) Frequency modulation; the R.F. carrier frequency is itself altered to a small extent, this variation being at audio frequency.
(*c*) Phase modulation; the phase of the R.F. carrier is varied at audio frequency.

The Principle of Detection. The tuned circuits of the receiver will, at a particular setting of the tuning condensers, accept a particular radio-frequency carrier for one transmitter selected from all others operating simultaneously. High selectivity is necessary nowadays to do this effectively in view of the large number of transmitters operating in Europe. Provided the bandwidth of these tuned circuits is adequate, there will appear across them a voltage at radio frequency, but varying in amplitude, frequency or phase in a manner depending on the audio modulation. The purpose of the receiver is to supply the amplified audio-frequency current only to the loudspeaker, whilst the R.F. currents need to be suppressed after having performed the useful operation of enabling the energy to be conveyed across space from transmitter to receiver. In order to extract the audio frequency component currents from the modulated R.F. signals, some form of *detector*, or *demodulator* circuit is necessary. This needs to respond to the A.F. modulation voltage impressed across it, and produce a corresponding output, whilst not responding to, but eliminating the R.F. voltage effects.

Amplitude Modulation. In this widely employed method, the peaks or crests of the R.F. current in the transmitter aerial are arranged, by special circuit methods, to vary at audio frequency.

In fig. 78*a* suppose the carrier frequency is 1 Mc./s., whereas the audio frequency is 1 kc./s. There will then be 1000 R.F. peaks occurring during one audio-frequency cycle, and these R.F. peaks will be of varying amplitude, in effect marking out 1000 points on the A.F. current curve. The modulation may be of more or less depth, depending on the relative amplitudes of the modulating current and the carrier current. If I_c is the carrier amplitude, and I_a the effective audio amplitude, then

$$I_a = I_{\text{max.}} - I_c, \qquad (211)$$

where $I_{max.}$ is the maximum amplitude occurring in the modulated carrier when I_a and I_c peaks occur simultaneously. The positive envelope will be seen to vary between I_c+I_a and I_c-I_a during each audio half-cycle, and correspondingly for the negative envelope.

The modulation depth$=I_a/I_c$, and % modulation$=I_a/I_c\times 100$.

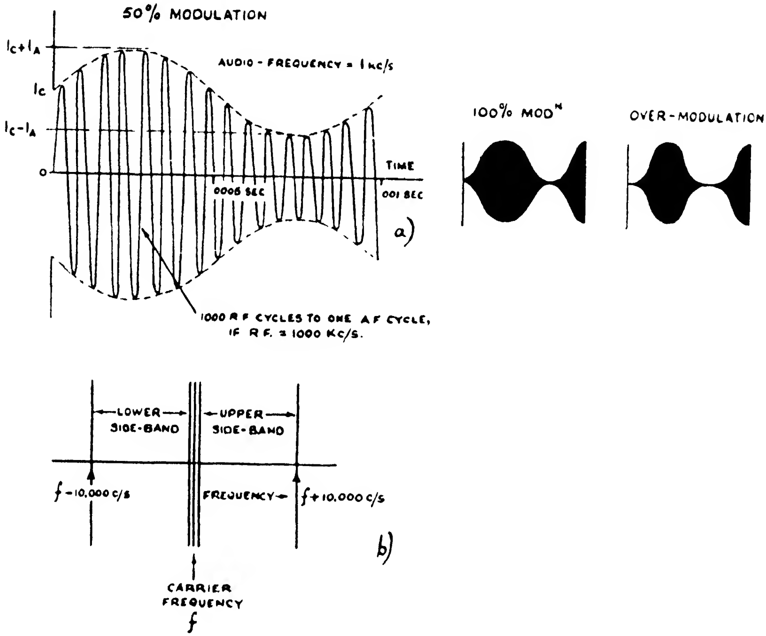


Fig. 78. *a*, An Amplitude-modulated Carrier Wave. *b*, Side-bands.

In fig. 78, percentage modulations of 50, 100 and over-modulation, giving distortion, are depicted.

A mathematical expression representing such a modulated carrier, where a single audio-frequency current is present, will be

$$i=(I_c+I_a \sin at) \sin \omega t, \quad . \quad . \quad . \quad (212)$$

in which i is the instantaneous transmitter aerial current occurring at time t, a is the audio pulsatance and ω the radio pulsatance.

From (212) $\quad i=I_c \sin \omega t+I_a \sin at \sin \omega t. \quad . \quad . \quad (213)$

Expanding this to put it in terms of currents of constant amplitude gives,

$$i = I_c \sin \omega t + \tfrac{1}{2} I_a \cos (\omega - a)t - \tfrac{1}{2} I_a \cos (\omega + a)t.^* \qquad (214)$$

Interpretation of (214) indicates that the effect of modulation is to produce three component R.F. currents,

(*a*) $I_c \sin \omega t$:	at the original carrier frequency, and with the original carrier amplitude.
(*b*) $\frac{1}{2} I_a \cos (\omega - a)t$:	at a radio frequency equal to carrier frequency minus audio frequency, and with half original effective audio amplitude.
(*c*) $\frac{1}{2} I_a \cos (\omega + a)t$:	at a radio frequency equal to carrier frequency plus audio frequency, and with half original effective audio amplitude.

Component (*b*) is called the lower side frequency; component (*c*) is the upper side frequency.

In practical radio-telephony, the microphone produces audio-frequency currents varying over the range from 50 to 10,000 c./s. After amplification it is these audio-frequency currents which give rise to amplitude modulation. Hence in broadcast practice, the audio pulsatance $a = 2\pi f_a$, where f_a is the audio frequency, and f_a will vary to represent musical sounds from 50 to 10,000 c./s. It follows that there are not simply upper and lower side frequencies to contend with, but upper and lower side-bands stretching on either side of the carrier frequency by 10,000 c./s. (see fig. 78*b*).

This side-band effect is important in considering the nature of the resonant tuned circuits employed in both receiver and transmitter. They must have a flat response over a band of frequencies 20 kc./s. wide, with the carrier as the mean frequency. This involves bandpass tuned circuits (see p. 68) which, if they pass too narrow a frequency band will result in high note loss in the receiver output, whilst if they pass too wide a frequency band will not be sufficiently selective to eliminate signals due to unwanted transmitters with a carrier and side-band frequencies near the carrier frequency of the required transmitter signal.

* Since $\cos (\omega t + at) = \cos \omega t \cos at - \sin \omega t \sin at,$
and $\cos (\omega t - at) = \cos \omega t \cos at + \sin \omega t \sin at,$
$\therefore \cos (\omega - a)t - \cos (\omega + a)t = 2 \sin \omega t \sin at.$

Power in Side bands. The power in the carrier is proportional to carrier-current squared; i.e. $\propto I_c^2$. Likewise the upper side-band and lower side-band powers will be proportional to the squares of their respective amplitudes. From equation (214) it is seen that these are both $I_a/2$, so the total side-band power is $\propto 2I_a^2/4 = I_a^2/2$.

For 100% modulation, $I_a = I_c$.

$$\therefore \quad \text{Power in carrier} = P = kI_c^2,$$

where k is a constant and

$$\text{power in side bands} = S = \frac{kI_a^2}{2} = \frac{kI_c^2}{2} = \frac{P}{2}.$$

The power in the side-bands is thus at the most half the carrier power, and one-third of the total power transmitted.

For 50% modulation, $I_a = \frac{1}{2}I_c$.

$$\therefore \quad \text{Power in carrier} = P = kI_c^2,$$

and power in side bands $= S = \dfrac{kI_a^2}{2} = \dfrac{k(\frac{1}{2}I_c)^2}{2} = \frac{1}{8}kI_c^2 = \dfrac{P}{8}$, and side power is only $12\frac{1}{2}$% of carrier power.

Basic Circuits for Amplitude Modulation. It is rare, except in simple transmitters, that the oscillator is directly modulated. The usual transmitter arrangement consists essentially of a master-oscillator with particularly stable frequency, generally crystal-controlled, followed by a class B or class C power amplifier stage. To avoid varying the oscillator output, and thereby almost inevitably varying its frequency undesirably, the modulation is performed at the class B or C amplifier.

Modulation is performed by injecting the audio-frequency output of the microphone and amplifier at:

(*a*) the anode of the class C power amplifier: anode modulation.
(*b*) the grid of the class B or C power amplifier: grid modulation.
(*c*) by cathode modulation of the power amplifier.
(*d*) by suppressor-grid modulation of a tetrode or pentode power amplifier.

Various other methods have been proposed and used.

Only the essentials of methods (*a*) and (*b*) are considered here.

The commonest technique employed in radio-telephony is to anode-modulate a class C power amplifier. This procedure allows a great percentage of modulation to be attained with little distortion. It demands, however, a large power output from the

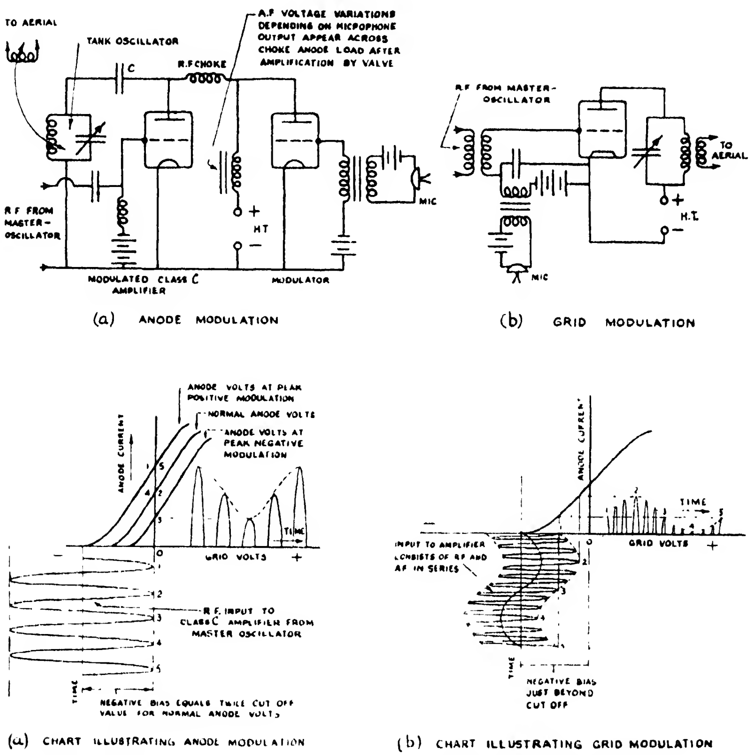


FIG. 79. *a*, Anode Modulation with Chart illustrating Circuit Action. *b*, Grid Modulation with Chart illustrating Circuit Action.

microphone amplifier. In the grid-modulated power amplifier, the output of the microphone (with or without amplifier) is put in series with the R.F. carrier input to the grid of a class B or C amplifier.

In the circuit of fig. 79*a*, the output from the microphone is amplified by the modulator valve, operating class A, so that a considerable voltage variation at A.F. appears across the iron-cored choke as anode load. This load is in series with the H.T.

supply which is common to both the modulator and the class C amplifier valves. At the same time the master-oscillator, radio-frequency input to the modulated amplifier produces, after amplification by the class C operated valve, a series of pulses of R.F. across the R.F. choke which feeds the tuned rejector circuit, or tank oscillator via a low reactance condenser *C*. This oscillatory circuit will only respond to the fundamental, or a harmonic of the periodic pulses of anode current which feed it. If sufficient make-up current is thus supplied, the alternating current at R.F. in the tank oscillator will be sinusoidal. The amplitude of the sine wave will, however, vary depending on the amplified A.F. across the iron-cored choke, because the voltage variations across this choke will cause the anode potential of the class C amplifier to swing at audio frequency about the D.C. value decided by the H.T. supply. Hence the chart of fig. 79*a* illustrating anode modulation is referred to a series of $I_A - V_G$ characteristics, of which three are drawn, one at the positive peak anode voltage when the voltage across the A.F. choke is at its positive peak, two when the anode voltage is at a mean value and the A.F. voltage is instantaneously zero, three when the alternating voltage across the A.F. choke is at its negative peak value, and the anode potential of the class C amplifier is at a minimum.

In the grid modulation method (fig. 79*b*), the input to a class B or C amplifier is due to the series combination of R.F. from the master-oscillator, and A.F. from the microphone. Again the anode load is a tuned circuit which has a resonant frequency equal to the fundamental of the periodically pulsating anode current. In all cases of modulation this output tuned circuit, or tank oscillator, must have a sufficient band-width at resonance to accommodate the side-bands which are produced by the modulation.

Frequency Modulation. The principle in this case is best understood by considering an example. Suppose a carrier wave of frequency 1 Mc./s. is to be frequency modulated by an audio-frequency current of amplitude I_a and frequency $f = a/2\pi$. Then the carrier at 1000 kc./s. is compelled to make a frequency excursion on either side of its mean value. The value of this frequency excursion, i.e. the departure from the mean carrier frequency in c./s., is made proportional to the audio amplitude I_a, whereas the number of times per second, or frequency at which this excursion takes place is decided by the audio frequency,

f. Throughout this process the carrier amplitude remains constant.

For example, suppose a frequency excursion of x kc./s. represents an audio amplitude of I_a at audio frequency f, then the carrier frequency will vary from $(1000+x)$ to $(1000-x)$ kc./s., f times per second. If the audio frequency is then altered to mf, where m is any number, integer or fraction, then this change is represented by the carrier frequency making this same excursion, but mf times per second. Again, if the audio amplitude changes to nI_a, where n is any number, then this alteration is represented by making the frequency excursion from $(1000+nx)$ to $(1000-nx)$ kc./s., at mf times per second. It is readily seen that, by a suitable choice of the excursion frequency, any range of audio-frequency currents with varying amplitudes can be represented by the frequency modulation of the carrier wave.

An expression representing a frequency-modulated carrier wave when the audio frequency has a single value (i.e. sinusoidal) is

$$i=I_c \sin(\omega t+m_f \sin at), \qquad (215)$$

where I_c is the carrier amplitude, which is constant, ω is the pulsatance of the unmodulated carrier and a is the audio pulsatance. m_f is the modulation index, which is defined as

$$m_f=\frac{\text{departure of the radio frequency from the unmodulated carrier frequency}}{\text{modulating frequency}}.$$

m_f depends on the amplitude of the audio-frequency current.

Frequency-modulation technique offers advantages for ultra-short wave communication working over the amplitude modulation method in that the signal/noise ratio is improved; a lower transmitter power can be used for the same receiver audio output; and a greater contrast is obtainable in the receiver audio output because there is less compression of the amplitude of the audio-frequency voltage in the modulation. A disadvantage of frequency modulation is that the side bands are more extensive than in the case of amplitude modulation. With an audio-modulating frequency of $f=a/2\pi$, side-band components at $2f$, $3f$, etc., are present. Fortunately the amplitudes of the higher harmonics of the fundamental side frequency diminish rapidly the farther removed they are from the carrier frequency.

Since the details of frequency modulation can only be studied in the light of modern radio-communication methods, they are considered here to be outside the scope of electronic principles. However, the methods whereby an oscillator may be frequency-modulated involve a valuable technique worth consideration; the use of a valve as a variable reactance.

Variable Reactance Valve Circuits. A valve can be made to act like an inductance (current lagging by 90° on voltage), or like a capacitance (current leading by 90° on voltage), in a circuit, where the effective inductance or capacitance can be made to

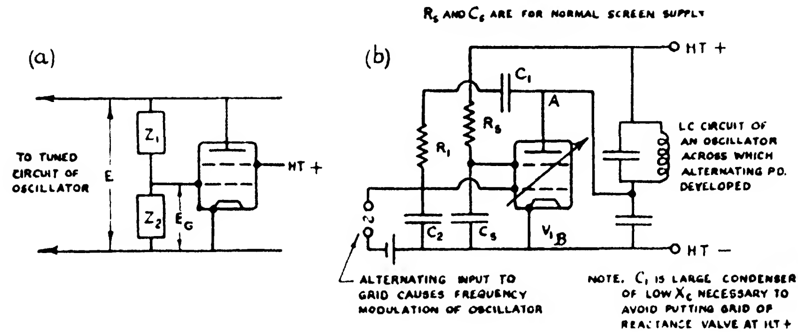


FIG. 80. Variable Reactance Valve Circuits.

depend on the valve grid bias. If such grid bias is then alternated, then a reactance which is made to vary in accordance with an applied signal can be achieved. The use of such a variable reactance across the tuned *LC* circuit of an oscillator enables a frequency-modulated oscillator to be produced.

Suppose a tetrode or pentode valve of high A.C. resistance is connected across the alternating supply of R.M.S. voltage E derived from the LC tuned circuit of an oscillator. An A.C. potential divider consisting of two series impedances Z_1 and Z_2 is connected across the A.C. supply, and the fraction of the alternating potential developed across Z_2 supplies the valve grid with an R.M.S. voltage of E_G. If g_m is the dynamic mutual conductance of the valve, then the R.M.S. anode current is $g_m E_G$.

Then the admittance Y of the valve to the A.C. supply is $(g_m E_G)/E$, where the valve A.C. resistance is so large in comparison as to be neglected.

But $E_G = \dfrac{EZ_2}{Z_1+Z_2}$.

$$\therefore \quad Y = \frac{g_m E Z_2}{E(Z_1+Z_2)} = \frac{g_m Z_2}{Z_1+Z_2} . \qquad . \qquad . \quad (216)$$

If Z_1 is a resistance R_1 and Z_2 is due to capacitance C_2, then if ω is the supply pulsatance,

$$Y = \frac{g_m \cdot 1/j\omega C_2}{R_1 + 1/j\omega C_2} = \frac{g_m}{j\omega C_2 R_1 + 1} . \qquad . \qquad . \quad (217)$$

Rationalising (see p. 56) by multiplying by $(1 - j\omega C_2 R_1)$ gives $Y = \dfrac{g_m(1 - j\omega C_2 R_1)}{1 + \omega^2 C_2{}^2 R_1{}^2}$.

The real part of this expression corresponds to a conductance of $\dfrac{g_m}{1+\omega^2 C_2{}^2 R_1{}^2}$, whereas the imaginary part corresponds to a susceptance of $\dfrac{-jg_m \omega C_2 R_1}{1+\omega^2 C_2{}^2 R_1{}^2}$. The valve is therefore equivalent to a resistance of $\dfrac{1+(\omega C_2 R_1)^2}{g_m}$ in parallel with an inductive reactance of $\dfrac{1+(\omega C_2 R_1)^2}{g_m \omega C_2 R_1}$. Note that a susceptance of $\dfrac{-jg_m \omega C_2 R_1}{1+\omega^2 C_2{}^2 R_1}$, since preceded by $-j$, corresponds to the case where the current lags on the voltage by 90°, i.e. is inductive.

If the valve has a large A.C. resistance, and high amplification factor, as in the case of a H.F. tetrode or pentode, then E_G needs be much smaller than E. Consequently Z_2 should be much less than Z_1. Hence, to a first approximation, Z_2 can be ignored in the denominator of (216).

So $Y = \dfrac{g_m Z_2}{Z_1}$, and (217) becomes

$$Y = \frac{g_m \cdot 1/j\omega C_2}{R_1} = \frac{g_m}{jR_1 \omega C_2} = \frac{-jg_m}{\omega C_2 R_1} . \qquad . \qquad . \quad (218)$$

Therefore the valve acts as an impedance of $\dfrac{j\omega C_2 R_1}{g_m}$, i.e. is equivalent to an inductance.

If, now, the grid bias can be alternated by a modulating signal, and a variable-mu valve is chosen so that g_m is decided by the grid potential, then the effective inductance of the valve depends on $1/g_m$, and is varied in accordance with the modulation. Such a varying inductance used across the LC circuit of an oscillator will then alter the effective inductance in the oscillator circuit, giving frequency modulation.

If instead of using a condenser at C_2, an inductance L_2 is chosen, then $Y=\frac{g_m j\omega L_2}{R_1}$ from (218) and $Z=\frac{-jR_2}{g_m\omega L_2}$. The valve is now equivalent to a capacity in the circuit.

In fig. 80*b* the valve V_1 is used as a variable inductance placed across the LC circuit of an oscillator.

Dealing with the action of this circuit non-mathematically and ignoring the low reactance condenser C_1. Consider an alternating potential across points AB. This is divided between R_1 and C_2, where the reactance of C_2 is much less than R_1. The alternating potential across C_2 and so across the grid cathode lags on the current through it by 90°. But the current through R_1 and C_2 in series is practically in phase with the alternating P.D. across points A and B, since the effect of the condenser C_2 is much less than the resistive effect R_1. Consequently the grid potential variations lag by 90° on the anode voltage variations at the valve. But the alternating current through the valve must be in phase with the grid voltage alternations, so the current through the valve lags by 90° on the anode voltage across it, i.e. the valve has an inductive effect in the circuit.

An entirely different method of using a valve as a variable reactance is to take advantage of the Miller effect (see p. 134). Since the effective input admittance to a valve depends on the gain afforded by the valve in a circuit, and using a variable-mu valve this gain can be made to depend on the grid bias, so it is possible to use a Miller valve circuit in which the effective input capacity to the grid depends on the grid bias. If this capacity is then part of that necessary to decide the frequency of working of a valve oscillator, then frequency modulation can be achieved.

Detection. Also called demodulation, is the opposite to modulation, and involves the extraction of the audio-frequency current component from the modulated R.F. appearing at the receiver. The detection of amplitude-modulated currents is achieved by some form of valve rectifier device combined with capacities of appropriate size to by-pass the unwanted R.F. currents to earth. The detection of frequency-modulated currents demands some form of valve circuit capable of frequency discrimination.

Detection must involve as little distortion of the A.F. current extracted as possible. Methods employed in the case of the much commoner amplitude-modulated signals are: (*a*) diode detection, (*b*) leaky-grid detection, (*c*) anode-bend detection, (*d*) infinite impedance detection.

(*a*) *Diode Detection.* As is considered on p. 80, the diode is a rectifier which only passes current when its anode is positive. Combined with a small condenser and large leak resistance in its anode circuit it forms, in these days of superheterodyne receivers capable of providing a large signal voltage, the commonest method of detection (fig. 81*a*).

The condenser C in the circuit is usually between 0·0001 and 0·0003 μF. so that it acts as a low reactance to R.F. of the order of 1 Mc./s., compared with the resistance R (order of 0·5 to 3 MΩ.). At audio frequency, however, its reactance is large compared with the value of resistance R.

On p. 133 the action of a condenser-resistance combination in the grid circuit of a triode was considered. The action in the case of a diode is the same, except that the condenser becomes charged up with rectified pulses of anode current instead of grid current. Consequently, as before considered, the P.D. across CR, which decides the anode bias, is proportional to the peak alternating potential input. If this alternating input is amplitude modulated, and the time-constant of the CR combination is approximately equal to the period of the highest audio-frequency component, then the P.D. across CR will vary at A.F. in accordance with the manner in which the R.F. input is modulated. This detected A.F. voltage can then be subsequently amplified by a triode or pentode valve, whilst in the coupling circuit used between the valves a filter eliminates the residual R.F. component.

$$\text{The detection efficiency} = \frac{\text{peak A.F. voltage across diode load}}{\text{peak amplitude of carrier} \times \text{depth of modulation}}.$$

This efficiency depends on the amplitude of the applied signal, being an optimum at about 5 V., and on $R/(R+R_a)$, where R is the anode load resistance, and R_A the A.C. resistance of the valve. For this reason R is usually made five to ten times the diode resistance.

(*b*) *Leaky-grid Detection* (fig. 81c) acts similarly to diode

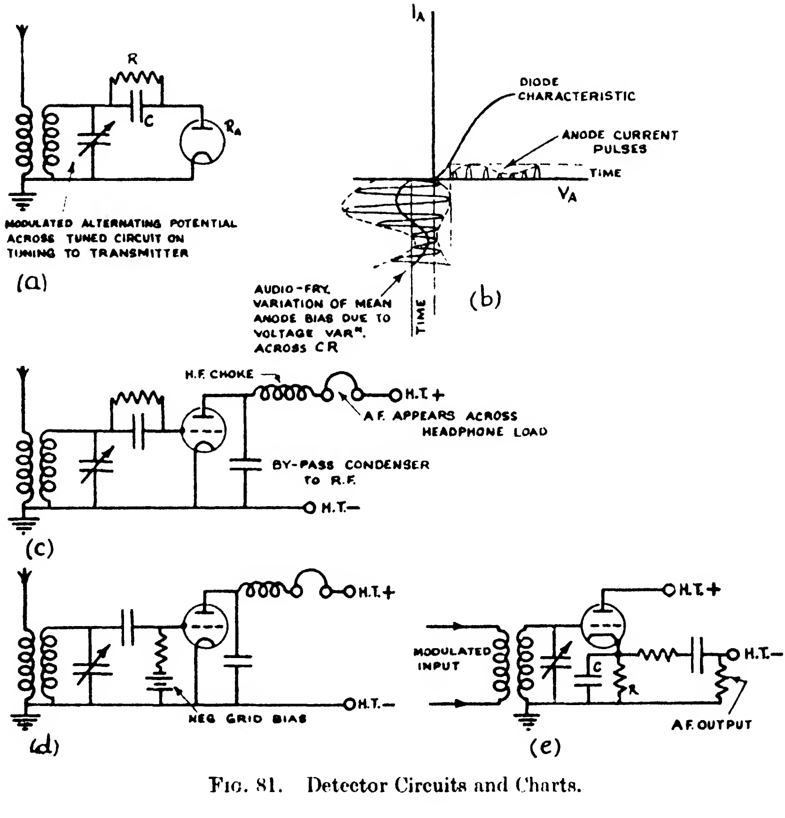


Fig. 81. Detector Circuits and Charts.

detection, except that the CR combination is placed in the grid-circuit of a triode or tetrode valve. By the same action as before, the grid potential will be made to vary at audio frequency depending on the modulation of the R.F. input. Such grid potential changes produce a corresponding change of anode current. Therefore a load resistance used in the triode anode circuit will have the necessary audio-frequency voltage variations across it. In this case distortion is more likely because of the lack of linearity of the grid volts-grid current characteristic of a triode. This is especially likely with inputs in excess of 0·5 V. R.M.S. using a triode of A.C. resistance greater than 20,000 Ω.

(*c*) *Anode-bend Detection*. In this case (fig. 81*d*) a triode or tetrode valve is used with a steady bias almost equal to the cut-off value (cf. Class B amplifier, p. 151). The curved portion of a triode characteristic is a parabola, and can therefore be represented by an equation connecting the anode current I_A with the grid potential V_G of the form,

$$I_A = A + BV_G + CV_G^2. \qquad (219)$$

Suppose an alternating input of constant amplitude represented by $E_0 \sin \omega t$ is applied to the valve in series with the fixed negative bias of value $-E$. Then putting $V_G = (E_0 \sin \omega t - E)$ in equation (219)

$$I_A = A + B(E_0 \sin \omega t - E) + C(E_0 \sin \omega t - E)^2$$
$$= A + BE_0 \sin \omega t - BE + CE_0^2 \sin^2 \omega t - 2CE_0E \sin \omega t + CE^2.$$

Putting $\sin^2 \omega t = \dfrac{1 + \cos 2\omega t}{2}$. (220)

$$\therefore \quad I_A = A - BE + CE^2 + \frac{CE_0^2}{2} + (BE_0 - 2CE_0E) \sin \omega t + \frac{CE_0^2}{2} \cos 2\omega t. \qquad (221)$$

In this expression (221), $A - BE + CE^2$ corresponds to the steady anode current depending on the steady bias E without the alternating input. $CE_0^2/2$ indicates that the steady current increases depending on the square of the amplitude of the input alternating potential.* The term involving $\sin \omega t$ indicates the magnitude of the A.C. component at the frequency of the input alternation, whereas the term $CE_0^2/2 \sin 2\omega t$ indicates that an alternating anode current component at the second harmonic of the input frequency is present.

In the case of the alternating input being amplitude modulated, then E_0 is not constant, but is represented by $(E + E_A \sin at)$ where a is the audio pulsatance and E_A the amplitude of the modulating audio signal. In this case the component $CE_0^2/2$ of equation (221) becomes

$$\frac{C}{2}\left(E + E_A \sin at\right)^2 = \frac{CE}{2} + CEE_A \sin at + \frac{C}{2} E_A^2 \sin^2 at.$$

This indicates that the mean anode current is caused to alternate

* This fact is used in the valve voltmeter.

at the fundamental audio pulsatance on applying a modulated signal, i.e. detection is achieved. Since there is also present a component $(CE_A{}^2/2)\sin^2 at=(CE_A{}^2/4)(1+\cos 2at)$, it follows that second harmonic distortion of the audio signal is produced. By suitable operation of the valve, with an input not exceeding 2 V., such second harmonic distortion can be kept to less than 10%. However, this method of detection gives greater distortion than either of the other two methods.

Note that on putting $E_0=E+E_A \sin at$ in the other terms involving $\sin \omega t$ and $\sin 2\omega t$ in equation (221), the result is to produce radio-frequency components of the anode current at pulsatances ω, $\omega+a$, $\omega-a$, 2ω, $2\omega+a$ and $2\omega-a$. Such radio-frequency currents are readily separated from the required audio-frequency component by a suitable condenser-choke filter (see fig. 81*d*).

(*d*) *The Infinite Impedance Detector.* The employment of cathode-follower technique (p. 161) in a detector stage (fig. 81*e*) leads to a type of circuit with very high input impedance and freedom from distortion. Such a circuit offers advantages over the diode and leaky-grid methods as there is no current demand on the input circuit during the time the input is positive, i.e. damping is negligible. Rectified anode current flows in the cathode load R on the application of a signal voltage, the result being a separate A.F. component across RC, as in diode detection.

Heterodyne Methods. In heterodyne techniques the object is to mix two alternating signals or currents so that a component at the difference frequency, or a "beat" frequency is produced. Thus two signals $V_1 \sin \omega_1 t$ and $V_2 \sin \omega_2 t$ are required to be so mixed that a signal of the form $V_3 \sin(\omega_1-\omega_2)t$ is obtained, where V_3 depends on V_1, V_2 and the nature of the heterodyne method.

This can be achieved if $\omega_1-\omega_2$ is small compared with ω_1 and ω_2. Thus put $\omega_1-\omega_2=\Delta\omega$, so that $\omega_1=\omega_2+\Delta\omega$, assuming that ω_1 is greater than ω_2.

Consider that two voltages occur simultaneously across the same points in a circuit, then

$$\begin{aligned} V_2 \sin \omega_2 t+V_1 \sin \omega_1 t &= V_2 \sin \omega_2 t+V_1 \sin(\omega_2+\Delta\omega)t \\ &= V_2 \sin \omega_2 t+V_1 \sin \omega_2 t \cos \Delta\omega t \\ &\qquad +V_1 \cos \omega_2 t \sin \Delta\omega t \\ &=(V_2+V_1 \cos \Delta\omega t)\sin \omega_2 t \\ &\qquad +V_1 \cos \omega_2 t \sin \Delta\omega t. \end{aligned} \tag{222}$$

This corresponds to two sinusoidal components of varying amplitudes, one depending on sin $\omega_2 t$ and the other on cos $\omega_2 t$. The two components are consequently 90° out of phase with respect to one another, so that their resultant is only obtainable by vector addition. Applying Pythagoras' theorem, the resultant is given by

$$\sqrt{[(V_2+V_1\cos\Delta\omega t)^2+(V_1\sin\Delta\omega t)^2]}\sin(\omega t+\alpha), \qquad (223)$$

where α is a phase angle depending on the amplitudes of the two components, which will vary.

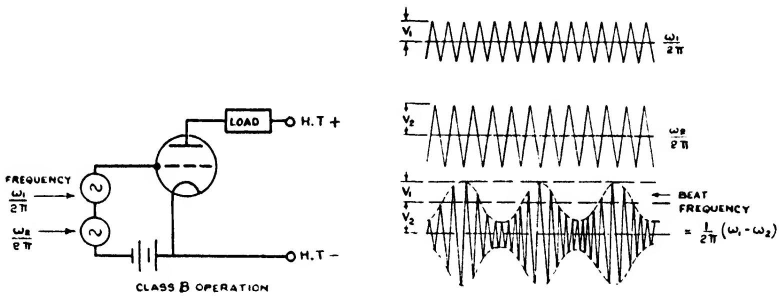


FIG. 82. The Heterodyne Method.

The amplitude of the resultant wave is, from (223),

$$\sqrt{[(V_2+V_1\cos\Delta\omega t)^2+(V_1\sin\Delta\omega t)^2]}$$
$$=\sqrt{[V_2^2+V_1^2(\cos^2\Delta\omega t+\sin^2\Delta\omega t)+2V_1V_2\cos\Delta\omega t]}.$$

But $\cos^2\Delta\omega t+\sin^2\Delta\omega t=1$.

$$\therefore \quad \text{Amplitude}=(V_1^2+V_2^2+2V_1V_2\cos\Delta\omega t)^{1/2}.$$

Expanding by the binomial theorem* gives

$$\text{Amplitude}=(V_1^2+V_2^2)^{1/2}+\tfrac{1}{2}(V_1^2+V_2^2)^{-1/2}\,.\,2V_1V_2\cos\Delta\omega t$$
$$+\frac{\frac{1}{2}\,.\,(-\frac{1}{2})}{1\,.\,2}\,.\,(V_1^2+V_2^2)^{-3/2}\,.\,(2V_1V_2\cos\Delta\omega t)^2+\text{etc.}$$

The resultant signal therefore has an amplitude which varies at the difference or beat frequency $\Delta\omega/2\pi$. Other terms involving higher powers of $\Delta\omega$ are also present, but these are small if either V_1 or V_2 is small. If such a combination signal is applied to any

* $(a+x)^n=a^n+na^{n-1}x+\frac{n(n-1)}{1\,.\,2}a^{n-2}\,.\,x^2+$ etc.

form of detector, it follows that the detector output will contain a beat frequency component at the difference frequency since the input is amplitude modulated at such a frequency.

Electron-mixing, and Frequency-changing. In modern practice, the production of a beat frequency oscillator in which a low-frequency output is obtained by combining two currents at high frequency, and in the superheterodyne receiver where the tuned radio-frequency signals at a receiver are all reduced to the same intermediate frequency value by combining them with a local oscillator signal of suitably varied frequency, the necessary heterodyne technique involves the use of an electron-mixer valve, containing at least two grids.

Suppose alternating potentials represented by $V_1 \sin \omega_1 t$ and $V_2 \sin \omega_2 t$ are applied respectively to grids 1 and 2 of the valve in fig. 83*a*. If the valve is operated with a positive anode potential, but with both grid potentials at zero, or somewhat negative, when no signals are applied, then electron space-charges form around both grids. Since the second grid, farthest from the cathode, is nominally at zero potential, so electrons are decelerated in the immediate locality of the grid wires, forming a space-charge, known as a virtual cathode. When an alternating potential is applied to grid 1 the virtual cathode in the region of grid 2 will contain a number of electrons per unit volume varying at a frequency decided by the potential on grid 1. With alternating potentials applied to both grids, the resultant anode current then depends on the product of the two signal voltages. Therefore

$$I_A = A + BV_1 \sin \omega_1 t V_2 \sin \omega_2 t, \qquad . \quad . \quad (224)$$

where A is constant.

$$\therefore \quad I_A = A + BV_1V_2 \sin \omega_1 t \,.\, \sin \omega_2 t$$
$$= A + BV_1V_2\left[\tfrac{1}{2}\cos\frac{(\omega_1-\omega_2)t}{2} - \tfrac{1}{2}\cos\frac{(\omega_1+\omega_2)t}{2}\right].$$

The anode current therefore contains a component

$$\tfrac{1}{2}BV_1V_2\left(\frac{\omega_1-\omega_2}{2}\right)t,$$

which is at the beat, or difference frequency, and with an amplitude depending on the product of the two signal amplitudes. This method is sometimes known as multiplicative mixing.

To prevent interaction between the two grids, and to achieve a valve of higher effective amplification and A.C. resistance, a hexode mixer valve is usually employed. This possesses two

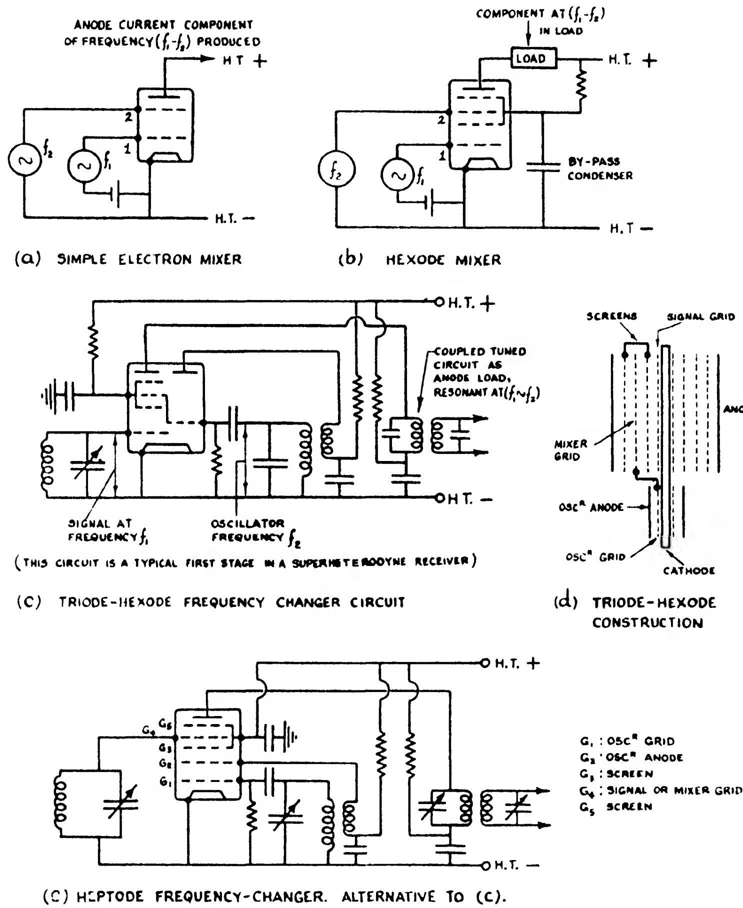


FIG. 83. Electron-mixing and Frequency-changing.

screen grids at a common positive potential, one between grid 1 and grid 2, the other between grid 2 and the anode (fig. 83*b*).

In the superheterodyne receiver, and in other electronic practice where a heterodyne effect is necessary, the mixer valve is usually combined with the local oscillator, the two electrode structures necessary being in the same valve envelope. In the triode-hexode, fig. 83*c*, a triode is mounted side by side with a hexode using separate filaments operated in parallel in the case of a directly heated valve, whereas the indirectly heated type employs a single cathode, the major upper part of which furnishes electrons to the hexode, whilst the lower part is the cathode for the triode oscillator portion. The heptode and octode valves are also used as frequency-changers. The heptode, or pentagrid converter, employs five grids mounted concentrically between a cathode and anode. The first two grids, nearest the cathode, form the grid and anode for an oscillator triode, whilst the fourth grid is the mixer grid around which the virtual cathode appears. Between the triode portion and the mixer grid, and between the latter and the anode, positive screens are introduced. In the octode a sixth grid, a suppressor, occurs just inside the anode; or a beam octode, employing beam plates, is used.

CHAPTER 9

Photo-Electric Tubes and Methods

The Conversion of Light Energy into Electrical Energy. In photometry, in the control of industrial machinery, and in the television and motion picture industries, the applications are manifold of a device in which electricity is produced by the action of light. There are three photo-electric effects, all of which have led to the development of photocells of various types; they are:

(*a*) The photo-emissive, or photo-electric effect proper, in which a cathode material of low work function and low atomic binding forces, produces electrons on being irradiated with light. This effect, already considered on p. 35, leads to the development of the most widely used types of photocell: the vacuum photocell, the gas-filled cell and the multiplier type of photocell.

(*b*) The photo-voltaic effect. Certain types of semi-conducting material on a metal base exhibit a barrier layer effect. Radiant energy falling on such a surface causes an E.M.F. to be generated across this barrier layer, even though no external battery is used. In particular, a layer of cuprous oxide on copper, and iron coated with iron selenide exhibit this phenomenon, leading to the development of the most sensitive type of photocell, which finds extensive application in photometry.

(*c*) The photo-conductive effect. Certain materials, e.g. selenium and the selenides and sulphates, possess electrical conductivities which depend on the intensity of illumination to which they are subjected. Cells based on this principle do not find such widespread application as those based on the other two effects, nevertheless photo-conductivity has been applied to a considerable extent, has played an important part in the history of photocell devices and is an effect which, by research, is still yielding information as to the fundamental nature of semiconductors and electrical conduction theory.

Vacuum Photo-emissive Cells. There are several types of construction used in these cells, three common forms of which are depicted in fig. 84. The photo-emissive material is deposited on a cathode consisting of a solid silver, or silver-plated nickel sheet, or silver deposited on the glass envelope, whilst the anode is an axial metal wire, or frame placed so that, when at a positive potential, it collects the electrons released by the light from the photo-cathode. At the same time the anode and cathode are so arranged as to allow, as far as possible, that the light beam has an unobstructed passage to the cathode. In fig. 84*c*, a type of cell useful in television transmission practice, the silver base layer is so thin as to be semi-transparent, so that the light beam used

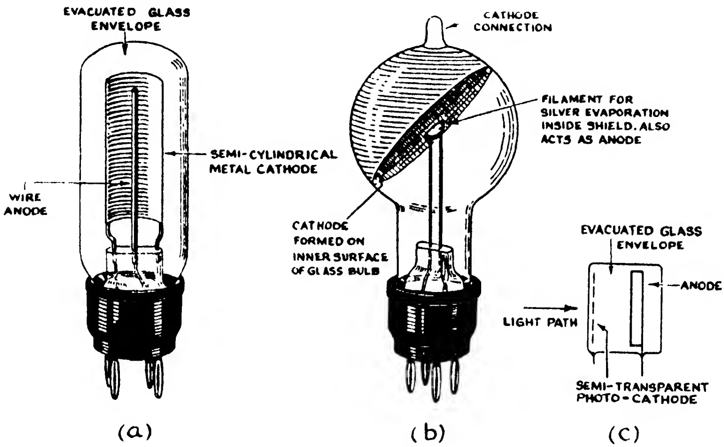


FIG. 84. Forms of Vacuum Photocell.

can conveniently irradiate the back of the photo-cathode, whilst the electrons released emerge from the front. Other base metals than silver have been used, especially antimony, gold and copper, the use of antimony leading to a highly sensitive type of alloy photo-cathode.

The preparation on the surface of the metal base of the photo-cathode can take many forms, several recipes having been developed, patented and used in the history of photocell manufacture. It is not proposed here* to deal with the preparation techniques; it must suffice to say that three types of surface are most widely used at the present time. These are:

(*a*) The silver-oxygen-caesium cathode, in which the silver base is oxidised and then treated with caesium, and sometimes extra

* See A. Sommer, *Photo-electric Cells*, Methuen and Co. Ltd., 1946, and J. Yarwood, *High Vacuum Technique*, Chapman and Hall, Ltd., 1945.

silver. During the preparation a carefully controlled baking treatment is necessary. This type of cathode, hereafter referred to as the Ag-O-Cs type, has been the most widely used in photocell practice. It is, however, giving way to some extent to

(*b*) The antimony-alkali alloy type of photo-cathode. First produced by Görlich, the antimony-caesium cathode, Sb-Cs, exhibits a high quantum yield at the short wave-length end of the visible spectrum, leading to a cell which, to blue and green light, is markedly more sensitive than the Ag-O-Cs type.

(*c*) The most recently developed photo-cathode is of more complex constitution than the other two, consisting of a mixture of bismuth, silver, oxygen and caesium. Though not so sensitive as the Sb-Cs photo-cathode, it has the advantage of being more uniformly sensitive through the spectrum.

A fourth photo-cathode which has been used to some extent when a spectral response curve approximating to that of the human eye is needed, is the silver-oxygen rubidium or Ag-O-Rb type, which is similar to Ag-O-Cs except that the element rubidium is used instead of caesium. Again, special photo-cathodes have been prepared using such materials as potassium in order that the peak sensitivity should be in the ultra-violet.

Spectral Response Curves of Photocells. In the selection of a photocell for use it is important to know how its sensitivity varies with the wave-length of the incident light. This is obtained in practice by the use of a monochromator, an optical apparatus employing a prism, or diffraction grating, for producing a beam of light occupying any desired narrow band of wave-lengths in the spectrum, in conjunction with such an instrument as a thermopile for recording the energy in the light beam. Various such measured light beams of wave-lengths occupying successive narrow regions of the spectrum are allowed to fall on the photocell, and the photo-electric currents recorded, at saturation anode potential, are obtained by a sensitive micro-ammeter. A curve is then plotted of the ratio of photo-current to light energy as recorded by the thermopile versus wave-length. In this way the colour sensitivity of any type of photo-cathode is obtained. Fig. 85 gives such "equal-energy" curves for the three main types of photo-emissive surfaces, (*a*) Ag-O-Cs, (*b*) Sb-Cs, (*c*) Bi-Ag-O-Cs.

These spectral response curves indicate the following practical points:

(*a*) Ag-O-Cs. This photo-cathode has a peak sensitivity in the ultra-violet, and a second peak in the infra-red. The threshold wave-length is slightly greater than 12,000 Å. A photocell using

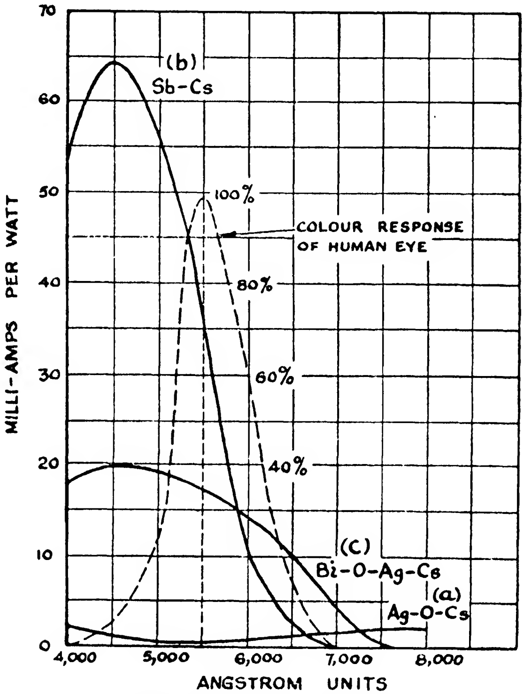


FIG. 85. Spectral Response Curves of important Photo-cathodes.

this cathode can then be used for the general recording of light, provided that comparison with visual photometry is not expected. It could be used in conjunction with an appropriate colour filter to arrange that its response curve through the spectrum approximates more closely to that of the human eye, but such a procedure would render its overall sensitivity much less than that of type (*c*). The pronounced sensitivity of this Ag-O-Cs photo-cathode in the near infra-red region makes it especially useful for recording this invisible radiation, a practical development being the remarkable picture transformer.*

* See V. K. Zworykin and G. A. Morton, *Jour. Opt. Soc. Amer.*, **26**, 181, 1936.

(*b*) Sb-Cs. This cathode has a sensitivity in the blue end of the spectrum which is about a hundred times greater than that of the Ag-O-Cs cathode, but its red sensitivity is very small. It is especially valuable in building light-activated measuring circuits and devices where the short wave-length end of the spectrum predominates, as in mercury discharge lamps and very high temperature incandescent bodies, such as the sun, and the carbon arc.

(*c*) and (*d*). The Ag-O-Rb and Bi-Ag-O-Cs photo-cathodes are particularly useful when the photo-electric device used needs to respond to colour in the same way as the human eye. The more recently developed Bi-Ag-O-Cs surface is much the more sensitive in this respect.

The Overall Sensitivity of Photo-cathodes. The most commonly quoted data regarding photocells is the overall sensitivity in micro-amps. per lumen* (μA./L.) to light from a tungsten lamp at 2848° K. This is highly misleading information if other than such an incandescent source is used for the cell illumination, since the radiant energy from a body at 2848° K. is at a maximum at approximately 10,000 Å, in the infra-red, and only some 10% of the total radiant energy is within the visible spectrum region. For this reason the Ag-O-Cs type shows to advantage compared with the Sb-Cs cell to a greater extent than it should do, the discrepancy being especially marked if a light source at lower temperature than the usual tungsten lamp is used.

The sensitivities in μA./L. for common photo-cathodes used in photocells are:

Cathode material	*Sensitivity* (μA./L.) *Source at* 2848° K.
Ag-O-Cs	40 approx.; max. 60
Sb-Cs	60 approx.; max. 100
Bi-Ag-O-Cs	40 approx.; max. 60
Ag-O-Rb	20 approx.; max. 30

The figures given are necessarily approximate, since much depends on individual manufacturer's preparation technique, and on the geometry of the photocell.

Though misleading in many circumstances owing to large differences in the spectral sensitivity curves of these cathodes, yet

* See Appendix, p. 320.

such figures are practically useful, owing to the frequent adoption of tungsten lamp sources at about 2850° K. in the construction of photo-electric devices.

Volts-current Characteristics of Vacuum Photocells. If graphs are plotted of photo-current against anode voltage at

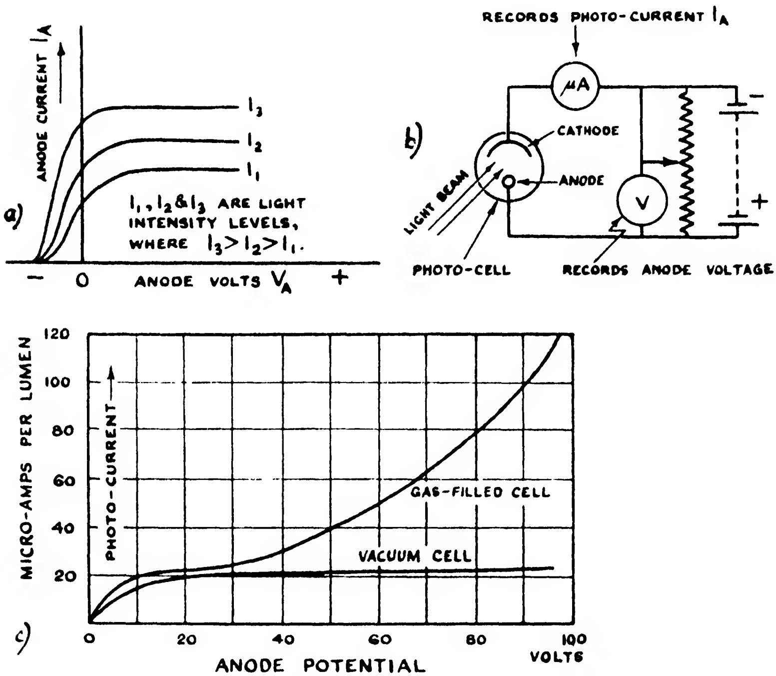


FIG. 86. *a*, The Current Characteristics of Vacuum Photocell. *b*, Circuit for obtaining Volt-current Characteristic. *c*, Characteristic of Gas-filled Cell.

various constant levels of illumination, characteristics of the type shown in fig. 86*a* are obtained, using the circuit of fig. 86*b*.

A small current exists at anode potential zero because of the initial velocities of the photo-electrons, and the effects of contact potential between anode and cathode. As the anode potential is raised so the current increases rapidly at first, non-saturation being due to two causes (*a*) a small space-charge effect, as in the thermionic diode, and (*b*) since the anode field is low at low potentials, so many of the released electrons miss the wire anode, the fraction of photo-emitted electrons not attracted by the anode decreasing as the anode potential is raised. At anode potentials of from 50 to 100 V., depending on the cell geometry, current saturation sets in; the value of this current increases with greater intensity of illumination of the photo-cathode. After saturation, a slight rise of current occurs with further increase of anode potential because of Schottky effect (see p. 87).

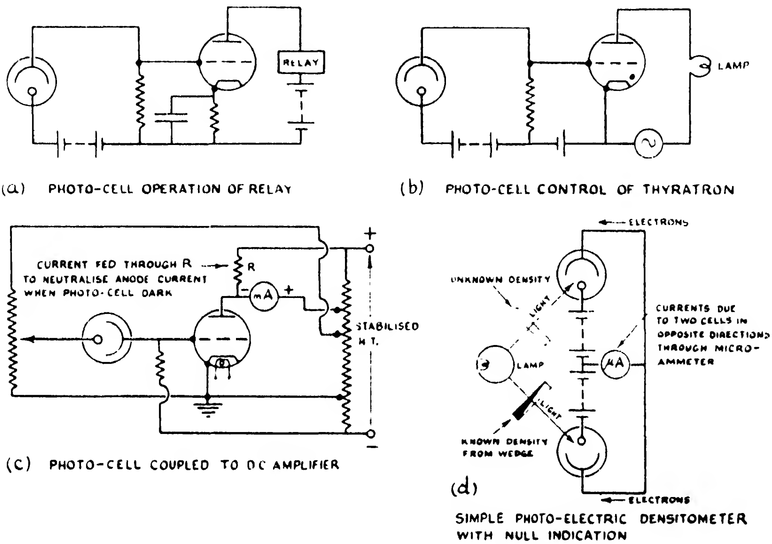


FIG. 87. Some Valve Circuits Employing Vacuum Cells.

The Gas-filled Photocell. The photo-current with given illumination is considerably increased by filling the cell envelope with an inert gas like neon or argon at a low pressure of 0·1 to 1 mm. Hg, the increased current being produced by the ionisation of the gas molecules by the rapidly moving electrons. The volts-current characteristics of these cells do not show saturation like the vacuum types, since increased electron speeds with increased anode potential result in further production of ions from the gas molecules (fig. 86*c*). The current amplification ratio obtained is of the order of five to ten times, depending on the intensity of illumination of the photo-cathode, and on the anode potential. Raising the anode potential too highly is ruinous to the photo-cathodes of such cells, since a glow discharge takes place, and the cathode becomes disintegrated by positive ion bombardment.

Two major disadvantages of the gas-filled cell are:

(*a*) Since positive ions are much heavier than electrons they travel comparatively slowly to the cathode, where they eject the secondary electrons which are largely responsible for the increased current. Hence the cell does not respond well to rapid changes of

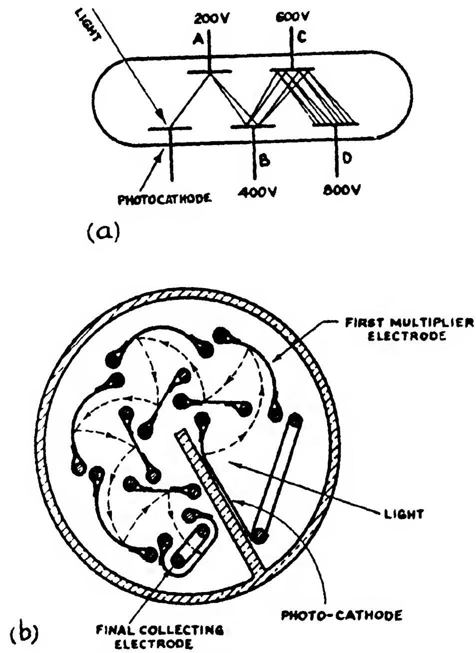


FIG. 88. The Multiplier Photocell.

light frequency, such as are encountered in television practice. A distinct fall-off of the so-called dynamic sensitivity of the cell takes place at frequencies above 3000 c./s. This gas-filled cell is mostly employed in cinema-projector sound track recording heads.

(*b*) The photo-current is not directly proportional to the incident light flux, rendering this type of cell inconvenient for photometric apparatus.

The Multiplier Photocell. To increase the limited output of the vacuum photocell, the electrons liberated by light are directed on to a surface with a high secondary emission ratio. From this target, or from successive such multiplying targets, the electrons are then directed to a collecting anode (fig. 88).

Though affording the advantage of enormously increased output current per lumen, yet the multiplier photocell suffers from disadvantages which make its application less widespread than would be expected. In order to achieve a satisfactory electron multiplication ratio at each target, the P.D. between successive targets is best of the order of 200 to 300 V.* (see p. 39). Hence with five such multiplying electrodes, the total P.D. across the cell needs to be some 1800 V. divided into six 300-V. steps. The provision of such a stabilised supply in conjunction with the complicated type of cell necessary makes the apparatus liable to difficulties due to drift and instability.

The final output current from multiplier photocells cannot exceed about 10 mA., otherwise, at the necessary high operating potentials, the heating effects of the electrons at the multiplier electrodes and collector will be excessive, destroying the low work function preparation on the multiplier targets. Hence the current from the initial photo-electric surface must be small, implying that such cells are best used with low levels of illumination. However, for such low illumination levels, the signal : noise ratio of the multiplier cell is superior to that of a single photocell with valve amplifier giving an equivalent total current gain.

Photo-voltaic Cells. The copper copper-oxide photocell consists of a disk of copper which is oxidised by heat treatment in air at 1100° C. so that its surface becomes black cupric oxide. This cupric oxide is then chemically reduced to reddish-brown cuprous oxide by the action of sodium nitrite. A semi-transparent metallic layer is then deposited by a vacuum process over the cuprous oxide, contact being made to this layer by a copper ring. The other electrode is the copper plate. The iron selenium type of photocell employs a base plate of iron thinly coated with iron-selenide or selenium, which is covered with a semi-transparent silver layer, contact being established by a ring, as before. These cells also act as rectifiers.

The response of these cells to light may be represented by

$$\log E = K \log I, \qquad . \quad . \quad . \quad (225)$$

* The commercially available multiplier photocells, such as the R.C.A. 931-A, employ only 100 V. approx. between stages, with a total H.T. supply of 1250 V., utilising nine secondary multiplier electrodes.

where E=E.M.F. generated across the cell, I is the intensity of illumination of the cuprous oxide, or selenium, and K is a constant.

When light is incident upon such a cell, it traverses the semi-transparent metallic layer and releases electrons from the selenium which travel across the "barrier layer" (fig. 89*a*) between the selenium and the thin metal layer. If an external resistance connects the metal contact ring to the back plate of the cell, then these electrons can traverse a closed circuit back to the selenium again. No external supply E.M.F. is necessary.

Fig. 89*b* shows the output current at various illumination levels using a 45 mm. diameter cell of the iron-selenium type. With a load resistance of 100 Ω., there is an almost linear relationship between output current and illumination level at both low and high intensities. As the load resistance is increased above 100 Ω., however, the current characteristic exhibits definite curvature. This curvature may be of advantage if a logarithmic response is required, as in a photographic exposure meter, or a densitometer. Such cells are capable of producing as much as 650 μA. per footcandle per square foot using a 100 Ω. load resistance.

The spectral response, or relative sensitivity of these cells to lights of various wave-lengths through the spectrum, approximates sufficiently well for many practical purposes to that for the human eye (fig. 89*c*). However, a filter should be used in front of the cell for correction purposes if measurements corresponding more exactly with human vision characteristics are necessary. A 5 mm. thickness of 10% copper sulphate solution serves well as such a filter.

The dynamic sensitivity of these cells is poor. At 50 c./s. light flux variations the current output in μA./L. is considerably less than its value with constant illumination, whereas at still more rapid light flux changes, a more serious decrease of sensitivity occurs. Moreover the self-capacity of the cell is a disadvantage with rapidly varying currents, though this effect can be offset by using a cell of small area, as is done in motion-picture sound head units.

Though the overall sensitivity of this type of cell is high, yet they suffer from the disadvantage compared with the vacuum type that they demand a low value of external load resistance for effective operation. As a result, the output cannot be amplified

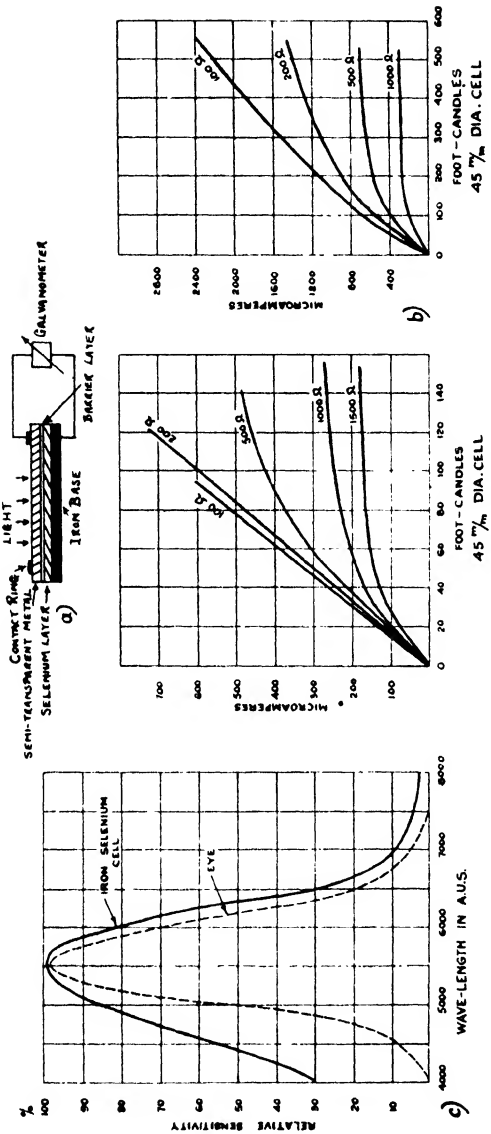


FIG. 89. *a*, A Photo-voltaic Cell. *b*, Current vs. Light Intensity Curves. *c*, Spectral Sensitivity Curve.

by a valve circuit (see p. 118). On the other hand, the convenient flat shape, not mounted in vacuum, allows these cells to be readily mounted in optical measuring instruments where the whole available light flux is readily collected by the cell.

Attempts have been made to build up a series combination of photo-voltaic cell elements so that they can be used with an external load resistance of as much as 1 MΩ., so enabling a D.C. valve amplifier to be coupled to the cell output.*

* Such cells are now marketed bv Evans Electroselenium Ltd.

CHAPTER 10

The Theory and Practice of Electron Optics

The Purpose of Electron Optics. In the text, so far, there has been developed discussions of the effects of electrostatic fields on the electrons emitted from conductors. Most of the analyses made have been straightforward since the electrostatic fields were assumed to be uniform; also the current due to the passage of the electrons has been of more concern than the shape of their trajectories. In fact, the electrons have had to be confined to but roughly defined beams. In many modern electronic devices, like the cathode-ray tube, electron microscope and Klystron oscillator tube, it is imperative, on the other hand, to obtain precise theoretical and practical information about the trajectories assumed by the electrons on traversing non-uniform electrostatic and also magnetic fields. Focusing of electron beams, and the regulated deflection of beams have to be brought about by a suitable disposition of electrodes of a specific geometry. At first acquaintance the problem seems to present insuperable difficulties; an exceedingly useful simplification of the treatment has been introduced, however, by establishing an analogy between the paths of electrons through electrostatic and magnetic fields, and the paths of beams of light through optical media. In fact an expression for the *refractive index* effect of an electrostatic field on a moving electron can be established, and shown to depend on the potential in the field. In several useful cases it has, therefore, become possible to apply optical computing methods to the solution of electron trajectory problems. Electro-dynamical methods have also been applied to the subject without introducing optical computing formulae, though optical ideas are involved. Various practical methods have been introduced, sometimes directly, sometimes indirectly determining the electron path. Such methods have proved invaluable in checking and correlating theoretical determinations.

In practically all cases of applied electron optics, a further simplification of the problem is introduced by considering only the cases of electrostatic and magnetic fields which are symmetrical about the principal geometrical axis of the electrode system; fortunately such fields are those of the greatest value. Again, as in ordinary optics, the use of apertures or stops to narrow down the cross-section of the beam to the paraxial case, implying that the beam only diverges by small angles from the axis of symmetry of the fields used, enables an analysis, which would otherwise involve formidable mathematics, to be tackled by the application of reasonably straightforward methods.

The object of the present discussion is to describe the electronic and optical principles involved to an extent enabling an accurate conception to be obtained of the applications of the subject to vacuum tube design.

The Refractive Index of an Electrostatic Field. Considering a particular case, suppose an electron beam is moving through an equipotential space of which V_1 is the potential with respect to the source of electrons. Let this beam be incident at an angle i on a surface which separates this region of potential V_1 from a neighbouring region of potential V_2. Actually, such a sudden potential increase in a field is unlikely to occur in practice, but this fact does not alter the validity of the general argument, particularly since the difference (V_2-V_1) can be made small to correspond to small displacements in the field. The subdivision of the practical case of a continuously varying field into convenient small sections, each of constant potential, where the value of this potential increases by small, discrete amounts from one region to the next is considered later.

Let v_1 be the velocity of the electron in the region of equipotential V_1, where PO is its linear path, and let v_2 be the electron velocity in the equipotential region V_2, and OQ its path. These velocities can be resolved in two directions, one along the equipotential surface, the other at right angles. Suppose such resolved components are v_{1x}, v_{1y} and v_{2x}, v_{2y}, where x is in the direction of the equipotential surface separating the two media, and y the direction at right angles (see fig. 90). Since there is no electrostatic force operating on the electron in the x-direction, it follows that $v_{1x}=v_{2x}$. But $v_{1x}=v_1 \sin i$ and $v_{2x}=v_2 \sin r$, therefore

$$v_1 \sin i=v_2 \sin r=\text{a constant}. \qquad (226)$$

$$\therefore \quad \frac{\sin i}{\sin r}=\frac{v_2}{v_1};$$

putting $\frac{1}{2}mv_2^2 = eV_2$, and $\frac{1}{2}mv_1^2 = eV_1$ from equation (57), p. 43, it follows that

$$\frac{\sin i}{\sin r} = \sqrt{\frac{V_2}{V_1}}. \qquad (227)$$

This is comparable with Snell's law in the ordinary optical case,

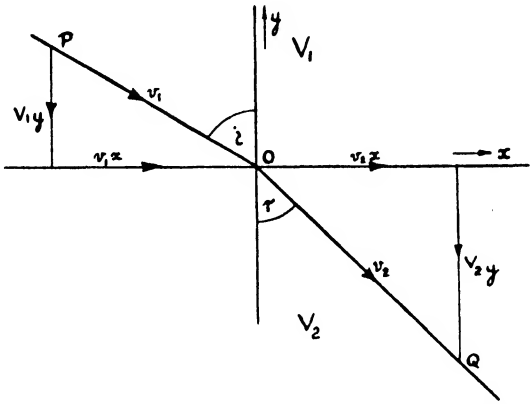


FIG. 90. The Refraction Effect of an Electrostatic Field on a Moving Electron.

defining the refractive index. So the refractive index effect in an electrostatic field can be written as

$$\frac{\mu_2}{\mu_1} = \frac{v_2}{v_1} = \sqrt{\frac{V_2}{V_1}}, \qquad (228)$$

where μ_2 and μ_1 are the refractive indices of the second and first media respectively.

Thus refraction depends on the square root of the potential in the field, and directly on the velocity of the electron.

A more complete mathematical comparison between electron optics and ordinary optics can be drawn by considering the appropriate forms of Lagrange's principle of least action, and Fermat's principle of least time.*

Plotting an Electrostatic Field. Before proceeding further with the discussion of electron optics, it is necessary to know how the map of the lines of force in the cross-section of an electrostatic

* See L. G. Maloff and D. W. Epstein, *Electron-Optics in Television*, McGraw-Hill Book Co., 1938.

field is obtained by experiment. The procedure involves the use of an electrolytic trough (fig. 91*a*).

The figure represents a glass trough filled with a weak electrolyte, a copper sulphate solution consisting of 10 mgm. of solid per litre of water being suitable. Models of the electrodes to be used are placed at the liquid surface. For example, suppose the

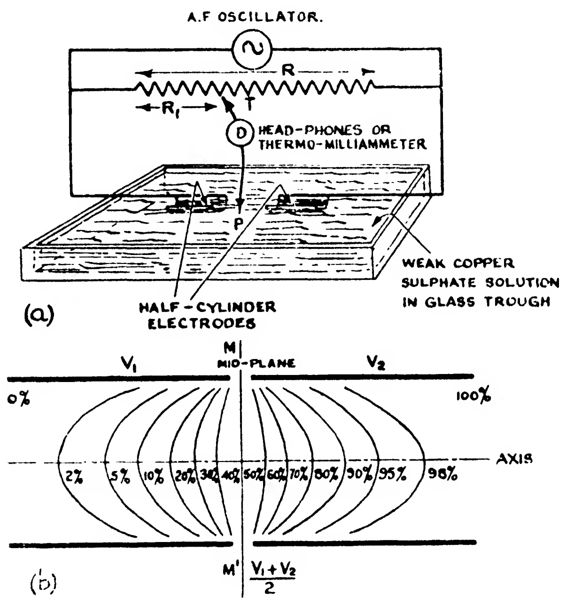


FIG. 91. *a*, The Electrolytic Trough. *b*, The Electrostatic Field Between Two Coaxial Cylindrical Electrodes.

electrostatic field due to two coaxial cylindrical electrodes at different potentials is required, see fig. 91*b*.

The electrolyte surface is arranged to correspond with the cross-section plane of these cylinders which includes the principal axis. This is done by placing half-cylinders of metal in the electrolyte so that their principal axis lies in the liquid surface. The method is based on the grounds that the physical law relating to the direction of the lines of force in an electrostatic field is of exactly the same mathematical form as the expression associated with the propagation of electric currents through an electrolyte.

To avoid polarisation effects, an audio-frequency alternating voltage source in the form of a valve oscillator is used. Across this oscillator is placed the potential divider arrangement formed by the variable resistance R, which has a value of about 10,000 Ω., and calibrated so that its tapping point T intersects a known fraction of the total resistance. The output terminals of the oscillator are joined to the electrodes in the trough as well as to the ends of the resistance, whilst the tapping point T is connected through a recording device to indicate a balance to a probe P which is placed in the surface of the electrolyte, so as to explore the variations of potential between the electrodes. The recording device can be a pair of head-phones, or a sensitive thermo-junction milliammeter. If desired, valve-voltmeters can be connected directly across the two arms of the resistance so as to directly measure the potentials between the probe and either electrode.

Suppose that the potential between the electrodes is V (its absolute value is not of importance), and it is required to move the probe along an equipotential in the electrolyte surface which corresponds to some fraction of V, say 40%. The tapping point T on the potentiometer is set so that the ratio of the resistance of the appropriate branches of the potentiometer to the total resistance is 4 : 10, say $R_1 = 4000$ Ω., and $R = 10{,}000$ Ω. The probe is then moved about in the electrolyte surface along such a line that a balance is maintained, i.e. a zero is recorded by the head-phones or thermo-milliammeter. If this probe is connected by an arm, or a multiplying pantograph, to a pencil, then the contour of this line can be traced out on a sheet of drawing paper so that the 40% equipotential is recorded. Selecting other ratios of $R_1 : R$ enables other equipotentials to be traced so that, in time, a complete plot of the electrostatic field map is obtained. If, for example, the total P.D. between the electrodes is 2000 V., then the various ratios selected correspond to equipotentials of definite values. Thus if one electrode is at 500 V., and the other at 2500 V. with respect to some zero decided by the vacuum tube under consideration, then the 40% equipotential line will correspond to a potential of $500 + 4/10\,(2000) = 500 + 800 = 1300$ V., and likewise for the other percentage equipotentials which have been plotted.

In this connection it should be noted that, in accordance with the analysis given on p. 242, the dimensions and voltages on the electrode system can be multiplied by any constant factor without producing any change in the distribution of the electrostatic field lines. So, in addition to considering the actual potentials in the form of a percentage of some arbitrary potential value (that of the oscillator output), the models of the electrodes used can be magnified replicas (say a 4 : 1, or 10 : 1 linear magnification), and the map of the electrostatic equipotentials magnified in the same proportions, i.e. the shape of the equipotentials, is unchanged. This allowed enlargement of the model system permits a greater accuracy to be obtained. Results correct to within $\pm 2\%$ are possible with care.

Precautions are necessary to avoid spurious contact potential effects, and surface tension troubles. Scrupulous cleanliness of the electrodes and electrolyte must be maintained, dust being a fruitful source of error. Some workers recommend that the glass walls of the trough be coated with a conducting metal layer which is earthed to eliminate wall charges. In any case the electrode system needs to be well in the centre of the trough, away from the trough walls, so as to avoid wall reflection or electric charging effects. The probe is conveniently a short platinum wire of about 0·001 inch in diameter. Silver plating the electrodes to be used to avoid contact potential and metal surface film effects is recommended. Particular care must be observed to ensure that the voltage output of the A.F. oscillator is constant.

Two Cylinder Electron Lens. It is possible to form an electron lens which compares immediately with the simple convex glass lens by using a system of convex and concave grids at various potentials. The use of grids in an electron lens invariably leads, however, to objectionable scattering of the electron beams, so that the simplest practicable form of electrostatic lens is formed by using a pair of coaxial cylinders.

The distribution of equipotentials between two such cylinders can be seen to follow the pattern shown (fig. 91*b*) on realising that the equipotential line at V_1 must cross the principal axis at a point well inside the cylinder at V_1, and likewise for V_2, whereas the equipotentials between V_1 and V_2 will lie in the gap between the two cylinders.

Suppose a parallel beam of electrons is incident on the lens at some height above the principal axis. On entering the electrostatic field between the cylinders the electrons will receive a force component directed along the electric lines of force in the field. At the left-hand cylinder of the lens the equipotentials are convex towards the incident beam, and application of the refraction formula (228) at each successive equipotential shows that the electron beam converges towards the axis. At the right-hand cylinder the electrons will encounter concave equipotentials, so the refraction effect will be divergence from the principal axis. It is necessary for the degree of convergence produced in the first lens element to be greater than the divergence in the second element if the two effects are not to cancel out, and therefore give no nett convergent action.

If V_1 and V_2 are the positive potentials of the first and second cylinders respectively, then the potential of some mid-plane MM' (fig. 91*b*) will be $(V_1+V_2)/2$. Suppose the nett refractive index effect resulting in convergence in the left-hand cylinder is put as μ_C, then the magnitude of μ_C is decided by $\mu_C=\sqrt{[(V_1+V_2)/2V_1]}$, from equation (228). If μ_D is the nett refractive index effect in the second, divergent element, then $\mu_D=\sqrt{[2V_2/(V_1+V_2)]}$.

$$\text{But}\quad \mu_C^2-\mu_D^2=\frac{V_1+V_2}{2V_1}-\frac{2V_2}{V_1+V_2}=\frac{(V_1+V_2)^2-4V_2V_1}{2V_1(V_1+V_2)}$$

$$=\frac{V_1^2-2V_1V_2+V_2^2}{2V_1(V_1+V_2)}=\frac{(V_1-V_2)^2}{2V_1(V_1+V_2)},$$

which is necessarily positive, whatever the values of V_1 and V_2 provided they are positive. Hence μ_C is always greater than μ_D, unless $V_1=V_2$, when $\mu_C=\mu_D$.

It follows that the amount of convergence produced by such a lens is always greater than the amount of divergence. The nett result is convergence. In the case when $V_1=V_2$ then no nett refraction occurs. If V_1 or V_2 is negative, then an electron mirror is produced, the electrons turning back into the direction from whence they arrived on encountering negative retarding potential regions.

As in glass lenses the beam through an electron lens of the electrostatic type is reversible.

In practice only converging* electron lenses are possible. It is therefore not practicable to avoid aberrations as in glass optics

* Except in the rather impracticable case of an electron beam encountering a single aperture.

by the method of using a doublet consisting of a convex and concave pair of elements.

The Cardinal Points of an Electron Lens. When a parallel beam of electrons enters an electron lens, such as that just considered, the nett convergence will bring them to a focus at a point F_2 on the principal axis. This point of focus is most readily determined by using the practical arrangement of fig. 92*a*. It is called the second focal point.

In fig. 92*a*, the electron lens L, consisting of cylinders C_1 and C_2, maintained at the required potentials by supplies connected to the external leads, is irradiated by electrons from an electron gun X. This gun is specially designed to produce a parallel beam of electrons entering the limiting aperture A of the lens. The focused image produced by the lens L is observed on the fluorescent screen T by a travelling microscope R. The screen T can be moved along the axis of the glass envelope tube by movement of the current energised solenoid M, which acts on the iron cylinder inside the glass extension tube. The glass vessel is exhausted to a pressure less than 10^{-5} mm. Hg.

Fig. 92*b* illustrates the case of a parallel beam from an object A_1B_1 falling on an electron lens, to converge towards an image at A_2B_2.

Extrapolating the straight part of the convergent electron beam backwards until it cuts the prolongation of the original parallel beam gives the position of the second principal plane, which intersects the axis at the principal point P_2.

Assuming parallel beams traverse the lens in the opposite direction, then, by a similar procedure, the positions of the first principal focus and the first principal plane and point are found. It is noteworthy that, compared with the usual optical case, the principal planes are crossed over with respect to their corresponding focal points.

In the straightforward manner in which thick glass lenses are treated to deduce the image at A_2B_2 (fig. 92*b*) of height h_2 formed by the lens of an object A_1B_1, of height h_1, draw a beam parallel to the principal axis through the point B_1 of the object; this will be refracted at the principal plane P_2 so as to pass through the focus F_2. A second beam, originating at B_1, and travelling through the focus F_1, is refracted at the principal plane P_1, and leaves the lens in a direction parallel to the axis. The two rays so traced

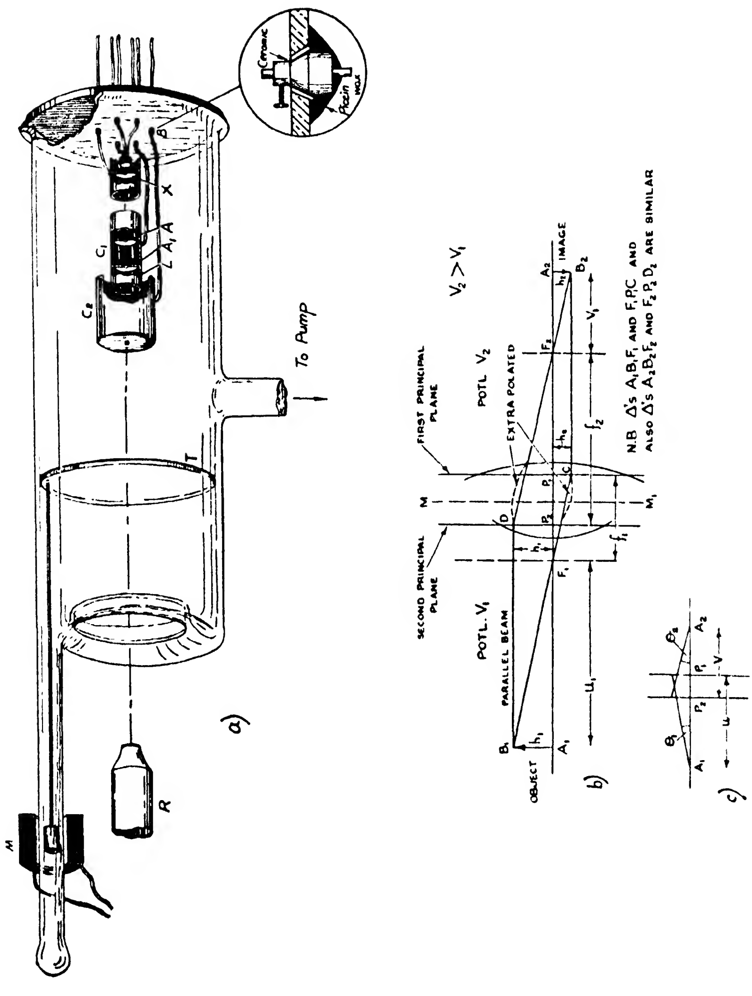


FIG. 92. *a*, Practical Method of observing Lens Focusing. *b*, The Cardinal Points of an Electron Lens.

emerge from the lens to intersect at B_2, giving A_2B_2 as the image.

Suppose $A_1F_1=U_1$ and $A_2F_2=V_1$; $F_1P_1=f_1$ and $F_2P_2=f_2$ are the focal lengths of the system.

Then $A_2B_2/A_1B_1=h_2/h_1=m$, the magnification of the system.

Referring to the geometry of fig. 92*b*, $\triangle A_1B_1F_1$ is similar to $\triangle F_1P_1C$, so $h_2/h_1=f_1/U_1$, also $\triangle A_2B_2F_2$ is similar to $\triangle F_2P_2D$, so $h_2/h_1=V_1/f_2$.

$$\therefore \quad \frac{f_1}{U_1}=\frac{V_1}{f_2}.$$

$$\therefore \quad U_1V_1=f_1f_2. \quad . \quad . \quad . \quad (229)$$

Measuring the object and image distances from the appropriate principal points as U and V respectively, so $U=f_1+U_1$ and $V=f_2+V_1$,

$$(U-f_1)(V-f_2)=f_1f_2.$$

Dividing by UV
$$\frac{f_1}{U}+\frac{f_2}{V}=1. \quad . \quad . \quad . \quad . \quad (230)$$

comparing with the glass optics formula.

Lagrange's Law. In glass optics a familiar law is

$$\mu_1\theta_1h_1=\mu_2\theta_2h_2, \quad . \quad . \quad . \quad (231)$$

where μ_1 is the refractive index in the object space, and μ_2 is the refractive index in the image space. θ_1 is the angle which the incident beam makes with the principal axis of the system, and θ_2 is the angle which the beam emerging from the lens makes with the axis. These angles are small, so that $\theta=\sin\,\theta=\tan\,\theta$. The object and image heights are h_1 and h_2 respectively.

In electron optics, $\frac{\mu_1}{\mu_2}=\sqrt{\frac{V_1}{V_2}}$.

Equation (231) becomes

$$\sqrt{(V_1)}\,.\,\theta_1\,.\,h_1=\sqrt{(V_2)}\,.\,\theta_2\,.\,h_2. \quad . \quad . \quad (232)$$

Hence for unit angular magnification, $\theta_1=\theta_2$, and

$$\frac{h_1}{h_2}=\sqrt{\frac{V_1}{V_2}}. \quad . \quad . \quad . \quad . \quad (233)$$

In the case of paraxial beams, θ_1 and θ_2 are small angles, and from the geometry of fig. 92*c*

$$\frac{\theta_1}{\theta_2}=\frac{A_2P_2}{A_1P_1}=\frac{V}{U}=\frac{f_2+V_1}{f_1+U_1}. \qquad (234)$$

From equation (229), $U_1V_1=f_1f_2$, and substitution in equation (234) gives

$$\frac{\theta_1}{\theta_2}=\frac{f_2+(f_1f_2/U_1)}{f_1+U_1}=\frac{f_2(f_1+U_1)}{(f_1+U_1)U_1}=\frac{f_2}{U_1}. \qquad (235)$$

Put $\frac{\theta_1}{\theta_2}=\frac{\sqrt{(V_2)}h_2}{\sqrt{(V_1)}h_1}$ from (232).

$$\therefore \quad \frac{\sqrt{(V_2)}h_2}{\sqrt{(V_1)}h_1}=\frac{f_2}{U_1}.$$

But $\frac{h_2}{h_1}=\frac{f_1}{U_1}$ from similar triangles $A_1B_1F_1$ and F_1P_1C.

$$\therefore \quad \frac{\sqrt{(V_2)}}{\sqrt{(V_1)}}=\frac{f_2}{U_1}\times\frac{U_1}{f_1}=\frac{f_2}{f_1}. \qquad (236)$$

It follows that the two focal lengths of the electron lens are proportional to the square roots of the voltages at the electrodes.

Trigonometrical Ray Tracing. So far it has been indicated that an electron lens is possible, and that a practical experiment can be devised to find its principal foci, and hence the lens' cardinal points. Such information is very valuable, and is much used in finding empirically the correct electrode system necessary for a particular focused image to be obtained. As in glass optics, however, it is also valuable to make a theoretical approach to the problem, and if possible devise methods of computing the electron trajectories for a particular electrode system, and vice versa, calculate the shape and disposition of the electrodes necessary to achieve a particular trajectory. Though much used, such computing methods are not so accurate in electron optics as in glass optics; they serve more to indicate the trend which electrode design should follow, rather than predict exactly the lens performance. The final criterion is a practical test, usually followed by intelligent modification.

If the path is required of the electrons through an electrostatic field which has been plotted by the electrolytic trough procedure, then the electron-optical form of Snell's law (227) can be applied.

Assuming a number of equipotentials have been plotted, then the spaces between the equipotentials at V_1, V_2, V_3, etc. have refractive indices of $\sqrt{V_1}$, $\sqrt{V_2}$, $\sqrt{V_3}$, etc., which can reasonably be assumed to be constant if the distances between neighbouring equipotentials are so small that the electron paths between them can be assumed to be short, straight lines. Knowing the angle of incidence which the original electron track makes with the first refracting equipotential surface, the angle of refraction can be calculated from (227) and (228). The angle of incidence at the next equipotential is then known, and the second refraction can be similarly calculated. Following out this tedious procedure through all the equipotentials concerned ultimately gives the approximate electron trajectory in the form of a series of contiguous short, straight lines.

The chief source of inaccuracy in this procedure is that it is necessary to draw the normals to the successive equipotentials in order to evaluate the angles of incidence and refraction. This is an inaccurate procedure, especially since the angles of incidence involved are necessarily small to limit the beams to paraxial ones.

To avoid the inaccuracy of this graphical method, Klemperer and Wright* have made use of the standard computing formulae employed in glass optics.

Irrespective of whether one is concerned with light or electrons, fig. 93 represents an incident beam impinging at an incident angle I' on a spherical glass surface, or on an equipotential, of radius r. After refraction, this beam makes an angle I with the normal. Let U be the angle between the incident beam and the axis of the system, whilst the prolongation of the incident beam cuts the axis at a distance L from the spherical surface. The refracted beam angle to the axis is U', and it cuts the axis at a distance L'. The sign convention used is that a distance from the refracting surface to the right is positive, and to the left negative. A positive angle is one generated by the rotation of a radius vector from the positive sense in the anti-clockwise direction, whilst a clockwise direction generates a negative angle. The incident beam is in a space N, whilst the refraction space is N'.

Now the ratio of the sine of an angle in a triangle to the opposite side is constant, so referring to fig. 93, $\sin I'/(L-r)=\sin U/r$, and $(L'-r)/\sin I=r/\sin U'$.

* O. Klemperer and W. D. Wright, *Proc. Phys. Soc.*, **51**, 296, 1939.

Again, since the exterior angle to a triangle equals the sum of the two interior opposite angles, so $U'=I'-I+U$ and, finally, from Snell's law, $\sin I'/\sin I=\sqrt{(V'/V)}$, where V and V' are the potentials of the spaces N and N'.

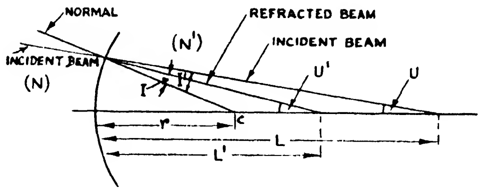


FIG. 93. Nomenclature in Case of Refraction at a Spherical Surface.

So the following computing formulae are available,

$$\left.\begin{array}{ll} \sin I'=\sin U\,(L-r)/r, & (a) \\ \sqrt{(V'/V)}=\sin I'/\sin I. & (b) \\ U'=I'-I+U, & (c) \\ L'=(r\sin I/\sin U')+r. & (d) \end{array}\right\} \quad . \quad . \quad (237)$$

The usual problem is that a set of equipotentials is known from an electrolytic trough plot for a given lens type, and an electron beam is to leave a point on the axis at a known distance L from the first equipotential, at an angle U with the axis, and the electron trajectory through the lens field is required.

By fitting to each of the equipotentials in turn a series of known templates, their radii of curvature are found. Then, to shorten the computation work, two graphs are drawn, as in fig. 95. One is log $\sqrt{V}$ against the distance from the mid-plane of the lens, the other of $1/r$ vs. this distance. Points are then marked on the log $\sqrt{V}$ graph for equal steps of log $\sqrt{V}$, so that the ratio $\sqrt{(V'/V)}$ is constant, implying that equipotentials through the field are selected in such a manner that the refractive index effect between neighbouring spaces, N and N', is constant. The necessary values of r corresponding to these equipotentials is easily found from the graph of $1/r$ vs. distance.

The computation of the path of the electron through the first equipotential of the lens is then performed by inserting the known values of L, r and U in 237*a*, to give I'. Then, since $\sqrt{(V'/V)}$ is a known, prearranged constant, insertion of I' in 237*b* gives I. In 237*c*, therefore, I' I and U are known, hence U' is found, whilst substitution of r, I and U' in 237*d* gives L'. It follows that the refracted ray, emerging from the first equipotential surface, is known, since the angle it makes with the axis, U', and the distance at which it cuts the axis, L', are found. This determined ray then becomes the incident ray with respect to the next equipotential of the lens. So the formulae 237 are used again, where L' and U' from the first computation become L and U for the second. Proceeding in this manner, the path of the electron through as many equipotentials as is required can be determined.

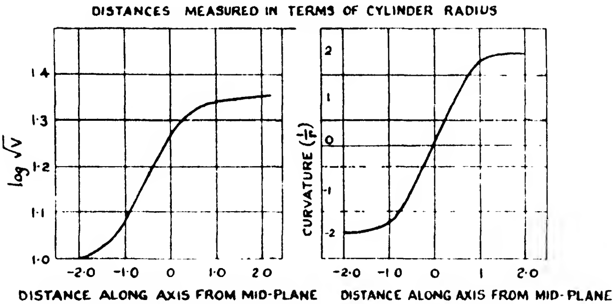


FIG. 94. Graphs of log $\sqrt{V}$ vs. z, and $1/r$ vs. z for a Two-tube Lens.

If the convergence by the lens of a pencil of incident rays is required, then commencing with values of U of, say, 1°, 2°, 3° etc., a long series of computations enables the series of trajectories resulting from the incident pencil of electrons to be ascertained.

L. Jacob* discusses the accuracy of determining a ray trace by a graphical method, using the simple Snell law formula.

Referring to fig. 95, suppose an electron beam emerging from a source S is incident on an electron lens at P_1 on the first refracting equipotential. Let v_0 be the incident electron velocity, and v_1 the velocity after the first refraction, where V_0 and V_1 are the corresponding equipotentials.

Let α = angle of incidence at the first equipotential and β = first angle of refraction, then the initial refractive index $\mu_1 = \frac{\sin\alpha}{\sin\beta} = \sqrt{\frac{V_1}{V_0}}$. Similarly $\mu_2 = \frac{\sin\gamma}{\sin\delta} = \sqrt{\frac{V_2}{V_1}}$, at the second refracting equipotential, $\mu_3 = \frac{\sin\theta}{\sin\eta} = \sqrt{\frac{V_3}{V_2}}$, etc., where α, β, γ, δ,

* L. Jacob, *Phil. Mag.*, **26**, 570, 1938.

θ, η, etc., are successive angles. In general, $\mu_n = \frac{\sin\phi}{\sin\psi} = \sqrt{\frac{V_n}{V_{n-1}}}$.

A useful simplification of the computing work involved is to select equipotentials from the plot so that

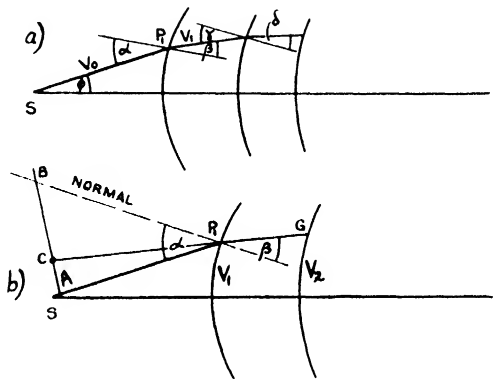


FIG. 95. *a, b,* Ray-tracing Procedure for an Electron Lens (Jacob).

$$\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_n, \text{ etc.}$$

$$\therefore \quad \frac{\sin\alpha}{\sin\beta} = \frac{\sin\gamma}{\sin\delta} = \frac{\sin\phi}{\sin\psi},$$

$$\text{i.e.} \quad \frac{\sqrt{V_1}}{\sqrt{V_0}} = \frac{\sqrt{V_2}}{\sqrt{V_1}} = \frac{\sqrt{V_3}}{\sqrt{V_2}} = \mu.$$

$$\therefore \quad \frac{V_1}{V_0} = \frac{V_2}{V_1} = \frac{V_3}{V_2} = \mu^2.$$

$$\therefore \quad V_1 = \mu^2 V_0; \quad V_2 = \mu^2 V_1 = \mu^4 V_0; \quad \text{and} \quad V_n = (\mu^2)^n V_0.$$

Jacob then applies a tracing method due to Bloch, which may be understood by reference to fig. 95*b*.

A paper scale *AB* is divided at *C* so that $BA/BC = \sin\alpha/\sin\beta = \mu$ (constant). This scale is laid along the incident ray *SP*, and normal so that, when its extremities *A* and *B* are on these lines, then the line CP_1 passes through P_1 to *G*. P_1G is then the refracted ray at surface V_1, and the incident ray at surface V_2. This same procedure is then adopted at equipotentials V_2, V_3, etc., in turn, tracing approximately the beam.

Generally speaking there is little to choose between this method of Jacob, and the procedure due to Klemperer. Both methods will only give sufficiently accurate results if the equipotentials are considered as arcs of circles, which precludes the accurate estimation of extra-axial ray paths.

Jacob's method has the advantage that it is quicker, and will therefore be less tedious when applied to a large number of surfaces through the lens. Since the number of equipotentials chosen should be large to give accuracy, this is a distinct gain. However, there is the outstanding advantage in the Klemperer and Wright method that it is unnecessary to rely on the inaccuracy of drawing the lines from one equipotential to the next on a chart as they are calculated.

The chief application of a trigonometrical procedure is to indicate the effects brought about by various suggested lens designs, and modifications of such designs.

The Differential Equation for the Electron Trajectory in an Electrostatic Field. Maloff and Epstein* have developed an ingenious method of computing the path of an electron through an electrostatic field, basing their computations on a discussion of the differential equations associated with the electrodynamics of a charge moving in an electrostatic field.

Consider the important case of an electrostatic field symmetrical about an axis, see fig. 96. Cylindrical coordinates are adopted as a means of obtaining relatively simple mathematical expressions.

Thus let point T in the field have cylindrical coordinates r, z, where z is the distance along the axis from some preselected origin, and r is the distance from the axis to the point T along the radius. The potential in the field at any such point as T can be expressed as some function of r and z, say $V(r, z)$, where each equipotential surface can be represented by

$$V(r, z) = \text{a constant.} \qquad (238)$$

In accordance with the fundamental definition of an electrostatic field (p. 7), the magnitude and direction of the force acting on a unit positive charge placed in the field is given by

$$F = \frac{-dV}{dn}, \qquad (239)$$

where dV/dn is considered along the normal to the equipotential surface.

* I. G. Maloff and D. W. Epstein, *Proc. Inst. Rad. Eng.*, **22**, 1386, 1934.

This force can be resolved in the direction of r and in the direction of z, so that the component forces on an electron of charge $-e$ become

$$Fr = e\frac{\partial V}{\partial r} = m\frac{d^2r}{dt^2}. \qquad (240)$$

$$Fz = e\frac{\partial V}{\partial z} = m\frac{d^2z}{dt^2}. \qquad (241)$$

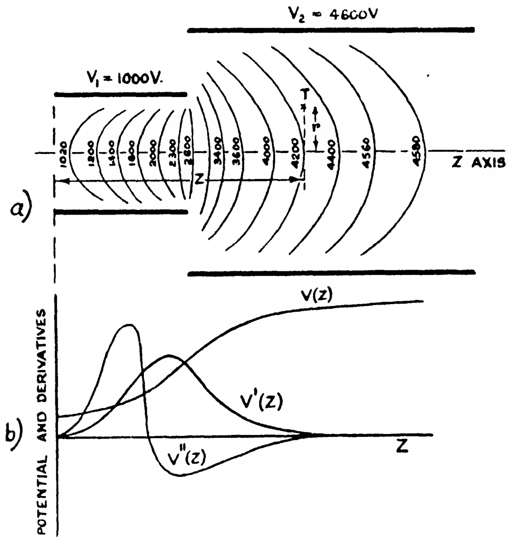


FIG. 96. *a*, Equipotential Line Plot between two Cylindrical Electrodes. *b*, The Distribution of Potential and its Derivatives along the Axis.

The voltages on the electrodes of the system are known. It is required to determine the potential distribution in the space between them. In the case of cylindrical electrodes, then for a plane drawn to include the principal axis, this problem can be stated in mathematical form in that the Laplace equation to be solved is:

$$\frac{\partial^2 V}{\partial r^2} + \frac{1}{r}\frac{\partial V}{\partial r} + \frac{\partial^2 V}{\partial z^2} = 0 \qquad (242)$$

on putting equation (15), p. 12, in cylindrical coordinates instead of Cartesians, two dimensions only being involved, and not three, since the potential distribution in a plane is sought, and not in a solid figure.

In the case of paraxial electron beams, Maloff and Epstein evolve a differential equation for the electron trajectory in the form

$$\frac{d^2r}{dz^2}+\frac{V_0'}{2V_0}\frac{dr}{dz}+\frac{V_0''}{4V_0}r=0, \qquad (243)$$

where $V=V_0$, $\partial V/\partial z=V_0'$ and $\partial^2 V/\partial z^2=V_0''$ since it can be proved that in the case of electron beams travelling close to the principal axis, then $V(r, z)$ can be considered as some function $V_0(z)$, written as V_0, implying that second and higher powers of dr/dz and r are neglected.

This equation (243) is the fundamental electron-optics equation relating to paraxial rays in axially symmetrical electrostatic fields. Its importance lies in the fact that if, from an electrolytic trough plotting, the axial distribution of potential, together with its first and second derivatives are known, then the trajectories of the paraxial electrons can be found. Vice versa, if the trajectory and its first and second derivatives are known, then the axial distribution of potential that will produce it can be determined.

It is to be noted that equation (243) is homogeneous in r and z, so if the electrode dimensions are all changed by a constant factor, the trajectory is increased by the same factor, its shape remaining the same. Likewise, multiplication by a constant factor of the electrode potentials operating an electron lens does not alter the electron trajectory shape.

In the case of estimating the electron path through an electrostatic field which has been plotted with the aid of an electrolytic trough, no simple form for $V_0(z)$ is likely. A step by step method of integrating equation (243) therefore becomes necessary, and the method begins to bear similarities with the previously discussed procedure of using optical computing methods.

From the electrolytic trough readings, a graph is drawn of $V_0(z)$ vs. z, and by graphical methods, the derivatives $V_0'(z)$ and $V_0''(z)$ are obtained from the initial curve. See fig. 96*b*.

Suppose the z axis is divided into n parts, where each section is sufficiently small to permit that $V_0''(z)/4V_0(z)$ and $V_0'(z)/2V_0(z)$ of equation (243) are constant during the interval. Let these expressions be $-2B$ and $-2A$ respectively during such an interval.

Equation (243) then becomes

$$\frac{d^2r}{dz^2}-2A\frac{dr}{dz}-Br=0. \quad . \quad . \quad . \quad (244)$$

Solution is

$$r=\frac{r_n'-r_nm_2}{2\sqrt{(A^2+B)}}\cdot e^{m_1z}-\frac{r_n'-r_nm_1}{2\sqrt{(A^2+B)}}\cdot e^{m_2z}, \quad . \quad (245)$$

where the nth interval is considered, r_n and r_n' being known values for r and dr/dz at the beginning of this interval, and $m_1=A+\sqrt{(A^2+B)}$, $m=A-\sqrt{(A^2+B)}$; r at any point during the nth interval is readily found, and r_n' can be obtained by differentiation of (245).

These values of r_n and dr_n/dz at the end of the nth interval are then used as r_{n+1}, $r_n'_{+1}$ at the beginning of the next interval; in this way the complete path is computed.

The chief disadvantage of this method lies in the inaccuracies associated with drawing a curve for the derivatives $V_0'(z)$ and then $V_0''(z)$ from a knowledge of the experimentally obtained graph of $V_0(z)$.

Electro-mechanical Ray Tracing Apparatus. D. Gabor* and D. B. Langmuir† have introduced ingenious contrivances for determining the trajectory of an electron moving through an electrostatic field.

For an electron moving with velocity v acquired by undergoing a potential drop of V

$$\tfrac{1}{2}mv^2=Ve. \quad . \quad . \quad . \quad . \quad . \quad (246)$$

If E_N is the field intensity normal to the path of the electron, then $E_Ne=mv^2/r$, where r is the radius of curvature of the electron path.

Combining these equations, therefore

$$r=\frac{mv^2}{E_Ne}=\frac{2Ve}{E_Ne}=\frac{2V}{E_N}. \quad . \quad . \quad . \quad . \quad (247)$$

This conveniently simple relation for determining the radius of curvature of the electron trajectory is applied in the methods of Gabor and Langmuir. The electron lens is represented by a model enlarged to some suitable scale (say 5 : 1) placed in an electrolytic

* D. Gabor, *Nature*, **139**, 373, 1937.

† D. B. Langmuir. *Nature*, **139**, 1067, 1937.

trough. The electron is represented by a probe, V being its potential with respect to the cathode source. The magnitude of E_N is determined by side-probes, placed closely on either side of the main probe at equal distances. By means of a mechanical link mechanism these side probes are maintained so that the line through them, and the main probe, is always perpendicular to the path. A specially designed measuring bridge is used for determining the necessary values of $2V/E_N$ at various points in the electric field, in conjunction with a mechanism for directly recording the electron trajectories on drawing paper.

Defects in Electron Lenses. The types of defects that may arise in electron lenses may be readily classified in two categories: (*a*) electron-optical defects, all of which have equivalent types in normal glass optics; (*b*) defects purely electrodynamic in origin.

In category (*a*) there are:

(1) Spherical Aberration.
(2) Chromatic Aberration.
(3) Astigmatism.
(4) Coma.
(5) Curvature of the Field.
(6) Pin-cushion and Barrel Distortion.

In category (*b*) there are:

(1) Space-charge effects, resulting from the mutual repulsions between like electrical charges.

(2) Defects are mechanical in origin, due to imperfect alignment of electrode apertures.

(3) Effect due to barrier layers at the surfaces of electrodes which cause a difference of potential between the actual electrode and the space immediately outside it.

The defects in category (*a*) are all illustrated in any good textbook of glass optics. It is therefore proposed to deal here with only particular information relating to the electron-optical cases.

Spherical aberration is the greatest source of difficulty in the design of an electron lens, since this aberration is concerned with the use of wide-aperture lenses, so that as many electrons as possible traverse the lens, so that non-paraxial beams are involved. This aberration is approximately proportional to D^2, where D is the diameter of the lens aperture. Generally, with a two-electrode cylindrical electron lens, the spherical aberration is positive on the low potential side, and negative in the high potential region, so the reduction of such aberration becomes a question of matching the positive against the negative component by a suitable modification of the equipotential surfaces. Some work has been done on removing spherical aberration by neutralising it with the opposite effect of controlled space-charges in the electron beam, and

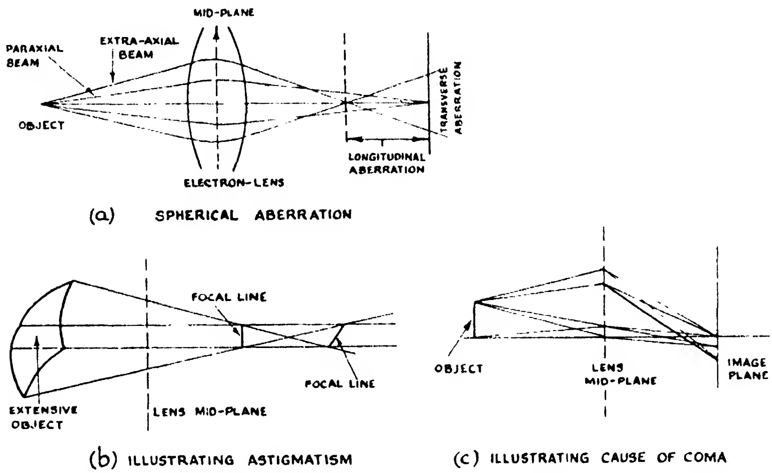


FIG. 97. Electron-lens Defects.

secondly by the use of electron mirrors. Generally speaking, however, the complete removal of spherical aberration defects from an electron lens is impossible. The only satisfactory remedy is to reduce the lens aperture to such a small size that this aberration is not marked.

Chromatic aberration arises from a non-uniform velocity distribution among the electrons forming the beam.* The effect can be either an axial displacement, or lateral transverse variation in the size of the focused image.

In the usual two- or three-cylinder lens, where the minimum potential with respect to the cathode is greater than 50 V., chromatic aberration due to variation in the incident electron beam velocity is small because the thermal velocities imparted to the emitted electrons at the cathode are negligible compared

* See p. 30.

with the velocities acquired by an electron undergoing a potential fall of 50 V. However, secondary electrons, released from the electrodes and apertures, may cause image defects because of the wide velocity distribution among such secondaries giving rise to chromatic aberration. Supply voltage fluctuations are also a cause of chromatic aberration, particularly in high voltage instruments such as the electron microscope. In the electron immersion lens (see p. 258) comprising a cathode, negative Wehnelt cylinder and positive anode, the chromatic aberration due to the Maxwellian distribution of velocities among the thermionically emitted electrons is considerable.

Astigmatism is produced in electron beams from an object point at a considerable distance from the axis. Instead of a point image being formed, rays in the image space converge towards two lines at right angles, at different distances from the object. The astigmatic difference, or distance between the focal lines, increases with the square of the distance of the object point from the axis, but is independent of the aperture. The length of the focal lines increases in direct proportion to the aperture diameter.

Coma is a source of bad imaging when an extended object is used. The amount of coma increases as the distance of the image point from the axis, and as the square of the aperture.

Pin-cushion and barrel distortion result when either the central portions of the image are more magnified than the outside portions (barrel distortion) or when the central portions are the less magnified (pin-cushion). These are due to a combination of the defects already discussed.

Some Typical Electrostatic Electron Lenses. In an electrostatic field, the refraction of moving electrons at each equipotential depends on V_1/V_2; moreover, equation (243) for the electron trajectory is homogeneous in V, r and z. These facts indicate that the absolute values of the potentials on the lens electrodes, and the dimensions of these electrodes are relatively unimportant in defining the focusing effects: it is the ratio of the potentials on the lens that is important.

Thus the shape of the electron trajectory is unaltered if the potentials are each multiplied by a constant factor, or if the diameters of the tubes of the lens are enlarged in the same proportion. This leads to a very useful simplification of the analysis of an electron lens; it is not necessary to consider all the infinite

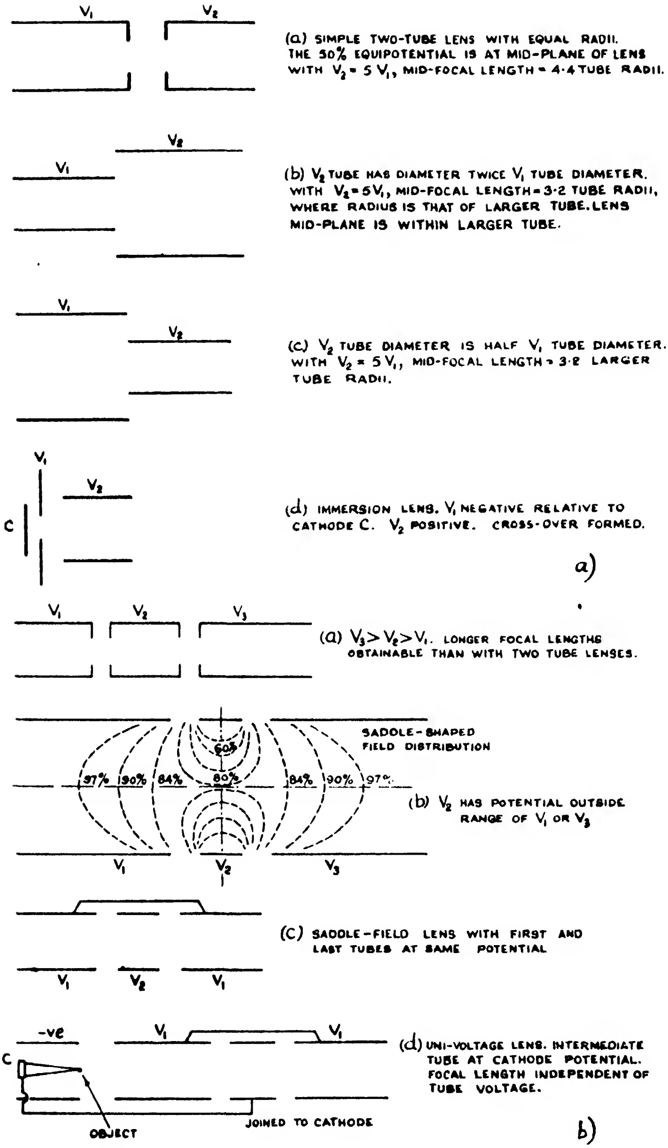


FIG. 98. *a*, Some typical Bi-potential Electrostatic Lenses.
b, Typical Three-tube Lenses.

variety of possible lens combinations by separate computing treatments. For example, in the case of the two-tube lenses, or so-called bi-potential lenses, if the cardinal points of such lenses are known for various voltage ratios, then any other similar lens, enlarged in scale, or with higher or lower voltages on the electrodes, can be related to the known case by the methods of simple proportion.

A three-tube lens on which the potentials V_1, V_2 and V_3 are such that $V_3 > V_2 > V_1$ is frequently used. Still further cylinders may be inserted, giving a very gradual change in the radii of curvature of the equipotentials of the electrostatic field, and consequently a great focal length.

A new state of affairs will exist if the intermediate tube potential V_2 in a three-tube lens is outside the range of V_1 to V_3. Then an electrostatic potential distribution like that illustrated in fig. 108 is obtained, which receives the name "saddle-shaped". The "einzel" lens is a particular type, where V_2 is outside the range of V_1 and V_3, and moreover $V_1 = V_3$. The symmetrical arrangement so obtained will ensure that the two focal lengths are the same.

A "univoltage" lens is one in which the intermediate electrode of an "einzel" lens is at cathode potential, so that only one value of voltage is necessary to achieve an electron lens—that of the two outside cylinders. Such a lens has a focal length that is independent of the single operating potential used, but depends only on the cylinders' dimensions and spacings. This is of great practical importance when a very stable focused image is required, focusing being independent of any possible supply voltage fluctuations. An example of its application is in the case of the electrostatic electron microscope (p. 280). The independence of the focal length on the voltage on the outside electrodes follows from the fact that the lens voltage ratio is necessarily infinity, irrespective of the actual potential used, because the central electrode is joined to the cathode, at zero potential. So the focal length must depend on the lens dimensions only, and in particular on the dimensions of the intermediate electrode. The only variations of focal lengths that can be brought about in practice are thus due to the small effect of the distribution of velocities amongst the thermally emitted electrons, and to the relativistic mass correction effect, (p. 43) necessary if the electron is travelling at great speed. This latter effect is only noticeable if a supply greater than 30 kV. is subjected to a variation of more than 3%.

The Magnetic Electron Lens. H. Busch,* in a paper which may be said to have laid one of the foundation stones of electron-optical science, discusses the effect of an axially symmetrical magnetic field, produced by a coil of wire carrying an electric current, in focusing a beam of electrons.

Consider the case illustrated in fig. 99, where an electron moves through a homogeneous magnetic field in a direction making an angle θ with the lines of force. Suppose v is the electron velocity, and is resolved into components v_x along the lines of magnetic force, and v_y at right angles to these lines.

Then $v_x = v \cos \theta$, and is not affected by the magnetic field, whilst $v_y = v \sin \theta$.

The time t taken by an electron to make one revolution is

$$\frac{2\pi\rho}{v_y} = \frac{2\pi m}{He} \quad \text{from (64).}$$

$$\therefore \quad t = \frac{2\pi m}{He} \quad . \quad . \quad . \quad . \quad . \quad (248)$$

The fact that a uniform magnetic field possesses a converging† property on electrons moving in a beam depends on the fact that this value for t is independent of velocity, radius of curvature and the angle which the beam of electrons originally makes with the field. During the time t the electron moves forward along the lines of force through a distance $v_x t = v \cos \theta \, (2\pi m/He)$. If a beam of electrons leaves the axis at point P, and the angle which the diverging beam makes with the axis is small enough to consider that $\cos \theta = 1$ (i.e. a paraxial beam), then all these electrons will cross the same line of force again at another point Q, since they all complete one revolution, returning to the axis, in the same time. Thus an object in the form of a thermionic or photo-cathode emitting electrons will be imaged at a fluorescent screen at some finite distance, this distance being governed by the axial velocity component of the electron, and the time t. The axial velocity can be imparted to the electrons by a suitably positioned electric field.

* H. Busch, *Arch. Electrotechnik*, **18**, 583, 1927; and *Ann. der Physik*, **81**, 974, 1926.

† It should be noted that a long solenoid does not possess a focusing action in the true sense of the term, since a *parallel* beam of electrons incident on such a field will not be converged, there being no radial component of the magnetic field in such a solenoid.

If the angle θ is not sufficiently small to allow $\cos\theta=1$, then the focusing action will not be so well defined, since

$$\cos\theta=1-\frac{\theta^2}{2!}+\frac{\theta^4}{4!}-\text{ etc.}, \qquad (249)$$

and θ^2 is not negligible. The result will be that the lens image will be spread out along the axis, giving longitudinal spherical aberration.

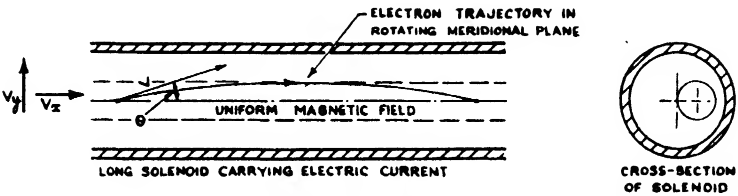


FIG. 99. Electron Trajectory in a Homogeneous Magnetic Field.

The mathematical treatment given by Busch for a short focusing coil is too complex to be reproduced here. It must suffice to quote useful results which are proved in his paper.

Assuming that (*a*) the electrons emerge from a point; (*b*) the magnetic and electric fields acting and the space-charge possess axial symmetry, and this axis passes through the source of electrons; (*c*) that the electron beam is narrow, and its axis coincides with the axis of symmetry; and (*d*) the electrons entering the fields have uniform velocity, then the following results are useful:

(1) A short, narrow, magnetic coil, electric field zero. A formula comparable with the case of a condenser lens in glass optics is

$$f=\frac{4v^2}{\eta^2\int H_0{}^2(z)dz}, \qquad (250)$$

where the principal axis of symmetry is the z-axis, and $H_0(z)$* is the magnetic field strength at the z-axis, being a function of the distance z along this axis. The integral is carried out over the length of the z axis through the coil for which the magnetic field is present. The constant $\eta=e/m$, where e and m are the electron

* Obtained from graph of field strength vs. distance along z axis, the result of practical measurement.

charge and mass respectively, whilst v_0 is the initial velocity of the electrons along the z axis. The focal length of the lens is f.

It is to be noted that the electron beam, after traversing the coil, does not travel in the original meridional plane, but in another plane which is turned through an angle θ with respect to the first.

$$\theta = \frac{-1}{\sqrt{f}} \cdot \frac{\int H_0(z)dz}{\sqrt{[\int H_0(z)dz]}} \quad . \quad . \quad . \quad . \quad (251)$$

The angle of rotation is increased if the extent of the magnetic field along the z axis is increased. Thus if z is increased in the ratio 1 : c, where $\rho = z/c$, then

$$\theta = -\sqrt{\frac{c}{f}} \cdot \frac{\int H_0(\rho)d\rho}{\sqrt{[\int H_0(\rho)d\rho]}} \quad . \quad . \quad . \quad (252)$$

evincing that the rotation angle is proportional to $\sqrt{c}$.

(2) A short magnetic coil with a superimposed electrostatic field. For simplicity, assume the electrostatic field consists of uniform equipotential surfaces perpendicular to the z-axis, i.e. the electric potential depends only on z.

Then
$$f = \frac{4V_0}{\eta^2 \int_{-\infty}^{+\infty} \frac{H_0^2(z)dz}{\sqrt{[v^2 - 2\eta g(z)]}}}, \quad . \quad . \quad . \quad (253)$$

where $V = g(z)$, a function of z, V being the electric potential at any point z along the axis of symmetry from the source of electrons. V_0 is the electron velocity in electron-volts where the electrostatic field is practically zero.

Bouwers* has introduced a much simpler treatment than Busch's for determining the angle of twist of the image compared with the object, and the focal length in the case of an axially symmetrical magnetic field, basing his analysis on the determination of the radial† and lateral equations of force acting on the electron.

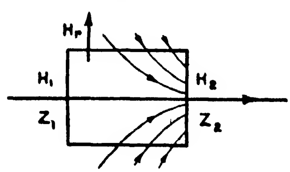


Fig. 100. Simple Treatment of Short, Magnetic Focusing Coil.

Let a magnetic field extend for a short distance along the z-axis of fig. 100, between points z_1 and z_2. Consider that

* A. Bouwers, *Physica*, **4**, 200, 1937.

† It is because there is a considerable radial component of the magnetic field in the case of a short coil that focusing of an incident parallel beam is possible.

a cylinder of radius r symmetrically disposed about the z-axis contains the non-uniform magnetic field, the magnetic field strength being H_1 at z_1, and H_2 at z_2.

Then the magnetic flux which passes through the circular cross-section at z_1 is $\pi r^2 H_1$, whereas the flux through the circle at z_2 is $\pi r^2 H_2$.

The difference between these two fluxes must be equal to the number of lines of force which permeate the cylindrical wall of the lens. If the radial magnetic field has an average value of H_r, then the total radial flux through the cylinder is $H_r \,.\, 2\pi r(z_2 - z_1)$, on multiplying the field strength by the area of the cylinder walls.

$$\therefore \quad H_r \,.\, 2\pi r(z_2 - z_1) = \pi r^2 (H_2 - H_1).$$

$$\therefore \quad H_r = \frac{r(H_2 - H_1)}{2(z_2 - z_1)} = \frac{r}{2} \cdot \frac{dH_z}{dz}. \qquad (254)$$

where dH_z/dz is the rate of change of the magnetic field along the z-axis.

If an electron with velocity v parallel to the z-axis enters this magnetic field at a distance r from the axis, then it will be acted on by the radial component of the magnetic field. The result will be a component of force acting on the electron in a direction perpendicular to both its initial direction, and to the radial lines of force, i.e. it will receive a lateral force component. The magnitude of this force will, from equation (62), be evH_r.

Hence if $\omega = d\theta/dt$ is the angular velocity with which the electron consequently rotates about the z-axis, then $r(d\theta/dt)$ is its linear velocity tangential to this circular motion, and its linear acceleration will be $\frac{d}{dt} \,.\, r\frac{d\theta}{dt}$.

$$\therefore \quad m\frac{d}{dt}\left(r\frac{d\theta}{dt}\right) = evH_r, \qquad (255)$$

since mass $\times$ acceleration $=$ force.

From (254) and (255)

$$\frac{d}{dt}\left(r\frac{d\theta}{dt}\right) = \frac{erv}{2m} \cdot \frac{dH_z}{dz}. \qquad (256)$$

Put $v = dz/dt$, since the velocity of the electron parallel to the z-axis is the rate of change of z with time.

$\therefore \quad \dfrac{dz}{v}=dt$, so (256) becomes $\dfrac{d}{dt}\left(r\dfrac{d\theta}{dt}\right)=\dfrac{er}{2m}\cdot\dfrac{dH_z}{dt}$.

Integrating,
$$r\frac{d\theta}{dt}=\frac{erH_z}{2m}. \qquad (257)$$

$$\therefore \quad d\theta=\frac{eH_z\,.\,dt}{2m}=\frac{eH_z\,.\,dz}{2mv}.$$

Integrating again, $\theta=\dfrac{e}{2mv}\displaystyle\int_{-\infty}^{\infty}H_z dz$* which, in practical units, gives the angle of rotation in radians, where the electron velocity is in volts, V, and the magnetic field is in oersted, as

$$\theta=\frac{0\cdot 5}{\sqrt{V}}\int_{-\infty}^{\infty}H_z dz. \qquad (258)$$

To find the focal length f of such a lens, consider that the centrifugal force acting on the electron must be equal to the force directing the electron towards the z-axis.

The centrifugal force $=\dfrac{mv^2}{r}=\dfrac{m[r(d\theta/dt)]^2}{r}$.

But $r\dfrac{d\theta}{dt}=\dfrac{erH_z}{2m}$, from (257).

$$\therefore \quad \frac{m[r(d\theta/dt)]^2}{r}=\frac{m}{r}\cdot\left(\frac{erH_z}{2m}\right)^2=\frac{e^2H_z^2r}{4m}. \qquad (259)$$

This must be equal to the electron mass multiplied by the radial acceleration, i.e. $m(d^2r/dt^2)$.

$$\therefore \quad \frac{d^2r}{dt^2}=\frac{e^2H_z^2r}{4m^2}. \qquad (260)$$

$$\therefore \quad \frac{d}{dt}\left(\frac{dr}{dt}\right)=\frac{e^2H_z^2r}{4m^2},$$

but $\quad dt=\dfrac{dz}{v}, \quad$ therefore $\dfrac{d}{dz}\left(\dfrac{dr}{dt}\right)=\dfrac{e^2H_z^2r}{4m^2v}$.

Integrating,
$$\frac{dr}{dt}=\frac{e^2r}{4m^2v}\int_{-\infty}^{\infty}H_z^2dz+C. \qquad (261)$$

* The integral is from ∞ to $-\infty$ since the magnetic field along the z-axis will extend thus far. In practice, the magnetic field will be insignificant at a short distance before and behind the lens, so the integration, usually performed graphically from a plot of H_z vs. z, is over a convenient distance, depending on the rate of decrease of the field away from the lens centre.

At the point where the initial electron beam enters the magnetic lens, the magnetic field is zero, and $dr/dt=0$ if the electron beam enters the magnetic lens parallel to the z-axis. Therefore $C=0$ in equation (261).

$$\therefore \quad \frac{dr}{dt}=\frac{e^2r}{4m^2v}\int_{-\infty}^{\infty} H_z^2 dz. \qquad (262)$$

With this radial component of velocity dr/dt, the electron will meet the z-axis after a time $t=\frac{r}{dr/dt}$, since it enters the lens at a distance r from the axis.

The distance vt will then be the focal length of the lens, since the electron beam is assumed to be initially parallel to the axis, and the axial velocity component v of the electron is unchanged.

$$\therefore \quad f=vt=\frac{vr}{dr/dt}=\frac{4m^2v^2}{e^2\int_{-\infty}^{\infty} H_z^2 dz} \text{ from (262).} \qquad (263)$$

Putting the electron velocity in volts V, and the magnetic field in oersted, then the focal length f in cm. is obtained as

$$\frac{1}{f}=\frac{0{\cdot}022}{V}\int_{-\infty}^{\infty} H_z^2 dz. \qquad (264)$$

Ruska* gives a formula

$$NA=326S\sqrt{[KV+(KV)^2/1000]}\sqrt{(R/f)}, \qquad (265)$$

for the number of ampere-turns NA necessary to obtain a focal length f *cm.*, where KV. is the electron-velocity in kilovolts, R is the coil radius and S is a constant depending on the geometry of the magnetic field, and amount of iron used.

* E. Ruska, *Zeits. Physik.*, **87**, 580, 1932, and **89**, 90, 1934.

CHAPTER 11

The Cathode-Ray Tube

In the cathode-ray tube as used for measuring purposes, or as a television receiving tube or radar indicating device, an electron gun is used to produce a fine beam of electrons which are focused

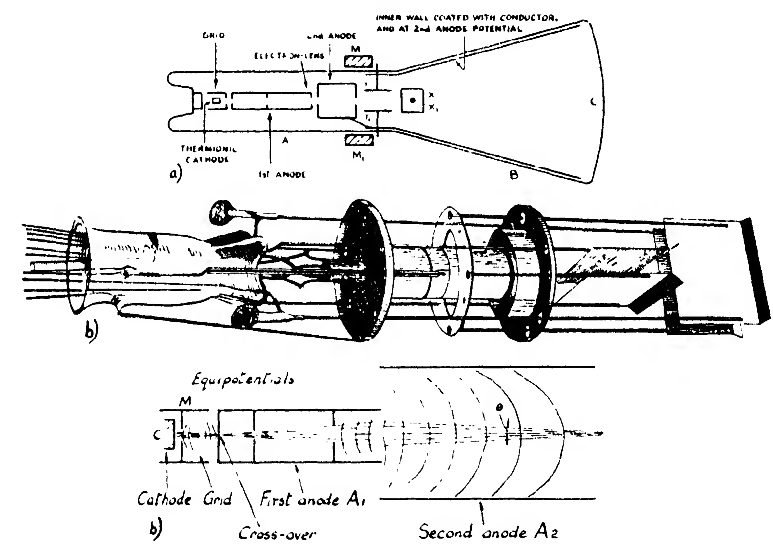


FIG. 101. *a*, The Cathode-Ray Tube. *b*, An Electron Gun.

to a point at a fluorescent screen, where their impact is made visible by the production of light. This electron beam has to be deflectable so that the fluorescent spot at the screen can be moved in two directions mutually at right angles. As a measuring oscillograph this instrument is capable of recording the wave-forms of all sorts of alternating potentials and currents at frequencies between 10 c./s. and 10 Mc./s.

Referring to fig. 101*a*, the electron gun is mounted axially along the glass tube *A*. This glass tube is sealed to the conical vessel *B*, on the curved surface of which is deposited a fluorescent screen at *C*. This envelope is highly evacuated. The gun produces a beam of electrons which are focused to give a bright pin-point of light at the centre of the fluorescent screen. This electron beam, and so the light spot, is capable of being deflected horizontally and vertically by the application of P.D.'s across, respectively, the pairs of electrostatic deflecting plates XX_1 and YY_1. Alternatively, deflection of the electron beam can be achieved magnetically by the passage of a current through the pair of coils MM_1 which surround the gun tube. If an alternating potential of any wave-form is applied to the vertical X-plates, a horizontal line is traced out on the screen by the rapidly moving spot. The simultaneous application of an alternating potential to the Y-plates, when the Ω-plate alternating potential has a particular saw-tooth form, will produce on the screen the wave-form of the potential applied to the Y-plates. To understand how this useful operation is brought about, the action of the essential parts of the tube are now considered in detail.

The Electron Gun. To form a basis for discussion, a typical gun due to Zworykin, fig. 101*b*, is considered. The important parts are:

(*a*) The thermionic cathode.

(*b*) The Wehnelt cylinder, grid or modulator.

(*c*) An electron lens consisting usually of two cylindrical, co-axial tubes, called the first and second anodes.

Other electrodes may be added to produce required features; these will be considered later.

(*a*) *The Thermionic Cathode.* The requirements in this case are similar, in most respects, to those in a valve. The same materials are used, commonly a mixture of barium and strontium oxides, activated in the vacuum, on the surface of an indirectly heated nickel base. The special requirements in a cathode-ray tube are a source of electrons of small area, the emission (in mA./sq. cm.) to be as great as possible, and the velocity distribution amongst the emitted electrons to be confined to a limited range of values. If alternating current is to be used to heat the filament inside the cathode, then the alternating magnetic field produced must be restricted to reduce the possibility of its affecting the electron motion immediately outside the cathode.

The only method of restricting the range of velocities amongst the emitted electrons, so avoiding chromatic aberration, is to use as low a temperature cathode as possible; the emitter material must hence have as low a work function as possible.

D. B. Langmuir* gives the following equation relating the beam current to the cathode current density:

$$I_B = I_C \left(\frac{Ve}{kT} + 1\right) \sin^2 \theta, \quad . \quad . \quad . \quad (266)$$

where I_B = current density in beam (electrons/sq. cm. of cross-section).

I_C = cathode current density (electrons/sq. cm. of cross-section).

T = absolute temperature of cathode.

V = final voltage acting on the electrons.

θ = semi-angle subtended by the beam at the final focused electron spot at fluorescent screen.

e = electron charge.

k = Boltzmann's constant.

This important equation has had a great influence on the trend of electron-gun design, since it indicates the lines along which development must proceed if a high beam current is to be obtained. The possibilities are:

(1) Increase the voltage V.

(2) Increase the angle θ, necessitating electron lenses capable of focusing wide angle beams of electrons, i.e. lenses in which large apertures are used. The difficulties attending this demand have been discussed in the chapter on electron optics.

(3) Produce a cathode of higher specific emission per sq. cm., i.e increase I_C. This is a physico-chemical problem: a higher free barium content in the cathode surface is necessary to obtain a lower work function. It is unlikely that a stable thermionic cathode more efficient than those at present available will be produced in the near future.

(*b*) *The Wehnelt Cylinder.* In most types of cathode-ray tube electron gun, an initial concentration into a beam, accompanied by control of the beam current, is accomplished by surrounding the cathode with a coaxial nickel cylinder maintained at a negative potential with respect to the cathode. The electrons, being negatively charged, cannot reach the negatively charged walls of this

* D. B. Langmuir, *Proc. Inst. Rad. Eng.*, 25, 977, 1937.

cylinder, so that their otherwise divergent paths are rendered convergent. Fig. 102*b* illustrates the advantage in that many more of the total number of electrons emitted will be able to penetrate the aperture in the first anode.

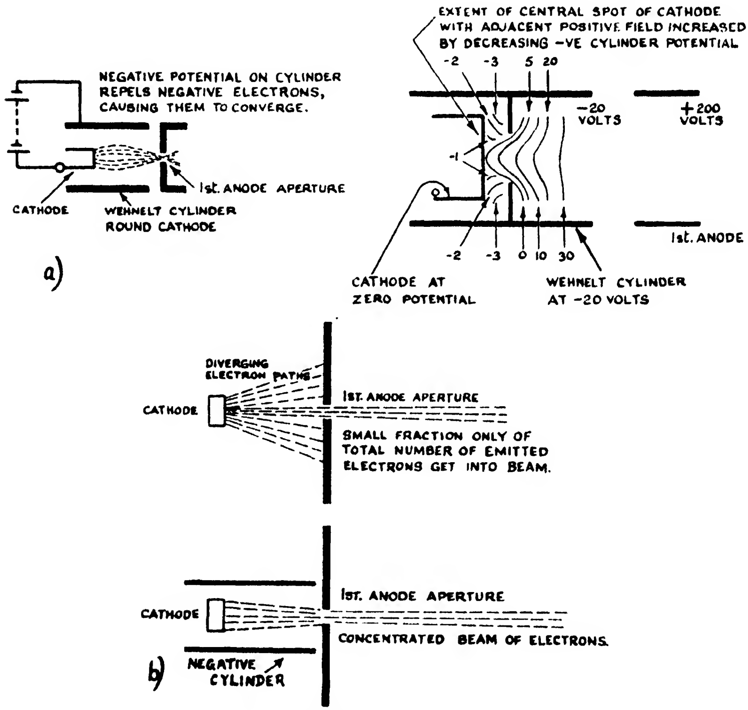


FIG. 102. *a*, The Wehnelt Cylinder and Electron Immersion Lens. *b*, Action of Wehnelt Cylinder.

This Wehnelt cylinder forms, in conjunction with the first anode, an "immersion" lens, so-called by analogy with the oil-immersion system used in a light microscope. Fig. 102*a* gives the distribution of equipotentials associated with this combination. The magnitude of the negative potential on the Wehnelt cylinder (frequently called the grid, or modulator) determines, in relation to the positive first anode potential, the portion of the cathode surface which can emit electrons. If the first anode potential is kept constant, then this area is decreased as the grid potential is made more negative. As the negative grid bias is increased, so the negative electron retarding field at the cathode surface extends inwards from the cathode surface periphery, so that with large negative bias, all but the central spot of the cathode is prevented from emitting electrons into the main beam. A sufficiently great negative grid bias will completely cut off the beam current. So the grid potential controls the magnitude of the beam current, and therefore the brilliance of the final fluorescent spot at the screen.

Maloff and Epstein* show that this immersion lens can be considered as two lenses separated by a mid-plane. The ratio of the refractive indices of these two lenses will be very large because the potential gradient immediately outside the cathode is small, whereas at the entry to the first anode region, it is very great. It follows that all rays coming from an object at any distance before the effective mid-plane will be focused close to the focal point of the lens after the mid-plane. A distorted image of the cathode will thus be formed in the "cross-over" region (fig. 102*a*).

The "cross-over" produced, formed usually at the first anode aperture by appropriate spacing, is the "object" of which the electron lens formed between the first and second anodes, produces a reduced image at the fluorescent screen.

E. Ruska† gives a formula for h, the radius of the "cross-over",

$$\frac{h}{h_0}=\frac{2}{\sqrt{[(V_A/V_{EM}).\sin 2\theta]}}, \quad . \quad . \quad (267)$$

where h_0 = radius of emitting cathode.

V_A = first anode voltage.

θ = semi-vertical angle of the cone of electrons in the beam forming the "cross-over".

V_{EM} = average initial electron emission velocity.

Substitution in this formula indicates a demagnification of ten to twenty times between the cathode and its "cross-over" image, with a first anode voltage of 200. Thus a 0·1 mm. diameter cross-over "object" of high electron density is obtainable using a cathode of 2 mm. diameter. In this formula no allowance is made for the considerable space-charge effect in the "cross-over".

(*c*) *The Electron Lens.* The usual electron lens in a cathode-ray tube electron gun consists of a two-tube construction in which the

* I. G. Maloff and D. W. Epstein, *Electron-Optics in Television*, McGraw-Hill, 1938.

† E. Ruska, *Zeits. Physik*, **83**, 684, 1933.

second anode diameter is greater than that of the first anode. The second anode potential is defined in relation to the first anode potential so as to produce the required focal length and demagnification of the electron cross-over object, in accordance with equations 230 and 236, for the focusing effects of electrostatic fields, and the cardinal points of an electron lens.

An investigation is necessary to determine the effect of the second anode potential on beam current, in relation to the first anode potential and grid bias values. Again, the second anode

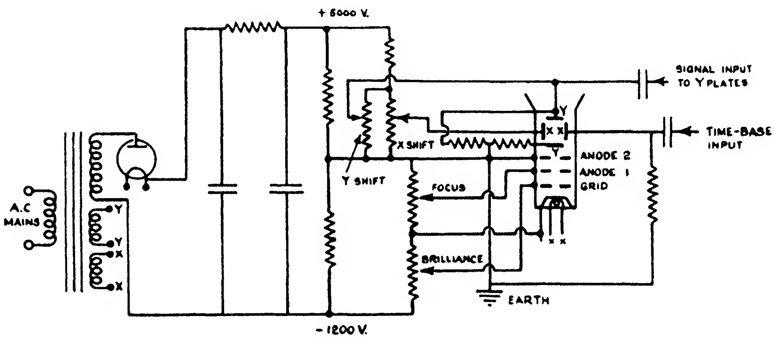


FIG. 103. Power Pack for Cathode-Ray Tube.

potential will decide the electron velocity, and hence the brilliance of the spot at the fluorescent screen, the velocity being proportional to $\sqrt{V_2}$, where V_2 is the second anode potential.

In attempts to obtain high beam currents, the factors demanding increase are the aperture sizes in all electrodes, the first and second anode potentials and the length of the Wehnelt cylinder. However, the values of these are set at an optimum by the focusing effects required. Thus the apertures need to be small enough to define paraxial beams, and to prevent secondary electrons from entering the beam and giving rise to chromatic aberration defects. An optimum length of Wehnelt cylinder is a factor well worth determining in practice, since it decides the efficiency of a gun in relation to the required cut-off bias, and the ratio of beam current to second anode current. Moreover, this length can be varied without too readily introducing aberrations in the focused image. The length and diameter of the first anode are important in deciding the minimum focused spot size, leading to typical electron guns in which the first anode is commonly five or six times the length of the second anode.

The Fluorescent Screen. Certain materials when irradiated by electrons emit light. This fluorescence persists for a time after the excitation has been removed, the "after-glow" being called phosphorescence. A common material used for cathode-ray tube screens is zinc orthosilicate, which occurs in nature as willemite. Generally this material is synthesised, and activated with manganese, since the natural form is unreliable because of the presence of impurities. The light given out on excitation is yellow-green, with a maximum intensity usefully occurring at a wave-length of about 5200 Å, near the peak sensitivity of the human eye. Some short-lived phosphorescence exhibited by this material results in an after-glow of the fluorescent screen, particularly noticeable after recording transients.

Calcium tungstate and cadmium tungstate are frequently used. These emit blue light, with very little phosphorescence. Zinc sulphide, and especially zinc phosphate, exhibit much phosphorescence, resulting in screen traces having long after-glows, in the second case of several seconds. This facilitates the inspection and photography of the traces produced by rapid transients. In television cathode-ray tube practice, an almost white fluorescence can be achieved by the use of zinc trisilicate as a screen material, or by the inclusion of small amounts of various substances in the normally green-blue zinc sulphide, or zinc cadmium sulphide.

The fluorescent screen material is mixed with a suitable non-active binder material and either sprayed on the inside face of the tube, or deposited from a chemical solution.

Deflecting Plates. Beyond the second anode of the cathode-ray tube gun there are supported two pairs of deflecting plates, or alternatively magnetic deflection is employed. In the usual measuring oscillograph, these pairs of plates are arranged with their mid-planes in the plane of the electron beam, the first pair being capable, on the application of a D.C. potential difference across them, of deflecting the beam vertically at the screen; whereas the second pair of plates, situated with their mid-plane at right angles to the first pair, can produce a horizontal deflection. So the first pair of plates are the Y-plates, and the second pair the X-plates. In normal use, where an oscillograph records the wave-form of an alternating source of E.M.F., the X-plates are subjected to the saw-tooth, wave-form voltage output of a time-base circuit which causes the beam at the fluorescent screen to trace out a uniform, horizontal time-axis, whilst the source of alternating E.M.F. of which the wave-form is required applies a P.D. across the Y-plates to produce a vertical deflection.

To investigate the electrostatic deflection sensitivity of the deflecting plates in a cathode-ray tube, consider fig. 104a.

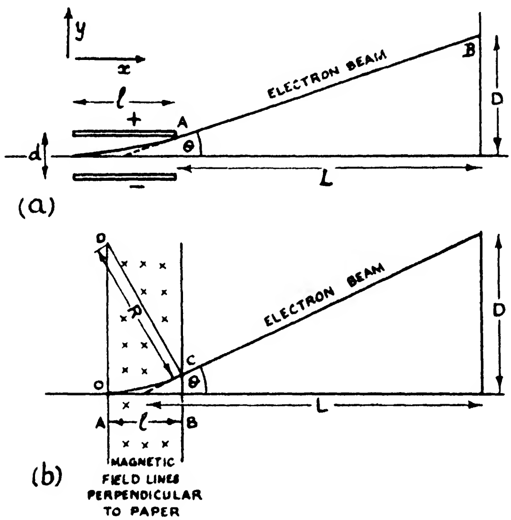


FIG. 104. a, Electrostatic Deflection of an Electron Beam.
b, Magnetic Deflection of an Electron Beam.

Suppose a constant P.D. E_d is applied across the deflecting plates. The electrons enter this deflecting field with a horizontal velocity v due to the accelerating potential V on the final anode of the gun. From (58) $v=\sqrt{(2Ve/m)}$ cm./sec.

Following the analysis given on p. 44, it can be seen that the path of the electrons in the deflecting field will be a parabola, of which the equation is

$$y=\frac{E_d ex^2}{2mdv^2}=Cx^2, \qquad . \quad . \quad . \quad (268)$$

where C is constant.

On leaving the deflecting field the electrons will travel in field-free space to the fluorescent screen. Their path will therefore be a straight line from the point of exit A at the edge of the plates to the point B on the screen. This straight line will be a tangent to the parabola at point A. Hence its slope is given by θ, where

$$\tan\theta=\left(\frac{dy}{dx}\right)_{x=l} \quad . \quad . \quad . \quad . \quad (269)$$

where l is length of deflecting plates.

From (268), put $\frac{E_d e}{md}$=constant C_1 for given deflecting P.D.

then $$\left(\frac{dy}{dx}\right)_{x=l}=\tfrac{1}{2}C_1\frac{d}{dx}\left(\frac{x^2}{v^2}\right)_{x=l}=\frac{C_1 l}{v^2}. \quad . \quad . \quad . \quad (270)$$

Hence the equation of the straight line AB is

$$y=\frac{C_1 lx}{v^2}+a. \quad . \quad . \quad . \quad . \quad (271)$$

where a is a constant.

To evaluate a, consider that at the point A, $x=l$, and $y=\frac{1}{2}C_1(l^2/v^2)$.

Substituting in (271)

$$\tfrac{1}{2}C_1\frac{l^2}{v^2}=\frac{C_1 l\,.\,l}{v^2}+a.$$

Therefore $a=\frac{-C_1 l^2}{2v}$, and the equation (271) becomes

$$y=\frac{C_1 lx}{v^2}-\frac{C_1 l^2}{2v^2}=\frac{C_1 l}{v^2}\left(x-\frac{l}{2}\right). \quad . \quad . \quad . \quad (272)$$

Note that when $y=0$, $x=l/2$, showing that, irrespective of the nature of the parabolic electron path in the deflecting field, the straight line AB can be considered as emerging from the point $x=l/2$, $y=0$ which is the centre point of the deflecting plates.

At the fluorescent screen point B, $y=D$, and $x=L+(l/2)$, where L is the distance from the exit point on the axis of the deflecting plates to the centre of the screen, and D is the y-deflection which the electrons undergo at the screen.

Substituting these values for x and y in equation (272) gives

$$D=\frac{C_1 l}{v^2}\left(L+\frac{l}{2}-\frac{l}{2}\right)=\frac{C_1 lL}{v^2}.$$

But $C_1 = \frac{E_d \cdot e}{md}$, and $v = \sqrt{\frac{2Ve}{m}}$, so that

$$D = \frac{E_d e \cdot lL}{md(2Ve/m)} = \frac{E_d \cdot lL}{2dV} \quad . \quad . \quad . \quad . \quad (273)$$

Equation (273) shows that the deflection on the screen of a cathode-ray tube is directly proportional to the deflecting potential difference E_d applied between the plates. This is valuable since it means that the oscillograph is a linear device, a very fortunate circumstance allowing alternating E.M.F. wave-forms to be recorded without distortion.

The deflection sensitivity S is defined as the deflection in cm. at the screen per volt applied to the deflecting plates.

Thus $$S = \frac{D}{E_d} = \frac{lL}{2dV} \text{ cm./V.*} \quad . \quad . \quad . \quad (274)$$

Note that the sensitivity to electrostatic deflection is increased if the final anode potential V on the electron gun is reduced.

From equation (274) for S it is obvious that the sensitivity is increased if the deflecting plates are close together, i.e. d small. To allow the plates to be separated by a short distance, and yet to prevent their exit edges from interfering with the electron beam when subjected to large deflecting voltages, the deflecting plates are frequently inclined to one another.

Magnetic Deflection. The use of a magnetic field extending over a short distance along the electron beam just beyond the gun may be employed to produce a deflection of the beam. Either one or both pairs of deflecting plates may therefore be replaced by deflecting coils as illustrated in fig. 104*b*. In a measuring oscillograph this has the advantage that a low impedance, current-sensitive deflecting means is obtained, whilst in a television receiving or transmitting cathode-ray tube, such magnetic deflection introduces less defocusing of the electron spot at the screen than does the use of an electrostatic field, especially when the spot is deflected to the edge of the fluorescent screen.

Referring to fig. 104*b* let H be the magnetic field strength over

* Note that the practical unit of P.D., the volt, can be introduced here without introducing a numerical factor depending on the ratio of practical to absolute units, since potential terms, E_d and V, occur in both the numerator and denominator of (273), so that the numerical factors cancel out.

the area indicated between A and B. Assume this field is uniform, and that in either direction beyond AB, H is zero.

According to equation (62), p. 46, an electron entering a field of strength H experiences a force Hev, where v is the electron velocity. The path OC of the electrons in the field will be the arc of a circle with centre at D. The velocity of the electrons will be constant, and equal to that which they receive as a result of the accelerating potential V on the final anode of the gun, i.e. $v=\sqrt{(2Ve/m)}$.

In fig. 104*b*
$$\angle\theta=l/R \text{ approx.} \qquad (275)$$

where R=radius of the arc, and l=distance along the axis over which the magnetic field extends.

From equation (64),
$$R=\frac{mv}{He} \text{ cm.} \qquad (276)$$

If L is the distance from the centre of the magnetic field to the fluorescent screen, then since $L \gg l$ in practice, so the deflection D of the electrons at the screen is

$$D=L\tan\theta=L\theta \text{ approx.} \qquad (277)$$

provided that θ is a small angle.

$$\therefore \quad D=L\theta=\frac{lL}{R}=\frac{lLeH}{mv}=\frac{lLH}{\sqrt{V}}\sqrt{\frac{e}{2m}}. \qquad (278)$$

The magnetic-deflection sensitivity Z is the deflection per unit field strength, i.e. D/H.

$$\therefore \quad Z=\frac{D}{H}=\frac{lL}{\sqrt{V}}\sqrt{\frac{e}{2m\,.\,10^9}} \text{ cm./oersted} \qquad (279)$$

on changing from absolute units and putting V in volts, and H in oersted.

Note that this sensitivity is now inversely proportional to $\sqrt{V}$, instead of V as in the electrostatic case. Hence, in employing electron guns, the use of magnetic deflection affords the advantage that large values of V do not reduce the deflection sensitivity so seriously as in the electrostatic case.

As compared with electrostatic deflection, the use of magnetic deflecting coils introduces two disadvantages: firstly, that the magnetic field cannot be caused to occupy so specifically a definite region of the electron path: it must exhibit more fringing at the edges of the field; secondly, the high field strengths required for appreciable deflections with large gun voltages, and the reduction of fringing, demand the use of iron-cored deflecting coils if the physical dimensions of the deflecting system are to be kept reasonably small. The presence of such iron may readily introduce distortion of wave-form, since the magnetic field produced by a coil with an iron core is not necessarily proportional to the coil current for wide variations of the current magnitude.

Time Bases, or Scanning Generators. If a cathode-ray tube is to operate so as to depict on its screen a graphical trace of voltage vs. time, or current vs. time when an alternating P.D. is applied across the Y-plates, or alternatively an alternating current is passed through the Y-deflecting coils, then the horizontal X-sweep which the electron beam undergoes must be so arranged that the distance traversed horizontally by the electrons at the screen must be directly proportional to time, so that, in effect, the horizontal axis is a time-axis. To achieve this it is obvious that D_x/t must be constant, where D_x is the X-deflection which occurs in time t. It follows that the beam must move across the screen in the horizontal direction with constant velocity. Moreover, if the oscillograph is to record wave-forms of widely varying frequency, then it must be possible to have control over the velocity of the X-sweep, so as to alter it at will over a wide range.

The achievement of such a time-base, or scanning generator demands the development of a circuit with an E.M.F. output in the case of electrostatic deflection (or current output for magnetic deflection), where such E.M.F. increases uniformly with time, and on reaching its peak value falls instantaneously to zero to repeat its uniform rate of increase. The wave-form of such an E.M.F. is obviously a saw-tooth, as indicated in fig. 105*a*.

The most elementary circuit for such a saw-tooth wave-form generator is shown in fig. 105*b*; it employs a neon tube.

A characteristic of a gas-filled discharge tube such as a neon tube is that its striking potential is greater than its extinction potential. Suppose that the neon tube used in circuit, fig. 105*b*, has a striking potential of 170 V., and an extinction potential of 130 V. Let the total D.C. supply voltage available be 300 V.

When such a circuit is switched on, the capacity formed by the condenser C with the small capacity between the neon tube electrodes across it, does not charge immediately; rather the P.D. across C increases with time according to the relationship equation (88), p. 59.

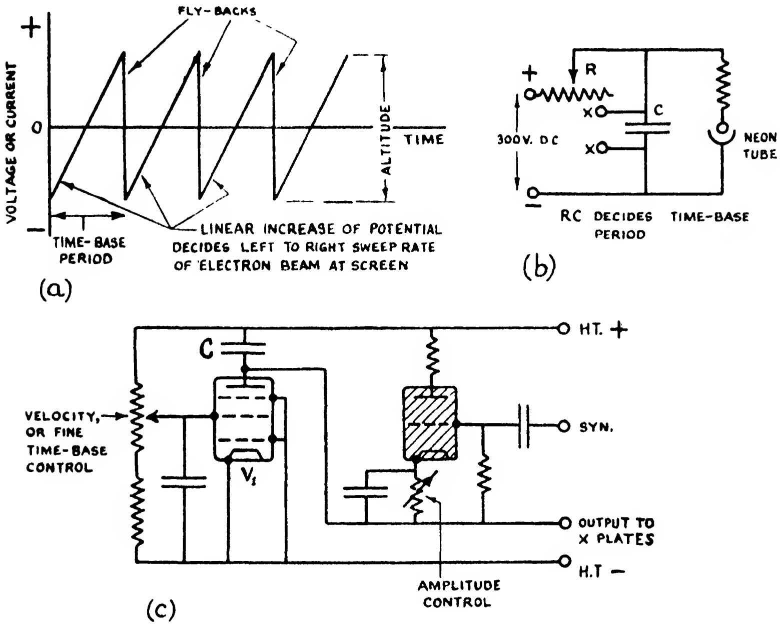


FIG. 105. *a*, Saw-tooth Wave-form of Time-Base Generator. *b*, Neon Time-Base Circuit. *c*, A Thyratron Time-Base.

$$v = V(1 - e^{-t/CR}). \quad . \quad . \quad . \quad (280)$$

As the P.D. across C grows exponentially at a rate decided by the time-constant CR, it will reach a value of 170 V. after a specific time interval. Then the neon will strike, and a low impedance gas path be presented across the condenser, allowing it to discharge at a much faster rate than it is charging, provided the resistance of the ionised gas is much lower than the series resistance R used in the circuit. This discharge will cause the P.D. across the condenser to fall rapidly to 130 V., when the neon tube will be extinguished, the ionised gas path will no longer be present and the charging action up to 170 V. will begin again.

If the X-plates of the cathode-ray tube are connected across the capacity C, they will be subjected to an exponentially increasing P.D. of which the rate of growth is decided by the product CR in the time-base circuit, whilst the rapid return to zero, or "fly-back" of the electron beam will take place when the condenser discharges via the ionised gas path. Such a circuit is therefore capable of giving an approximately saw-tooth wave-form E.M.F. output for time-base purposes of which the total sweep voltage is 40.

The disadvantages of such a circuit are:

(1) The amplitude of the sweep voltage cannot conveniently be varied.

(2) The minimum period of time occupied by the horizontal sweep, and so the maximum time-base fundamental frequency, is limited by the finite recombination time of the gas ions to a maximum frequency of some 10 kc./s.

(3) The rate of increase of P.D. across the condenser is not uniform but decided by the exponential growth of the condenser charge. This defect is not specially noticeable if the sweep amplitude is less than 10% of the neon striking potential.

To avoid the lack of linearity associated with the use of a condenser charged via a resistance, a pentode or beam tetrode valve is used in place of the resistance. On referring to the characteristic (fig. 59), it is seen that such a pentode operated with an anode voltage well in excess of the "knee" voltage, acts as a constant-current device in that the current through the valve is practically independent of the voltage across it. Moreover, the value of this current can be altered by variation of the screen potential. On charging a condenser from a source of D.C. via a pentode, the current passing through the valve is then constant, hence the rate of growth of charge, and so the P.D. across the condenser, is uniform.

To arrange that the amplitude of sweep of a time-base circuit can be varied at will, a thyratron circuit is used, fig. 105*c*. Since the discharge device is still an ionised gas path in this case, so the second disadvantage cited above still applies.

In the study of thyratrons considered in Chapter 5, it was shown that the anode potential at which the valve gas ionised was equal to the control ratio multiplied by the negative grid bias present.

Thus if K is the control ratio, and $-V_G$ the negative grid bias, then the thyratron anode potential must reach $K.V_G$ before the gas ionises, and anode current is obtained. Again, it was considered that once the ionisation of the gas had begun, then the grid became inoperative since it became surrounded by a sheath of positive ions, and that to extinguish the ionisation, and so cut off the anode current, it was necessary to reduce the anode to 10·4 V., the ionisation potential of mercury.

In the time-base circuit of fig. 105*c*, the condenser C will charge linearly via the pentode valve V_1, and at a rate decided by the screen potential of V_1. If the thyratron grid bias is initially set at $-V_G$, then the condenser potential will reach $K.V_G$ before the thyratron passes current to rapidly discharge C, giving the "flyback" stroke. This discharge action will continue until the thyratron anode potential is reduced to 10·4 V., then the anode current and the condenser discharge cease. The charging action then recommences as before. In this way a true saw-tooth waveform P.D. is set up across C, of which the frequency is decided by the capacity C and the pentode screen potential, and of which the amplitude is $(KV_G-10{\cdot}4)$ V. The amplitude can now be varied by a potentiometer adjusting the grid bias, $-V_G$.*

In view of the limitations to which all gas-filled tubes are subject in that they cannot be satisfactorily operated in less than 0·1 msec., the most generally used types of time-base nowadays employ hard valves. An ingenious circuit due to Puckle is shown in fig. 106.

In this circuit the condenser C charges via a constant-current pentode V_1, as before. The discharge of the condenser after a suitable period is achieved by the combined action of the valves V_2 and V_3. This time-base is capable of a maximum repetition frequency as great as 1 Mc./s.

The condenser C is connected between the cathode and anode of the triode valve V_2, the cathode being free to make a voltage excursion. Initially, on the first application of H.T. to the circuit, there is no P.D. across condenser C, so that the cathode of V_2 is at H.T.+ potential. The grid of V_2 is, however, at a positive

*** In Fig. 105*c*, note that the bias on the thyratron is cathode bias achieved by a condenser and resistance in the cathode circuit. The bias on the grid of the thyratron is due to the condenser charge obtained when the thyratron passes current. In order that the P.D. across this condenser maintains a negative grid during the operation, the time-constant of this *CR* combination must be large, compared with the maximum time-base period. See *Time Bases* by O. S. Puckle, Chapman and Hall Ltd., 1943.**

potential equal to H.T. + minus the voltage drop across R_2, since this grid is connected to V_3's anode. Hence the valve V_2 has an initially negative grid bias equal to $I_{A3}R_2$ with respect to its cathode, where I_{A3} is the initial anode current of valve V_3. This grid bias on V_2 is arranged to be more than the cut-off bias.

As condenser C charges up via the pentode V_1, the cathode of V_2 is made progressively more negative with respect to its anode, until a point is reached where the P.D. across C is sufficiently

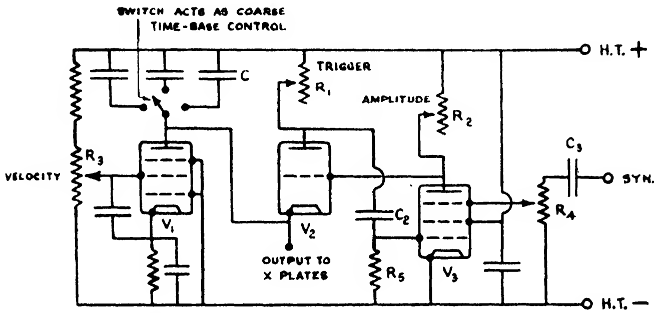


FIG. 106. Hard Valve Time-Base Circuit (Puckle).

large to ensure that the grid bias on V_2 is just less than its cut-off value, since V_2's cathode potential has been driven less positive to make it more nearly equal to V_2's grid potential. At this point a cumulative action takes place due to the connection of V_2 to V_3, whereby the anode current of V_2 suddenly becomes great enough to discharge C rapidly, giving the "fly-back" stroke. Then C commences to recharge, and the cycle is repeated.

That this cumulative action occurs, making the anode current I_{A2} much greater than it would otherwise be, can be considered as follows. When V_2 just begins to pass anode current, the P.D. developed across R_1, the anode load of V_2, passes a negative potential impulse on to the grid or suppressor of V_3 via the condenser C_2. As the grid of V_3 is made more negative, so the anode current of V_3 is reduced, and the P.D. across R_2 falls. But this P.D. across R_2 decides the negative grid bias on V_2, hence this bias is reduced, so that I_{A2} of V_2 grows still further, the action being cumulative, ultimately bringing the grid of V_2 to a positive potential with respect to its cathode, allowing a rapid discharge of the condenser C_1. In following out this argument it must be realised that the circuit currents are continually and rapidly changing, so that the coupling condenser C_2 is capable of depressing the grid potential of V_3 as a result of the decrease of the anode potential of V_2. Again, owing to the practically infinite speed of the electron velocities in the valves, the whole of this cumulative action occurs practically instantaneously, enabling "fly-back" times to be so short that repetition frequencies up to 1 Mc./s. are realisable. Greater frequencies than this are not easily obtained because of the presence of inter-electrode capacities in the circuit.

Synchronisation. If, for example, a 1000 c./s. sinusoidal wave-form is delineated on the screen of a cathode-ray tube, and the time-base is adjusted so that the electron beam sweeps from left to right of the tube in 1 msec., then a single sine-wave will be shown on the screen. For most purposes it is necessary that this wave-form be maintained stationary so that it can be examined, or photographed at leisure. This is not possible without some method of synchronising the time-base to the source applied to the Y-plates, since the unaided time-base circuit will usually be subjected to temperature, valve or voltage changes which cause its repetition frequency to vary slightly, though the values of C and R are nominally fixed. As a result an aggravating drift of the delineated trace to one side or the other of the fluorescent screen occurs.

Synchronisation is effected by injecting a portion of the E.M.F. output to be applied to the Y-plates into the time-base circuit. Thus a link is usually provided on an oscillograph to enable the Y-voltage to be coupled to the grid of the thyratron in a time-base, or via a condenser, to the grid of V_3 in the Puckle circuit. Considering the latter case more fully, the terminal marked SYN on the circuit (fig. 106) is electrically connected to one of the oscillograph Y-plates, whilst the other plate of the pair is earthed. By variation of the synchronising control potentiometer R_4, more or less of the total alternating E.M.F. to be examined is made to apply a corresponding P.D. to the grid of V_3. Since the instant of time at which the time-base condenser C discharges is determined by the bias on V_2 which in turn depends upon the current through R_2, and so on the bias on V_3, it follows that the synchronising voltage applied to the grid of V_3 can also affect the time at which the time-base fly-back begins. Suppose, for some reason, the time-base repetition period should be 0·001 sec., but tends to increase or decrease by 0·0001 sec. The application of a suitably adjusted synchronising P.D. to V_3 grid can then cause the grid bias of V_2 to drop or rise prematurely, as the case may be, so that the period of 0·001 sec. is rigidly maintained.

Circular Time-Base. The electron beam of a cathode-ray tube can be made to execute a circular trace at the screen by the application of sinusoidally varying E.M.F.'s to the two pairs of deflecting plates, where the alternation applied to the X-plates

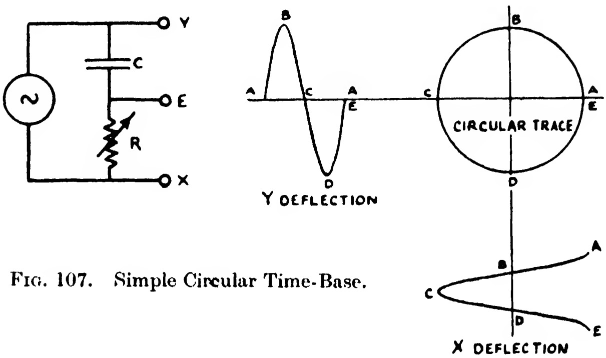


FIG. 107. Simple Circular Time-Base.

is equal in amplitude, but 90° out of phase with the E.M.F. applied to the Y-plates.

This is most easily achieved by applying A.C. of sine-wave form across a condenser and resistance in series, where $R=1/\omega C$, R being the non-inductive resistance value in ohms, C being the condenser capacity and ω the pulsatance. By such a device, two alternating P.D.'s can be applied to the plates differing in phase by 90°, the action being illustrated in fig. 107.

If the amplitudes of the applied P.D.'s are not equal ($R \neq 1/\omega C$) then an elliptical trace is produced. If the phase difference is not 90° (R possesses inductance) then the elliptical trace has an axis at an angle to the horizontal.

Applications of the Cathode-Ray Tube. The uses of this versatile piece of apparatus are far too numerous to consider here in detail: in fact the number of uses to which the oscilloscope can be put is chiefly limited by the lack of ingenuity or knowledge of the electronic engineer who has an instrument available. Some of the chief applications of which details can be found in the literature (see Bibliography) are worth mentioning here to complete an introductory account of this most useful apparatus.

(1) Measurement of D.C. and A.C. voltages, using a tube of known electrostatic deflection sensitivity, provided with a screen calibrated both horizontally and vertically in mm.

(2) The measurement of phase difference, either by using an elliptical trace procedure, or by employing a double-beam tube. Such a tube has a plate placed just outside the final anode of the gun. This plate is edge-on to the beam, with its plane containing the tube axis, and extends between the Y-plates, producing two deflecting systems, so that each Y-plate affects one half of the beam separately. A common X-plate time-base system is provided. The most widely used example is the Cossor double-beam tube, possessing terminals Y_1 and Y_2 for the application of two vertical deflecting potentials.

(3) The measurement of frequency, either by Lissajou's figures, or by using a calibrated time-base, or by comparison of an unknown with a standard source of A.C.

(4) The examination of wave-forms, leading to Fourier analysis for evaluating percentages of the various harmonics present.

(5) Using a frequency-modulated oscillator for producing the X-deflection (see p. 202), the resonance curves of tuned circuits can be reproduced.

(6) Measurement of modulation depth and distortion.

(7) Examination of frequency-response characteristics of distortion and noise arising in amplifier systems in general.

(8) The delineation of valve characteristics.

(9) The measurement of resistance, inductance, capacitance and power factor.

(10) The production of B-H curves in examination of ferromagnetic specimens.

(11) The investigation by the echo method of the ionosphere, where a calibrated time-base is used as a method of measuring very short time intervals. This leads to the innumerable applications of the cathode-ray tube as a radar navigational device in general.

(12) As a television receiver tube.

(13) In electro-cardiography and electro-encephalography in medicine.

(14) Indication of pressure variations in heat engines.

(15) In all forms of work on acoustics where the wave-forms of sounds produce via a microphone an equivalent electrical current.

(16) Examination of the characteristics of gas discharges in general.

CHAPTER 12

The Electron Microscope

The Resolution Limit of the Microscope. Some seventy years ago, after the light microscope had been known and developed for more than two hundred years, Abbe showed that the limitation to the resolving power, i.e. the ability to distinguish detail in the field of view, was set by the finite wave-length of the light employed. The smallest distance d between two points in an object which can be viewed as distinctly separated in the microscope field is given by the equation,

$$d = \frac{\lambda}{2\mu \sin \alpha}, \qquad . \quad . \quad . \quad (281)$$

where λ = wave-length of light employed, μ = refractive index of object space, α = semi-vertical angle of cone of rays entering objective.

In the light microscope the design of objectives has enabled values of $\sin \alpha$ as great as 0·9 to be obtained, implying that light is gathered from an object point over an angle as large as $2 \sin^{-1} 0{\cdot}9 = 128°$, and by immersing the object in an oil of the same refractive index as the objective glass, μ may be uniformly 1·6. In such an instrument, equation (281) will reduce to

$$d = \frac{\lambda}{3} \text{ approx.} \qquad . \quad . \quad . \quad (282)$$

It follows that, using the best modern optical technique, detail finer than one-third of a wave-length of the illuminating light employed cannot be achieved. Using white light considered as of mean wave-length 5000 Å., this sets the resolution limit at 1700 Å. $= 1{\cdot}7 \times 10^{-5}$ cm. A linear magnification of more than 500 brings about no increased advantage, therefore, since the unaided eye can perceive detail as fine as 10^{-2} cm. Even using an ultraviolet light microscope, with the attendant disadvantage that viewing must be via a fluorescent screen, or by photographic recording, yet $\lambda = 2000$ Å., and cannot be much less because of the limited transmission of the quartz lenses* to lower wave-lengths. Then the resolution limit is some 700 Å.

In many fields of work, especially biology, objects of interest such as bacteria are of finer detail than 700 Å. The provision of a microscope of higher resolution would thus be of outstanding advantage.

In 1924, de Broglie advanced the hypothesis of the wave-nature of the electron: an electron of mass m and velocity v has an associated periodicity of frequency ν and wave-length $\lambda = c/\nu$,

where
$$\frac{h}{\lambda} = mv, \qquad (283)$$

and $h = 6{\cdot}55 \times 10^{-27}$ erg sec., is Planck's constant.

By substitution for the constants in this equation, and realising that $v \propto \sqrt{V}$, where V is the electron-voltage, this equation leads to

$$\lambda = \frac{12{\cdot}25}{\sqrt{V}}\ \text{Å}. \qquad (284)$$

and at $V = 150$ V., $\lambda = 1$ Å. approx.

By using electrons accelerated by potentials of the order of 50 kV., a much shorter wave-length can be obtained, of the order of 100,000 times smaller than for visible light. Equation (284) needs a relativistic correction for the electron mass at these enhanced speeds, giving

$$\lambda = \frac{12{\cdot}25}{\sqrt{V}} \frac{1}{\sqrt{(1 + 9{\cdot}8 \times 10^{-7} V)}}. \qquad (285)$$

An electron microscope of very great resolving power then suggests itself as a practical possibility, particularly in view of the fact that, as is shown in Chapter 10, electron lenses of the magnetic or electrostatic types are available. In practice the extraordinary resolution limits suggested by considering the very short electron wave-length at 50 kV. are not achieved. This is because it is no longer possible for the angle α in equation (281) to be anything like so large as in the ordinary optical case. The modern electron lens with its spherical aberration, coma and chromatism is less perfect than a moulded glass simple box camera lens. Its aperture

* It is worth noting that the reflecting mirror microscope, particularly owing to the development work of C. R. Burch, will probably lead to the use of U.V. light of wave-length less than 2000 Å.

must be therefore much limited if it is to give good focus. In electron microscopes the numerical apertures (μ sin α) used are about one-thousandth of those used in light microscopes.* The possible improvement of 100,000 times in resolution due to shorter wave-length is thus offset by a factor of 1000 due to limited numerical aperture, so that electron microscopes are only capable of some 100 times better resolution than light microscopes, allowing a total magnification of 50,000 to be realised. The best results achieved so far have a resolution limit at 10 Å., much greater than should be theoretically possible, but nevertheless well worth while because it brings within the miscroscopist's field of view the whole range of minute organisms with detail between 10 and 1000 Å. which were formerly hidden from his inspection.

An Early Electron Microscope. In 1932 Knoll and Ruska† produced the first practicable electron microscope taking advantage of de Broglie's hypothesis with regard to electron waves, combined with Busch's treatment of the magnetic focusing coil. The instrument they made had little greater magnification than a good pocket magnifier, yet it paved the way for future developments in this field, and in basic outline was similar to its successors. It is a noteworthy tribute to the rate of progress nowadays that, within four years, the electron microscope was made as good as the best light microscope, whilst by 1940 the light-microscope resolution limit had been exceeded by 100 times.

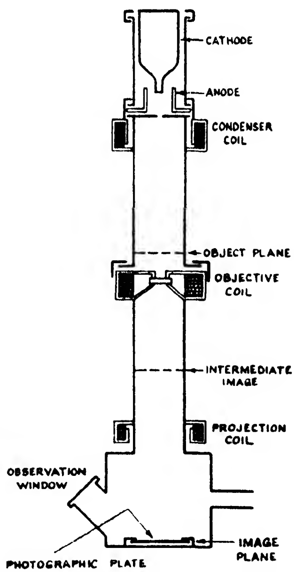


FIG. 108. Early Electron Microscope (Knoll and Ruska).

The lens arrangement adopted by Knoll and Ruska compares with that of the usual light microscope. The specimen to be examined is in the form of a film sufficiently thin to be transparent to fast-moving electrons. The electrons are furnished by a cold-discharge method using an anode at 50 kV. This is equivalent to the source of

* Values employed range from 0·6 to 6×10^{-1} degrees, or 10^{-3} to 10^{-2} radians.
† M. Knoll and E. Ruska, *Ann. der Physik*, **12**, 607, 1932.

illumination. Between the source and the object to be viewed is situated a condenser coil, comparable with the usual sub-stage light condenser. The electrons which traverse the object are focused to form an intermediate image by the objective coil, and finally the intermediate image acts as the object for a projection coil, comparable with an eye-piece, to give an electron-optical image at the fluorescent screen, or photographic plate where the electron impact is rendered visible by fluorescent action, or by subsequent development of the photographic emulsion.

The whole system has to be maintained at an adequately low pressure by means of a mechanical backing pump and diffusion pump combination.

Two Types of Object used in Electron-microscopy. In general there are two types of electron-microscope objects which can be viewed:

(*a*) A cathode emitting electrons, i.e. a self-luminous object, corresponding with the examination of an actual source of light in ordinary microscopy. Such cathodes are usually thermionic emitters, but may be photo-cathodes, radio-active sources or secondary emitters. Electron microscopy employing such objects is necessarily limited to the examination of actual sources of electrons, e.g. used in the study of cathode activation.

(*b*) The second, much more important type of object is in the form of a thin film illuminated by fast electrons from a separate cathode. In light microscopy the object either reflects light to the objective from the source, or light passes through the object. In electron microscopy the use of electrons from the source which are reflected by the object is practically prohibited because of their varying velocities, and admixture with secondary electrons. Most electron microscopes employ a separate thermionic cathode which is so placed that electrons from it pass through the sufficiently thin object. The fact that the object has to be in the form of a thin film* mounted in vacuum is a limitation to the applicability of electron microscopes in the examination of specimens.

The Formation of the Image. It is not obvious that an electron microscope should be capable of forming an image at all. The image in light microscopy is apparent because of variations

* At an electron voltage = 30 to 60 kV., specimen must be between 1000 and 5000 Å. thickness, i.e. between 10^{-5} and 5×10^{-5} cm.

in light absorption, refraction or colour over the area of the object being examined. The object specimens used in electron microscopy are, however, so thin that they cannot produce sufficient differential absorption through the cross-section of a high velocity irradiating beam to give the contrast in the image that is actually obtained. The object, in fact, produces an image by scattering of the incident electrons.

When fast electrons collide with an atom in the object either an elastic or inelastic collision occurs. Elastic collision will mean that the electron scatters from the atom without loss of energy; this is rare. It is more usual for inelastic collisions to occur whereby the electron loses energy in discrete steps of about 20 V. at a time.

On impinging with the object, electrons will be scattered at all angles to the object surface and in a manner depending on the object structure. If the electron lenses used have very small limiting apertures, as is common to cut down aberration, those electrons which are scattered outside this aperture will be missing from the final beam at the fluorescent screen, and so produce contrast, giving image detail. According to Gabor,* this aperture is not, however, necessary to the argument since the electron lens will act on scattered electrons of varying velocity as a lens with chromatic aberration, and will focus such deflected electrons in a different plane from that of the image, so that they will eventually reach the fluorescent screen or photographic plate over a greater area than that of the point in the object from which they arose, and simply give low-level background illumination of the screen or plate. In fact Gabor asserts that chromatic aberration arising within the objective lens with respect to scattered electrons is an essential condition for the working of an electron microscope, whilst demonstrating that such chromatism must be clearly separate from the harmful chromatic effect due to variation of electron velocities brought about by accelerating voltage fluctuations. This argument is certainly borne out by the fact that a limiting aperture is not essential in the production of contrast in the image.

Electron Lenses Employed. Since Knoll and Ruska's pioneer work the development of the electron microscope has been along two main lines:

(*a*) using magnetic electron lenses;
(*b*) using electrostatic lenses.

* D. Gabor, *The Electron Microscope*, Hulton Press, 1944.

At present electron microscopes of both types are commercially available, but the former type (*a*) has been more generally successful.

To obtain the high magnification necessary to take advantage of the high resolving power of these microscopes, it is necessary that the objective lens should be of as short a focal length as possible. Even so, to obtain the highest magnifications possible, the microscope needs to stand some six to seven feet high.

Coils used are clad with soft iron, and operate at about 2000 to 3000 ampere-turns. The iron terminates in two pole-pieces leaving a small gap within which the focusing field is concentrated. Usually the winding is completely enclosed by an air-tight iron shield so that the coil can be within the microscope vacuum chamber. In order that the object be sufficiently close to the field between the pole-pieces, it is supported within the coil, as shown in fig. 109*a*.

In accordance with equation (264), p. 254, the focal length f of a magnetic focusing coil is given by

$$\frac{1}{f} = k.\frac{I^2}{V}, \qquad (286)$$

where I is the coil current deciding the magnitude of the field H_z, and V the accelerating voltage, k being a constant depending on the coil size, shape and iron disposition.

Variation of the focal length f will be produced by changes of I or V. It is therefore necessary, to avoid chromatic* aberration, and focal length changes, for I and V to be kept particularly constant. Modern electronic practice has produced stabilised circuits capable of maintaining a current constant to within 0·002%, whereas voltage can be maintained constant to 0·001%. This is sufficient for modern needs in view of the other inherent defects of electron microscopes.

In 1932 Brüche and Johannson† introduced the first electron microscope employing electrostatic lenses. The outstanding advantage is that the univoltage type of lens (p. 248) can be employed giving freedom from chromatic aberration due to supply

* From de Broglie's formula $h/\lambda = mv$, the electron wave-length depends on the electron velocity, and so accelerating potential. Variations of the latter therefore cause wave-length changes. The comparable aberration in light optics is colour or chromatic aberration.

† E. Brüche and H. Johannson, *Naturwissenschaft*, **20**, 353, 1933.

voltage fluctuations so that elaborate circuit stabilising is avoided. Moreover, to achieve the short focal length lens necessary to produce great magnification in a single stage, the unfortunate saturation of the iron of a magnetic focusing coil with increased

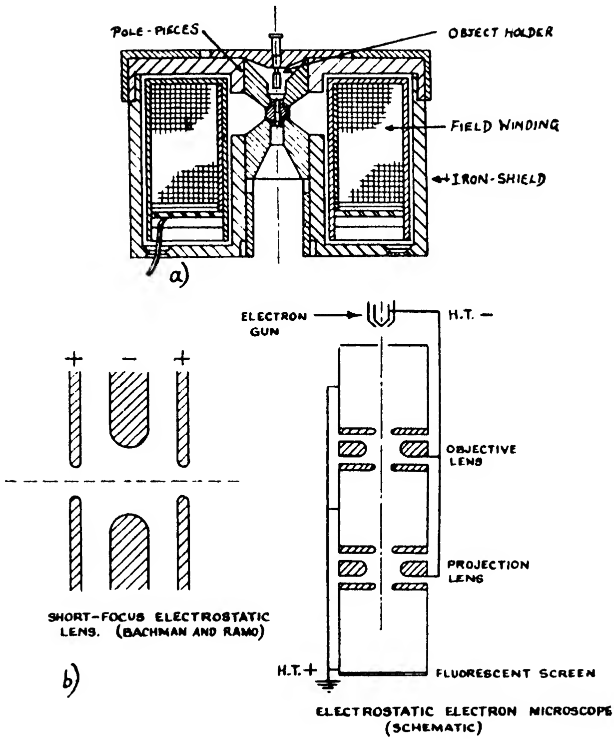


FIG. 109. *a*, A typical Magnetic Objective Lens. *b*, A typical Electrostatic Lens.

coil current is avoided. Unfortunately, however, the production of very short focal length electron lenses of the electrostatic type is limited by the high potential gradients necessary, giving rise to field emission from the electrodes, insulation difficulties, arcing and brush discharge.

Requirements demanded of a suitable electrostatic lens are:

(*a*) The specimen must be situated outside the field of the objective lens.

(*b*) The mechanical arrangement of electrodes must be considered to minimise the insulation problem, and sharp points and edges avoided to prevent field emission.

(*c*) To minimise chromatic aberration defects an excessive drop in axial potential in the central region of the lens must be avoided.

(*d*) Considerable freedom from spherical aberration with reasonable numerical aperture.

For an electron lens, magnetic or electrostatic, the longitudinal spherical aberration increases with α^2, whilst the tranverse type increases with α^3, where α is the semi-vertical angle of the cone of rays at the objective. The only way in which this aberration can be sufficiently reduced is by making α very small. Hillier and Vance* quote the formula

$$D=4f\times 10^{-3}. \qquad (287)$$

where D is the diameter of the aperture, and f the focal length of the electron lens. At such apertures the magnitude of the lens aberration is made equal to the diffraction defect. Such an equality is desirable since diffraction defects increase with diminution of aperture diameter.

If f is 5 mm., then D in accordance with the above formula will be $4\times5\times10^{-3}=2\times10^{-2}$ mm. The smallest apertures used in practice are $2{\cdot}5\times10^{-2}$ mm.

The Brightness of the Image. Since small apertures are demanded of electron lenses in order that they be adequately free of spherical aberration, it would seem that the prospect of obtaining a sufficiently bright image was very low. However, on realising that the electrons emitted from the cathode with thermal velocities of about 0·2 eV. are finally accelerated to some 50 keV., it is seen that this limitation is not serious. The ratio of image to object light intensities is as the ratio $(\mu/\mu_0)^2$, where μ is the refractive index in the image space, whilst μ_0 is the index in the object space. This cannot exceed three to four times in light optics, but in electron optics, where enormous effective refractive indices are obtainable, this ratio of intensities can be more than 100,000 times.

Modern Electron Microscopes. Three of the most noteworthy electron microscopes available are:

(*a*) The R.C.A. Magnetic Type (American).

* Hillier and Vance, *Proc. Inst. Rad. Eng.*, **29**, 167, 1941.

(*b*) The Metropolitan-Vickers Magnetic Type (British).
(*c*) The G.E.C. Electrostatic Type (American).

(*a*) *The R.C.A. Magnetic Type Electron Microscope.* First placed on the market in 1941, and developed by Zworykin, Hillier and Vance, this instrument operates at 60 kV., and is capable of a magnification of 20,000 times, so that detail in the object of 30 Å. linear dimension will be rendered at 0·06 mm. in the final image. This is readily photographed by a plate with a resolution of 150 lines/mm., i.e. grain size about one-tenth of the recorded detail.

This instrument is nearly seven feet high. At the back of the microscope proper is mounted the electronic gear for supplying the 60 kV. H.T. and the coil currents, whilst at the base of the assembly there is a 3-stage oil-diffusion pump and backing pump for producing a vacuum of less than 10^{-5} mm. Hg.

The microscope itself is built up in a series of chambers, each accurately machined and fitted together, on top of one another, by rubber gaskets between flanges. The uppermost chamber is the discharge tube containing an electron gun fitted with a V-shaped directly heated tungsten filament, cathode shield and corona shield, the whole being maintained at -60 kV. Below this appears a flexible union, provided with two sets of adjusting screws so that alignment with the following condenser lens chamber can be accurately achieved. A second flexible joint, with screw adjustment, then separates the condenser lens chamber from the objective lens, allowing alignment of the electron beam to be obtained by viewing its impact on a perforated fluorescent screen set just above the projector lens. Beneath the objective lens there is a flexible joint connecting this lens to the cylindrical chamber above the projection lens. The chamber in which the electron beam travels after passing the projection lens is finally followed by the photographic chamber. The final image can be observed on a fluorescent screen mounted just over the photographic plate, exposure being effected by tilting this screen out of the electron path by an external control operating via a conical joint.

Air locks for the specimen and for the photographic plate are so arranged that the changing of either of these can be done without the necessity of bringing the whole of the interior of the microscope up to atmospheric pressure. This enables re-evacuation, after changing the specimen or the plate, to be performed in less than three minutes.

The specimen to be examined is mounted on a thin film of collodion, prepared by flowing collodion solution over a water surface. A small disk of this collodion, picked up by a fine wire frame from the water surface, is dried, a drop of the suspension to be studied is placed on the collodion and the arrangement placed over the opening of a small cartridge holder which fits snugly into a small thimble situated about 5 mm. above the centre of the objective lens. Many other techniques for specimen mounting, decided by the nature of the specimen, have been employed.

The object stage within the objective chamber can be moved in two directions at right angles to one another in the horizontal plane, permitting any desired portion of the object to be viewed.

A 25×5 cm. photographic plate is used, suitable for obtaining 5×5 cm. negatives side by side.

The H.T. and coil current supplies have to be particularly stable to avoid chromatic aberration. A voltage variation of less than 0·004% during the exposure period (a maximum value is 30 sec.) is achieved by a circuit due to Vance in which the A.C. supply to the rectifier has a frequency of 32 kc./s., generated by a valve oscillator. The use of a supply at such a frequency makes the problem of smoothing the rectified output much simpler, because the high working voltage condensers are required to have a capacity 1/640 of that necessary at 50 c./s. to achieve the same freedom from ripple, so that the physical dimensions of the condenser are reasonable (see p. 101). Moreover, it is much easier to shield the microscope efficiently from variations at 32 kc./s. than it is at 50 c./s., since magnetic shielding at the higher frequency is efficiently arranged without using thick and heavy ferromagnetic sheets of material. A third point in favour of the 32 kc./s. supply is the faster action of the voltage regulator necessary to counteract the effect of long period line voltage fluctuations. The overall size of the apparatus for producing 60 kV. is smaller if 32 kc./s. is employed than if 50 c./s. mains is used, despite the necessity for providing an auxiliary oscillator in the power pack.

The high voltage output is regulated by a negative feed-back circuit. A potential divider appears across the H.T. supply. This divider passes a load current at least ten times greater than that required to operate the electron microscope (0·5 mA.) so that the load presented to the power pack is practically constant. The smaller part of the P.D. across the divider is balanced against a dry battery, which has a practically constant E.M.F. Any fluctuation of the total H.T. supply then appears as a voltage difference relative to the battery E.M.F. This voltage difference is amplified by a D.C. amplifier, and used to control the screen grid voltage of the output power amplifier fed by the 32 kc./s. oscillator. The polarity of this screen potential is arranged so as to reduce the 32 kc./s. input to the rectifier if the total H.T. supply tends to rise, and vice versa.

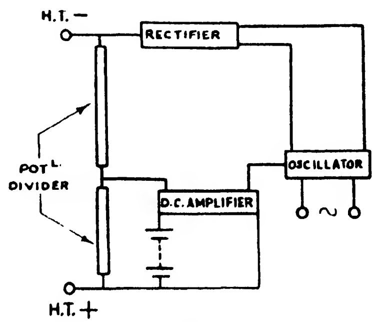


Fig. 110. Outline diagram of H.T. Supply for R.C.A. Magnetic Electron Microscope.

To provide a smaller and cheaper instrument capable of a resolution of about 100 Å., and so enable the adequate examination of a wide range of interesting specimens with detail ten to twenty times too small to be viewed by a light microscope, the R.C.A. Co. have also marketed a desk model microscope. This is considerably simpler than the model just described. It dispenses with a condenser lens by employing electrostatic focusing as the sub-stage arrangement, provided by impressing a negative potential on the cathode shield. A novel feature of this instrument is that a single coil is used to energise both the objective and projection lenses which form one unit, provided at each end with pole-pieces. A fixed magnification of ×5000 is obtainable, reducible to ×500 by removing a cartridge from the pole-piece forming the projection lens. The supply to the energising coil is variable over only a small range to give a fine focus adjustment. The fluorescent screen is viewed by transmission, as in an ordinary cathode-ray tube. This instrument operates at 30 kV.

(*b*) *The Metropolitan-Vickers Magnetic Type.** The type EM 2 electron microscope which operates at 50 kV. has been produced

* By courtesy of the Metropolitan-Vickers Electrical Co. Ltd., England.

in England, chiefly as a result of the work of Haine of the Metropolitan-Vickers Co. It has a specified resolving power of better than 100 Å., with direct magnifications in excess of $\times$ 10,000.

This instrument, illustrated in fig. 111, is somewhat similar to the R.C.A. type in broad outline, but differs markedly in constructional features, and is generally alleged to afford easier operation. Three main components form the microscope complete, the main tube mounted on a cast brass pedestal, the electrical control cubicle, and a separate oil-immersed 50 kV. D.C. set with rectifiers operating with an input at 50 c./s. mains frequency. The necessary rotary vacuum pump is mounted apart from the main tube to avoid vibration interference.

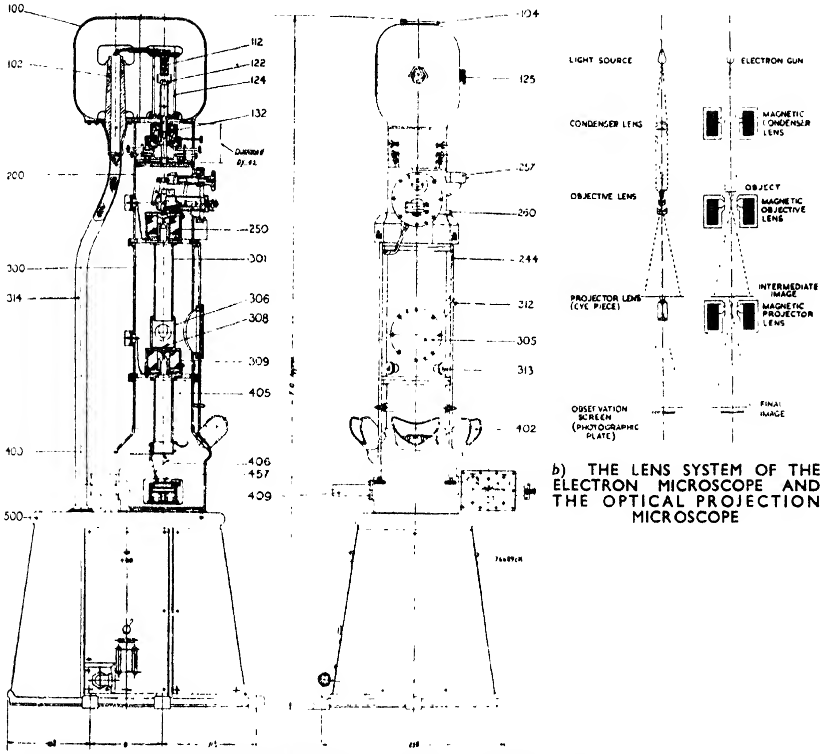
The specimen under examination and the photographic plate can be readily inserted into and removed from the vacuum without the necessity of breaking the vacuum to the whole microscope chamber. Such demountability is arranged by the use of an all-metal construction with rubber cord vacuum seals between flanges, or, where it is necessary to transmit a drive within the vacuum, by rubber-packed glands, or flexible metallic bellows.

As in the R.C.A. type, the electron gun and condenser lens positions can be adjusted laterally by means of two sets of knurl-headed screws. Tilting of the gun and condenser assembly can also be achieved to regulate the direction of the electron beam.

The specimen is normally mounted on a plastic film some 10^{-6} cm. thick supported on an $\frac{1}{8}$-in. diameter 200-mesh copper grid. This grid is carried in a cartridge which fits snugly into the mechanical stage which is clamped on top of the objective lens coil. This stage is provided with screw adjustments for positioning, operated from outside the vacuum. The specimen is inserted into, and removed from, the main tube through an air lock.

A magnification of $\times 60$ to $\times 100$ is furnished by the objective lens, and forms an intermediate image which can be viewed on a fluorescent screen through an eye-piece tube set obliquely into the main instrument tube. The central portion of the electron beam passes through a small hole in the intermediate screen, and is then focused by the projector lens ($\times 150$ to $\times 200$) on to a 3-in. square final fluorescent screen. This screen is swung out of the path of the electrons when desired.

The camera uses 10×2 in. photographic plates enabling five 2×2 in. exposures to be made.



b) THE LENS SYSTEM OF THE ELECTRON MICROSCOPE AND THE OPTICAL PROJECTION MICROSCOPE

a) CROSS SECTIONAL DIAGRAM OF MAIN ELECTRON MICROSCOPE ASSEMBLY

Fig. 111. The Metropolitan-Vickers EM 2 Electron Microscope.

KEY TO ITEM NUMBERS

100—Gun Protective Cover.
102—H.T. Bushing.
104—Earthing Bar.
112—Cathode Assembly.
122—Anode Assembly.
124—Glass Cylinder.
125—Gun Cover Latch.
132—Condenser Lens.
200—Object Chamber
244—Mechanical Stage Driving Shaft.
250—Objective Lens
257—Object Chamber Illuminator.
260—Viewing Port.
300—Intermediate Tube
301—Mumetal Screening Tube.
305—Access Port.
306—Mumetal Screening Tube
308—Intermediate Screen
309—Projector Lens.
312—Intermediate View Port.
313—Projection Lens Adjuster.
314—Tube for H.T. Cable.
400—Viewing Chamber.
402—Binocular Viewing Window.
405—Mumetal Screening Tube.
406—Masking Device.
409—Camera Sealing Box.
457—Final Viewing Screen.
500—Pedestal.

Recently (1947) an improved type of Metropolitan-Vickers electron microscope, the EM 3, has appeared. This makes use of a second intermediate projector lens enabling the microscope tube

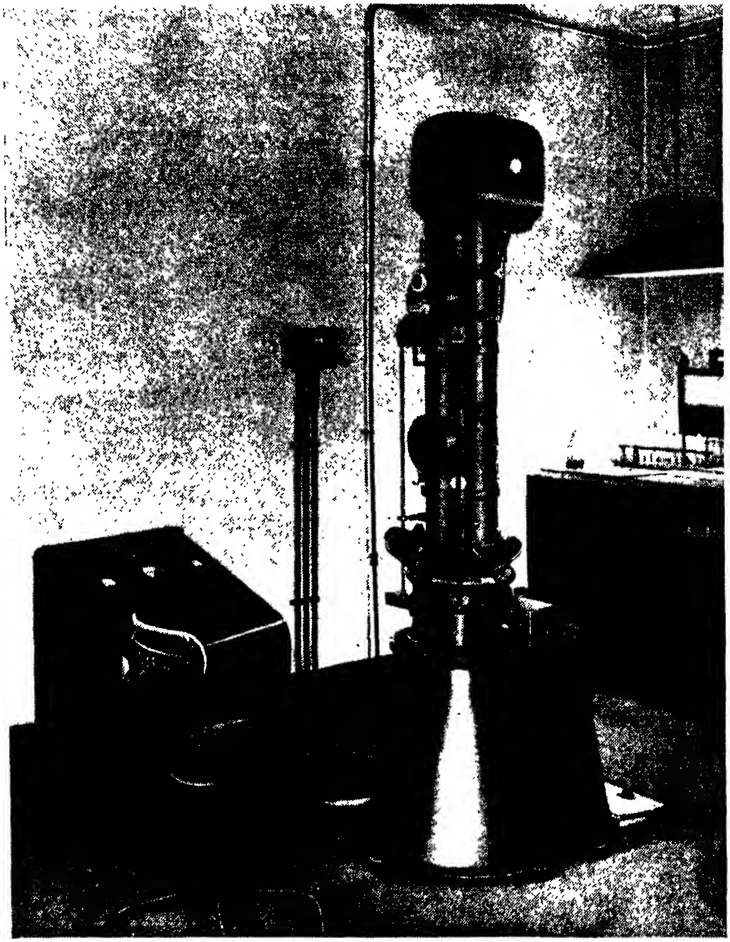


FIG. 112. Metropolitan-Vickers Electron Microscope, Type EM 2.

to be reduced in length, and permits wide variations of magnification without alteration of the focal length of the objective lens.

Since the microscope is smaller, evacuation is much quicker. So the use of air locks for specimen and photographic plate insertion is avoided, since the production of the necessary vacuum in the whole system only takes $2\frac{1}{2}$ minutes. The microscope is therefore more trouble-free as regards air leaks, and easier to align and control.

An electron gun giving much higher beam currents than that

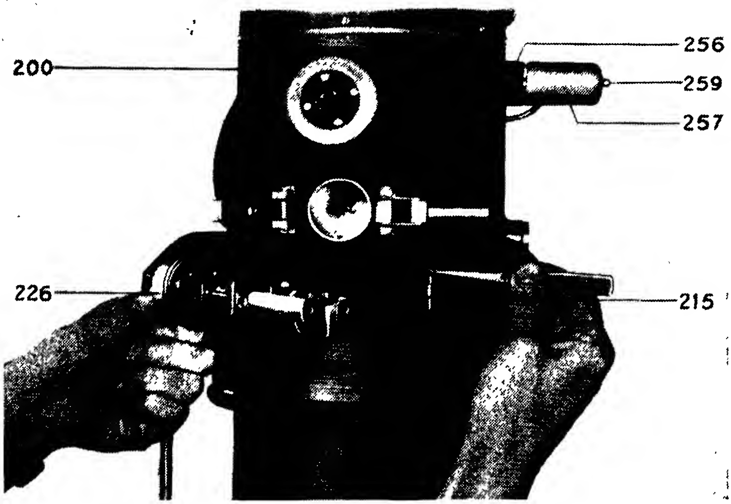


FIG. 113. Removal of Specimen Holder (EM 2 Microscope).

KEY TO ITEM NUMBERS

200—Object Chamber.
215—Specimen Holder.
226—Insertion Block.
256, 257, 259—Object Chamber Illuminator.

used in the EM 2 enables electronic magnification up to $\times 100{,}000$ to be realised, with a sufficiently bright image.

The magnetic coil windings are outside the vacuum, permitting efficient cooling, with large coil currents. The necessary short focal lengths have been achieved by redesign of the pole piece shapes.

Other innovations in this new instrument are:

(i) A new specimen stage providing greater mechanical stability, incorporating a stereoscopic device which enables the specimen to be tilted whilst under observation.

(ii) A physical aperture can be fitted to the fixed focus objective lens so that the image contrast can be improved. Since the objective lens has fixed focus it is possible greatly to improve the stabilisation of its coil current.

(iii) As in the R.C.A. instrument, the high voltage oil-immersed constant potential unit is fed from a high-frequency source to enable its physical size to be reduced. Stabilisation to 1/50,000 is achieved by means of a feed-back amplifier. The electron gun filament is also fed by low voltage high-frequency A.C.

(iv) The magnification can be varied from $\times 1000$ to $\times 100{,}000$ without changing the specimen holder. The instrument will operate at up to 100 kV.

(v) The microscope is arranged in the form of a desk, with easy access to all controls.

(*c*) *The G.E.C. Electrostatic Electron Microscope.* Developed in America by Bachman and Ramo,* this instrument, like all electron microscopes with electrostatic focusing, employs univoltage lenses (see p. 280). This simplifies greatly the problem of high voltage supply since fluctuations of voltage do not produce any change of the focal length of such lenses, except for the chromatic aberration due to the relativistic mass correction necessarily applied to electrons moving with high velocity, which effect gives negligible aberration at the present standard of development of electron optics. The focal lengths of the lenses are a function only of their electrode geometries.

The electron gun used to provide a beam to illuminate the object is similar to that used in the magnetic type microscope. This gun is adjusted to give a collimated illuminating beam. Mounted between the gun and the specimen there is a molybdenum disk provided with an aperture which subtends a solid angle of 10^{-5} steradians at the sample.

Three lenses are used, one objective and a double projection lens, all of the univoltage type, each with their two outer electrodes connected to the earthed casing, and their central electrodes at cathode potential, which is 30 kV. negative with respect to earth. Since these lens have fixed focus it is necessary to focus by movement of the object along the instrument axis.

A total magnification of $\times 1000$ is obtained electronically. This is improved in practice because the final image at the

* C. H. Bachman and S. Ramo, *Jour. of Appl. Phys.*, **14**, 155, 1943.

fluorescent screen is photographed through an optical microscope. This technique necessitates a fluorescent screen with a resolving power of 100 lines/mm., and an exposure time of a few minutes compared with the thirty seconds maximum necessary for direct photographic recording. The necessary long duration electron irradiation of the specimen may cause harm to its structure.

The microscope envelope is an accurately machined tube made from alternate layers of magnetic and non-magnetic material to provide efficient electric and magnetic shielding against stray fields.

The electron-lens system and object stage form a single unit, removed from the microscope when the specimen is changed. The specimen can be moved in three directions, one along the instrument axis for focusing, and in two directions in the plane perpendicular to the axis, for positioning.

Such an instrument is valuable when a magnification about ten times that obtainable with the best optical methods is required.

The difficulties associated with electron microscopy can be tabulated as follows:

(1) The usual lens aberrations, including space-charge troubles.

(2) In obtaining a picture of a structure which the electrons penetrate, a large proportion of the moving electrons may not pass sufficiently close to any of the electrons or nuclei in atoms to suffer perceptible changes in energy or direction.

(3) Electrons may pass so close to a massive nucleus as to suffer large changes in direction, but comparatively small energy changes.

(4) Chromatic aberration defects may arise due to a velocity distribution imparted to the electrons when penetrating the film, as distinct from those necessary to show the specimen structure.

(5) Effects due to stray electric and magnetic fields, both due to the earth, and man-made. These are nowadays adequately avoided by efficient screening.

(6) The structures examined may be changed, or destroyed by electron bombardment heating, or by intense electric charges produced on them.

(7) At the pressures obtainable in a demountable vacuum system, a semi-conducting, carbonaceous material is deposited on all the diaphragms in the projection system, giving a polarisation

layer effect which modifies the electric fields produced. These diaphragms must therefore be periodically removed and cleaned.

(8) Fluctuations of voltages applied to accelerating electrodes, or of magnetic coil currents.

Additional Information. Recently, the ever-increasing application of the electron microscope has led to improvements, not of any fundamental character, but chiefly to afford the greatest facility and flexibility in operation. The R.C.A. universal model,* p. 283, type EMU, now has all controls within easy reach of the operator during specimen examination. The microscope column, oil diffusion pump and power supply are built into the main unit, while the voltage regulator, and mechanical first stage pump, liable to cause vibration, are separate, and may be housed in a different room. Magnifications from $\times 100$ to $\times 20{,}000$ are obtainable, whilst the most frequently used range, 7500 to 20,000 is covered in ten fixed steps. Three lower ranges $\times 100$ to $\times 800$, $\times 800$ to $\times 3200$ and $\times 3700$ to $\times 11{,}000$ are obtainable by lens adjustment. The resolution is 100 Å. with facility, resolution as good as 30 Å. being obtainable. The depth of focus is 10 to 25×10^{-3} mm. A stereoscopic photographic attachment is available, and a built-in electron diffraction unit incorporated.

The desk, or console model, type EMC, see above, p. 285, is a smaller instrument which dispenses entirely with water cooling for the electrical equipment and pumps, the complete microscope being in one unit, and easily moved from one laboratory to another. Ready adjustment of the controls on the desk panel is made whilst the image is viewed at eye level by a seated operator. The resolution is comparable with the universal model, whilst the depth of focus is 5×10^{-3} mm.

Electron microscopes are now also available from Philips' Lamps Ltd., who manufacture a magnetic type instrument, type 11980, stated to have a resolution of 25 Å. Two extra magnetic lenses are fitted in addition to the normal set of condenser, objective and projector lenses, giving a shorter and more compact instrument which, by virtue of the reduction of the microscope tube diameter between lenses, is of sufficiently small volume to be evacuated between specimen insertions in 15 sec. The magnification is continuously variable from 1000 to 50,000, without

* By courtesy of R.C.A. Photophone Ltd., London, W.1.

changing lens pole-pieces. A special, adjustable diaphragm enables any small area of any part of the specimen to be examined at will. Photography is done on 35 mm. film, enabling forty exposures to be made before reloading.

The Plessey Co. Ltd. have also produced an electron microscope of the magnetic type, a special feature being a form of mechanical stage incorporating an indexing mechanism enabling the position of any part of the specimen to be fixed within 0·00005 in.

CHAPTER 13

Ultra-High Frequency Thermionic Tubes

Capacity and Inductance Associated with Valve Structure. At radio frequencies, particularly those greater than 10 Mc./s., the inter-electrode capacities of valves become such low reactances that they pass appreciable currents. Since such currents do not depend on the normal influence of the electrode potentials on the electron streams within the valve, so these capacitative currents vitiate the valve performance. Suppose the capacity between a pair of valve electrodes is 10 $\mu\mu$F. At a frequency of 10 Mc./s. the corresponding reactance is given by

$$X_C=\frac{1}{2\pi fC}=\frac{1}{6{\cdot}28\times10^{7}\times10\times10^{-12}}=\frac{10^{4}}{6{\cdot}28}=1590\,\Omega.$$

If the H.F. applied potentials are of the order of a few volts, then capacitative currents of a few milliamps. will be entailed: such currents being as significant as the usual electron currents due to the normal electronic action of the valve. The grid input impedance of even the best types of H.F. pentode, adopting the usual constructional methods, becomes lower than the dynamic resistance of a 10 Mc./s. rejector circuit anode load.

Again, suppose a wire used as an electrode lead in the valve construction is 10 cm. long and 1 mm. diameter. This will have an inductance of about 0·1 μH. Consider the reactance of such an inductance at 10 Mc./s.: $X_L=2\pi fL=6{\cdot}28\times10^{7}\times0{\cdot}1\times10^{-6}=6{\cdot}28\,\Omega$. Such a value is usually low enough to be insignificant, but at ten times this frequency, at 100 Mc./s., the inductive reactance will be 62·8 Ω., and may be an appreciable source of loss of potential.

Suppose L_C=cathode lead inductance in a particular case, and I_C=alternating component of cathode current. Then the volts drop across this lead at a pulsatance ω is $\omega L_C I_C$, and this voltage, say E_C, lags by 90° in phase on the input signal to the valve. Let the grid-cathode capacity concerned be C_{GC}. The volts drop across the inductive lead produces a capacitative current through C_{GC} which leads the voltage E_C by 90°, and is therefore in phase with the input signal voltage. The current through C_{GC} is

$\dfrac{\omega L_C I_C}{1/\omega C_{GC}} = \omega^2 L_C C_{GC} I_C$. Since this current is in phase with the input voltage, it corresponds to a resistive effect involving a power loss which damps the input tuned circuit. If ω is high then, since this current depends on ω^2, so the effective resistance placed

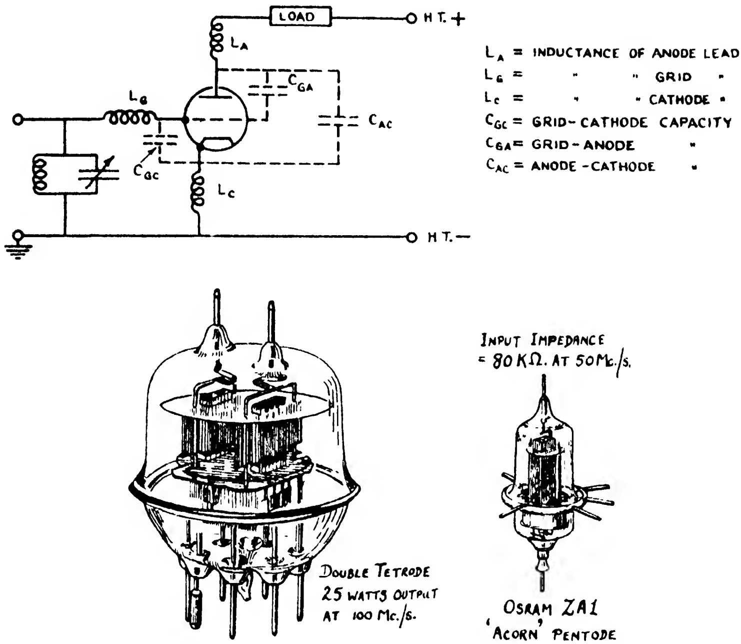


FIG. 114. *a*, Inductances and Capacities associated with a Valve Circuit. *b*, U.H.F. Valves.

across the anode load of any preceding amplifier valve will be so low as to reduce seriously the gain of this previous stage.

Electron Transit Time. At very high frequencies, especially at those in excess of 100 Mc./s., the time taken by an emitted electron to traverse the distance between the electrodes in a valve, known as the *electron transit time*, becomes appreciable compared with the period occupied by one H.F. cycle. Emitted electrons then no longer follow out instantly changes in electrode potentials. In a valve used as a normal H.F. amplifier, or oscillator, the number of electrons approaching the grid at any instant is equal to the number of electrons departing from the grid towards the anode, provided that the grid is negative. However, when the frequency is so high that the grid potential changes can occur before the electrons they influence have time to traverse the necessary space, then the number of electrons approaching the grid may not be equal to the number receding from it. When a negative electron approaches a grid it induces a positive charge on it, by the usual electrostatic effect; vice versa, on leaving the grid, a negative charge is induced. In valves operating at normal radio frequency, since the number of electrons approaching a negative grid equals the number departing from it, so these induced charges cancel out; the nett result on the grid potential is zero. In the very high frequency case, however, since these electron numbers are not equal, so the resultant grid charge may be positive or negative. If the induced grid charge due to electron motions is positive, then grid current will flow due to attracted electrons, even though the grid bias is nominally negative. The grid current loss resulting is so great at very high frequency as to damp out the input signal to the valve.

The phase displacement θ between the grid potential peak and zero instantaneous current depends on the transit time and the frequency, whilst the grid current produced increases as the frequency squared. The energy given up to the grid raises its temperature.

The usual form of feed-back oscillator requiring a valve electron current in phase with the grid voltage and 180° out of phase with the anode voltage for efficient operation, will be upset at frequencies where the electron transit time occupies a considerable fraction of the time of one cycle, since the current due to electrons arriving at the anode lags behind current due to electrons leaving the cathode. If this lag is about 30°, oscillations can occur, but with low efficiency. A lag due to this cause of as much as 90° will prevent the circuit oscillating. Thus it becomes impossible to build the usual feed-back oscillator in which the electron transit time is as much as 25% of the oscillation period.

Types of Valve Used at U.H.F. The difficulties cited above have been fairly adequately overcome at frequencies up to 50 Mc./s., and partly overcome at still higher frequencies up to 750 Mc./s. in the case of receiver valves, by the construction of tubes with very small electrode sizes and short leads. Examples are the "Acorn" type valve, like the Osram ZA 1, and the "all-glass" H.F. pentode such as the Mullard EF 50. Thus the ZA 1 has an input impedance of 80,000 Ω. at 50 Mc./s.

It is unfortunate that the demands of low capacity between electrodes combined with short electron transit time are somewhat in conflict with one another. If the distance between any two electrodes concerned is increased then, for a given area, the capacity is reduced, but at the same time the transit time necessary is increased. The tendency in receiver valve design has therefore been to ensure low capacity by the use of small electrodes, whereas in transmitter valves, where severe limitations of electrode size limit the power handling capacity of the tube, the remedy has been to increase the electron velocity, and so reduce the transit time, by the use of higher electrode potentials.

To produce oscillatory currents at frequencies above 500 Mc./s., the ordinary feed-back method is abandoned. Oscillators are constructed in which electron transit time is not a bugbear, but is the factor deciding the production of oscillations. Such oscillators, capable of working at frequencies up to 1000 Mc./s., and more, are the Barkhauszen-Kurz, the Magnetron and Klystron oscillators.

Barkhauszen-Kurz Oscillator. A triode valve is operated with its grid at a high positive potential, and the anode at cathode potential, or slightly negative. Emitted electrons travel towards the grid, and those which do not hit grid wires will pass towards the anode. Since the grid spaces cover a much greater area than the grid wires, so the majority of the electrons will go near to the anode. On approaching the anode, the electrons encounter a retarding field and are consequently decelerated to zero velocity, reverse their path and travel back to the grid again. Depending on their initial path relative to the grid wires (see fig. 115) so the electrons will oscillate some number of times back and forth about the grid. In the end, however, the electrons are removed from active operation by hitting a grid wire.

The electrons will oscillate about the grid even if an external circuit, other than the power supply, is not present.

Consider how power is supplied to an appropriate *LC* circuit, which usually takes the form of a pair of lines connected as shown in fig. 115. On switching on the H.T. supply, a transient oscillation of this *LC* circuit takes place. Suppose the potential of the *LC* circuit at a particular instant is such that it makes the valve grid positive during the passage of a given electron from cathode to grid, and is negative when the same electron is moving from grid to anode. The *LC* circuit frequency must then be approximately twice the electron oscillation frequency. In such circumstances the *LC* circuit voltage will increase the electron acceleration between cathode and grid, but decrease its deceleration between grid and anode. The electron will thus be absorbing energy from the *LC* circuit. But such an electron will be collected by the anode, and so be removed from active operation, because

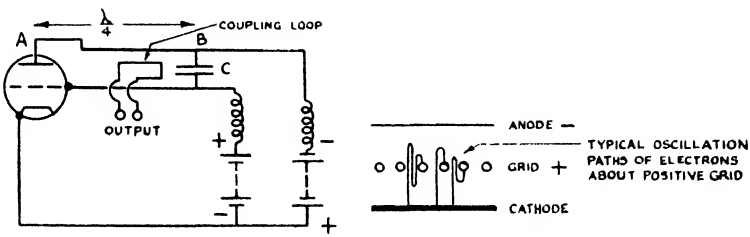


FIG. 115. The Barkhauszen-Kurz Oscillator.

it travels through the grid faster than it would do in the absence of the *LC* circuit, and is not decelerated so much as it would be by the negative anode.

On the other hand, consider the operation 180° earlier or later in the *LC* oscillation cycle, when the *LC* potential is reversed. Then this potential will decrease the acceleration of the electron in going from cathode to grid, and increase its deceleration between the grid and anode: the electron motion is opposed throughout its path. Now the electron will not reach the anode, but will oscillate about the grid, delivering energy to the *LC* circuit. The energy delivered to the *LC* circuit in this case is greater than that absorbed by an electron in the previous case, since the electrons will now be able to supply energy for a number of cycles before they are eventually collected by the grid. The nett result is to tend to maintain *LC* circuit oscillations.

Such an oscillator can generate wave-lengths between 1 cm. and 1 m. but the efficiency is only 2 to 3%; it therefore finds very little use in transmitter practice. Moreover, the positive grid tends to rise in temperature, and the heat dissipated at the valve grid cannot be effectively removed by an efficient cooling device. Generally, the triodes used as Barkhauszen oscillators are of a concentric cylinder construction to ensure that a continuous virtual cathode is maintained just inside the slightly negative anode.

For a tube with planar electrodes, neglecting space-charge and thermal emission velocities, and assuming the electrode alternating potentials are small,

$$\lambda = \frac{2000}{\sqrt{E_G}} \cdot \frac{d_A E_G - d_G E_A}{E_G - E_A}, \qquad (288)$$

where λ = oscillation wave-length, E_G = grid potential (D.C.), E_A = anode potential (D.C.), d_A = anode-cathode spacing, d_G = anode-grid spacing.

If E_0 = alternating grid potential, and the grid is midway between cathode and anode, then

$$\lambda = \frac{4000 \sqrt{(E_G - E_0)}d}{E_G - (4E_0/\pi^2)}, \qquad (289)$$

where $2d$ = anode-cathode spacing.

For a cylindrical array, $\lambda = \dfrac{670d}{\sqrt{E_G}}$, $\qquad (290)$

where d = anode diameter, and the grid-cathode spacing equals the grid-anode spacing, both being small.

Gill-Morrell Oscillations. If E_0, the grid potential fluctuation, is large, then the frequency is decided chiefly by the LC circuit, and not, as in the case of Barkhauszen oscillations, by the electron transit time. Now the LC circuit potential is sufficiently great compared with the steady grid voltage that it is capable of so modifying the extent of the excursions of the effective oscillating electrons that their amplitude and frequency adjust themselves to suit the tuned circuit. A band of frequencies is then produced of which the lower limit has approximately a period equal to twice the electron transit time from grid to anode.

The Magnetron. This tube, which is capable of generating very high frequency oscillatory currents, consists essentially of a cylindrical form diode whose anode current is controlled by an external magnetic field, of which the lines of force are directed along the diode axis.

Since electrons leave the thermionic cathode with virtually zero velocity, their paths will be cycloids (see equation (69), p. 47), assuming the electrostatic field is uniformly radial, and space-charge is neglected.

For a given anode potential (in practice about 1000 V.), and weak magnetic field, the curvature of the cycloid will be low so that the electrons will readily reach the anode, forming an anode current. As the magnetic field is increased, the generating radius of the cycloid will be reduced until a point is reached at which the electron path is a closed curve only just reaching the anode surface. Magnetic fields greater than this critical value will completely cut off the anode current. For a cylindrical type of anode, the cut-off magnetic field is given by the formula

$$H_C = \frac{6{\cdot}74\sqrt{V}}{R}, \qquad . \quad . \quad . \quad (291)$$

where V is the anode voltage, and R is the anode radius in cm.

The magnetron characteristic, anode-current vs. magnetic field strength, will then be as shown in fig. 116*c*.

Though formula (291) indicates that the cut-off effect of the magnetic field should be sharp, yet a characteristic curve obtained in practice indicates that a small anode-current is still obtained for values of field slightly larger than the cut-off value given by (291). This is due to the distribution of velocities amongst the thermally emitted electrons, whereas the formula (291) assumes such velocities are zero.

The magnetron principle is applied to the development of two types of U.H.F. oscillator. Firstly, it can be utilised as a negative-resistance device, acting in the so-called *dynatron régime*, where the frequency of oscillation is decided by an external *LC* circuit, and the maximum frequency which has been developed is 600 Mc./s. ($\lambda = 0{\cdot}5$ m.), the limitation being set by electron transit time; secondly, it can be used in the *electronic régime*, in which the electrons spin round the cathode, inducing U.H.F. currents on the anode electrodes. In this case, the oscillation frequency is governed by the electronic transit time, and frequencies as high as 30,000 Mc./s. ($\lambda = 1$ cm.) have been generated, using a multi-segment construction.

The Dynatron Régime. The anode is split into two equal halves about a diametral plane containing the cathode axis. Across these halves, an oscillatory circuit, often in the form of a parallel pair of bar terminated transmission lines, or Lecher wires, is connected, with H.T. fed at a centre-tap as shown in fig. 116*e*.

Consider that the external *LC* circuit begins to oscillate freely, and that at a particular instant the upper half *A* of the anode

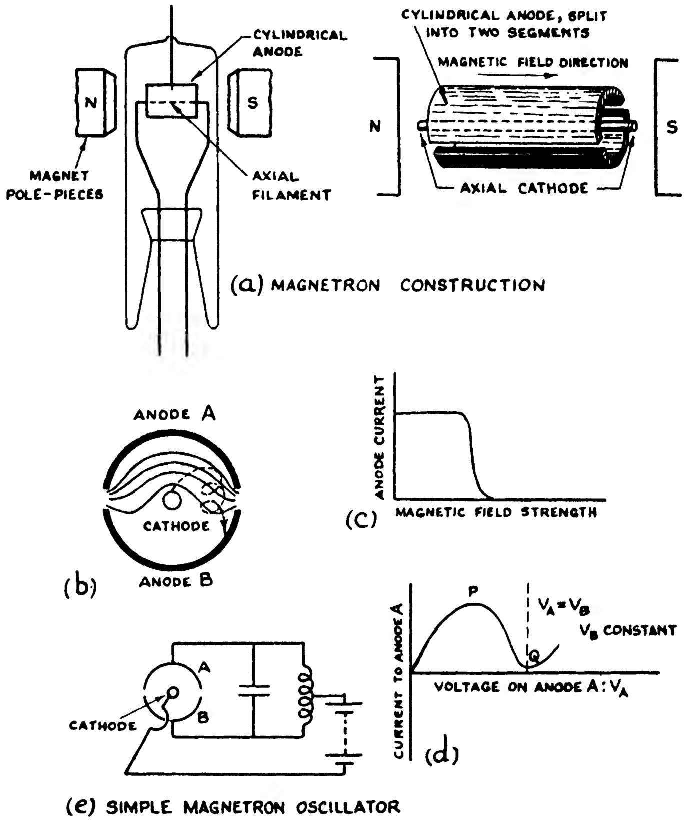


FIG. 116. The Magnetron (Dynatron Working).

indicated in fig. 116*b* is positive with respect to the lower half *B*. If the applied magnetic field is greater than the cut-off value, an electron on entering the region of half-anode *A* will be deflected to follow a path of a certain radius of curvature. It will curve towards the half-anode *B* where, since the potential relative to the cathode is lower, it will traverse a path of less radius of curvature because of the then greater effect of the magnetic field compared with the electric field. The electron will then, under suitable operating conditions, execute a series of small loops, finally reaching the lower potential plate. Such electrons have then been forced to work against the action of the existing electric field; they correspond to an electric current flowing against the direction of an applied P.D., i.e. they constitute a negative resistance. This negative resistance can be balanced against the positive resistance of the *LC* circuit, reducing losses to zero, and so enabling continuous oscillations to be performed.

Such oscillators can be made up to 60% efficient by using resonant circuits of high impedance, with a magnetic field considerably in excess of cut-off. Under these conditions the one anode will exceed the H.T. battery E.M.F. by a considerable fraction of its total voltage, whilst the other anode is correspondingly negative. The losses in the tube are then small, whilst the use of a high magnetic field strength ensures that once electrons enter the lower potential half of the split-anode system, they will execute loops of such small radius of curvature as to have little chance of escaping to the more positive anode region.

That such a split-anode magnetron does possess a negative-resistance characteristic can be shown by plotting a static characteristic curve for the tube of P.D. on one anode, V_A, against the current collected by it, where the other anode potential, V_B, is kept constant. Thus in fig. 116*d*, where the magnetic field is above the critical cut-off value, when V_A is very small there will be little current collected by it, most of the current going to the other anode. As V_A is increased the electron flow to anode *A* will increase, the current to anode *B* decreasing. But when $V_A = V_B$, then the critical magnetic field operates, the electrons follow cycloidal paths and miss both anodes. It follows that the current to *A* is zero when $V_A = V_B$. There must therefore be a peak current to anode *A* at some point *P* on the characteristic curve, fig. 116*d*, whilst between *P* and *Q* the anode current to *A* decreases with rise of potential V_A on *A*; equivalent to a negative resistance effect (cf. the dynatron oscillator, p. 183).

Since a strong magnetic field is required for this oscillator, so the total anode length is limited to a few centimetres if the overall magnet size is to be of reasonable dimensions. Moreover, the diameter of the pair of anodes should not be too great if very high frequency oscillations are to be produced. These two considerations limit the size of the usable anode, and hence seriously limit its power handling capacity. To offset overheating of the anode when considerable power dissipation is required, a thick metal anode, air or water-cooled, is used.

A second difficulty is that many electrons, after performing several loops in the magnetic field, return to, and bombard the cathode, causing its temperature to rise. Some circuits incorporate an automatic device which lowers the anode potential when the cathode temperature rises inordinately.

The Electronic Régime. Oscillations in this case depend on the transit time of the electrons, typical electron paths being shown in fig. 117. In order that oscillatory conditions are achieved it must be shown that such electrons do, on balance, deliver energy to an external resonant circuit.

To simplify the consideration of the mechanism, examine the case of a single cylindrical anode magnetron, with the magnetic field just beyond cut-off, and, due to the free oscillation of the external circuit, a small alternating voltage is superimposed on the steady anode potential of which the period is equal to the transit time of an electron following a complex spiralled path from cathode to anode. An electron leaving the cathode when the anode potential is just about to become more positive under the action of the applied alternating potential will undergo a greater acceleration towards the anode than an average electron, and will therefore either be collected by the anode, or approach closely to it before returning towards the cathode. On the return journey the applied A.C. potential will be executing a negative half-cycle, the cathode then being positive with respect to the anode, so that the electron will return to the cathode. Such electrons will acquire energy from the external alternating source, and they therefore retard rather than assist the possibility of self-oscillation of the circuit.

On the other hand, electrons which leave the cathode during the time when the anode potential undergoes the negative half-cycle of the superimposed alternating potential will not reach the anode or readily return to the cathode again. This is because their path towards the anode is retarded by the decreasing anode potential whereas, half a period later, their return towards the cathode is decelerated by the increasing anode potential. These electrons therefore deliver energy to the external circuit, and can make several excursions of diminishing amplitude to and fro before

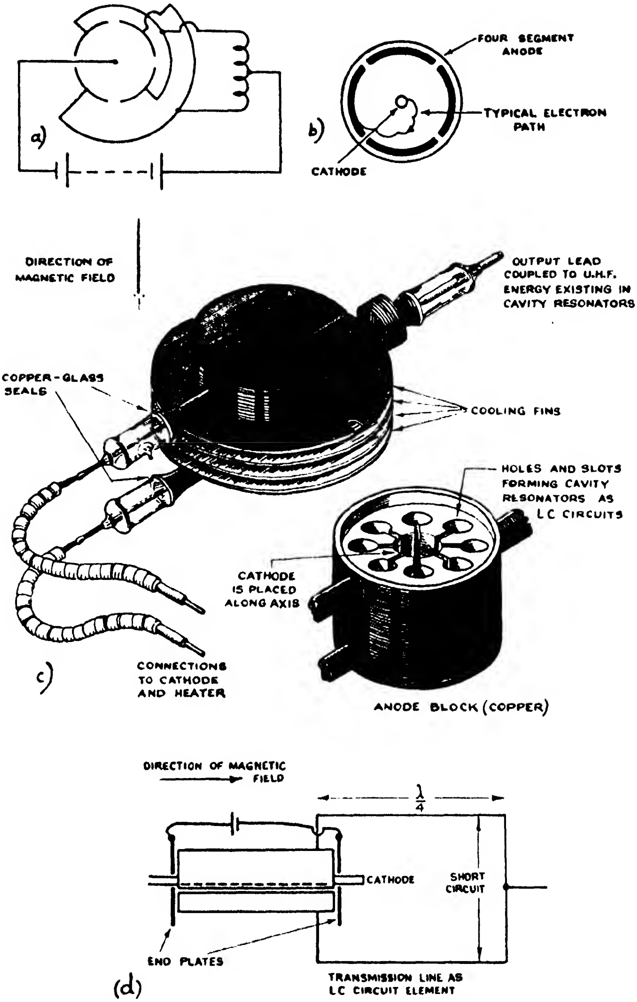


FIG. 117. Magnetrons Used in Electronic Régime.

coming to rest. Over the whole cycle a total positive amount of energy is delivered to the external circuit, sustaining oscillations.

In the absence of space-charge, the wave-length corresponding to a single cylinder construction is given by the formula

$$\lambda=\frac{12,300}{H}, \quad . \quad . \quad . \quad . \quad . \quad (292)$$

where λ is in cm., and H in lines per sq.cm. If the space-charge is saturated, then

$$\lambda=\frac{16,700}{H} \quad . \quad . \quad . \quad . \quad . \quad (293)$$

In practice, with emission limited space-charge, a wave-length somewhere between those given by formulae (292) and (293) is obtained.

Better results are obtained with a split anode arrangement. The basis of the mechanism remains the same, though the electron orbits are the more complex. To achieve short wave-lengths without restricting the anode size so greatly as to make its power dissipation possibilities very limited, the modern multi-segment magnetron has a four- or eight-segment anode, where the electrode structure itself forms the resonant LC circuit. In this way magnetrons of the constructional type illustrated in fig. 117*c* have been built capable of power outputs of 100 to 200 W. with an efficiency of over 50%, at wave-lengths of a few centimetres. Under the pulse conditions, used in radar practice, outputs of the order of megawatts have been achieved.

To remove from the field between the cathode and anode the electrons which finally come to rest so that they do not cause the accumulation of excessive space-charge, an end-plate at positive potential is added to the tube (fig. 117*d*), or the magnetic field lines of force are arranged to have a few degrees of tilt with respect to the cathode axis so that these "dead" electrons receive a small component of velocity along the tube axis to remove them from the electronically active region.

In place of the usual axial cathode, the electron supply for the magnetron oscillations can be furnished from an electron gun which fires electrons from the end of the cylindrical structure into the anode space, the electrons entering through an end-plate aperture so that the magnetron anodes are shielded from the gun. This entering electron beam may, moreover, be controlled by a grid so that the magnetron can be operated as a U.H.F. amplifier as well as having oscillatory properties.

Velocity-modulated Electron Tubes. Of recent years, particularly since the Varians'* invention of the Klystron in 1939, the trend of development of electron tubes capable of generating

* R. H. and S. F. Varian, *Jour. of Appl. Phys.*, **10**, 321, 1939.

and amplifying centimetre waves has undergone radical changes. The chief new techniques which are being adopted are the use of a rhumbatron as an LC circuit, forming part of the valve electrode structure, and the use of a velocity-modulated electron beam as a means of controlling the introduction of U.H.F. energy into such a resonant LC circuit.

In the velocity modulation of an electron beam, varying velocities are imparted in a controlled manner to a stream of electrons. This has predominantly the effect of changing a continuous stream of electrons into a "bunched" beam in which the electrons travel forwards in regions of high space-charge density separated by gaps in which very few electrons are present.

The manner in which this can be achieved is illustrated in fig. 118*a*.

Suppose a thermionic cathode in a vacuum tube emits electrons which are roughly concentrated into a beam by a cylinder at earth, or negative potential. Beyond this emitter is situated a grid G at positive potential. The electrons which penetrate this grid will have practically constant velocity if they enter field-free space. This velocity is given by the usual energy relationship

$$\tfrac{1}{2}mv^2 = Ve, \quad . \quad . \quad . \quad . \quad (294)$$

where v is the velocity, and V the positive potential of the grid G.

Within the nominally field-free space which the electrons enter, let a pair of grids E and F be situated which the electrons penetrate. Suppose an alternating potential is connected across these grids so that their potentials vary in sinusoidal fashion about the mean potential V. This alternation is assumed to have very high frequency, yet such that the time of flight of the electrons between the grids is short compared with the alternation period.

Consider an electron passing through the centre of the H.F. field just as that field is changing from opposing to aiding electrons, i.e. H.F. field effect is instantaneously zero (point X on fig. 118*a*). Since this electron passes through the space EF between the grids before the H.F. potential has time to change significantly, it will reach the third grid F with much the same speed as it had at the second grid E.

A second electron, travelling initially just ahead of the first, will necessarily meet with a retarding effect due to the negative half-cycle of the R.F. field (point Y), so its speed after the third grid F will be reduced. On the other hand, a third electron initially behind the first, will encounter an aiding field due to the positive half-cycle of the R.F. field (point Z). It will have its speed increased beyond grid F.

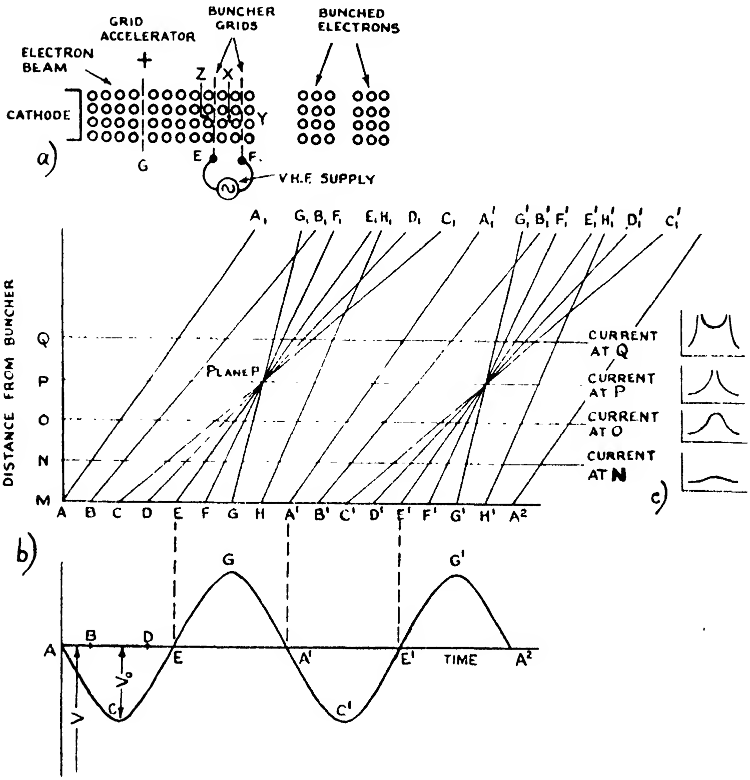


FIG. 118. *a, b, c,* The Velocity modulation of an Electron Stream.

In the field-free space beyond the third grid these differences in velocity cause electrons ahead and behind the one of unchanged speed to draw nearer to it: the one which was lagging speeds up, tending to catch the electron of unchanged velocity, the one which was leading slows down.

An opposite state of affairs will exist a half-cycle earlier or later. Reference to fig. 118*a*, combined with deduction similar to that already made, indicates that the electron in front of the one of unchanged speed will be accelerated, whilst the lagging electron will be decelerated. Now the electrons will move farther apart beyond the third grid F.

So the electron stream beyond the grid F traverses field-free space in "bunches", regions of a high number of electrons per c.c. will be separated by regions of very low electron numbers.

The Mathematics of Electron Bunching. At the buncher grids, electrons encounter an accelerating voltage which varies sinusoidally with time. Their velocity on emerging will be proportional to

$$\sqrt{(\text{Accelerator potential} + \text{alternating potential across buncher})}$$

or
$$v = \sqrt{\left[\frac{2e\,.\,10^7}{m}\left(V + V_0 \sin \omega t\right)\right]} \text{ cm./sec.}$$

$$= 5{\cdot}93 \times 10^7 \sqrt{(V + V_0 \sin \omega t)} \text{ cm./sec.} \quad \text{(ref. p. 43)}, \quad (295)$$

where V=accelerator grid potential in volts, V_0=amplitude of H.F. potential across grids of buncher, ω=pulsatance of H.F. supply to grids.

Equation (295) can be written

$$v = 5{\cdot}93 \times 10^7 \sqrt{[V(1 + m \sin \omega t)]}, \quad . \quad . \quad (296)$$

where $m = V_0/V$=the depth of modulation.

Here the small effects of mutual repulsions between the electrons due to their like charges have been neglected.

These results can be usefully expressed in terms of the Applegate diagram of fig. 118*b*, from which the variation in current in the beam at various points along it can be deduced. In this diagram the horizontal axis is a time axis, whilst the vertical axis represents distance beyond the buncher. Lines AA_1, BB_1, etc. drawn on this chart have slopes which are proportional to the velocities of the electrons. Thus AA_1, EE_1 and $A^1A_1{}^1$ are lines all of the same slope, since they correspond to electrons released when the instantaneous buncher potential equals the D.C. accelerating potential, the A.C. potential being zero. On the other hand, lines such as CC_1 have a minimum slope, corresponding to times in the H.F. cycle when the A.C. potential across the buncher grids is at a negative peak, whilst lines such as GG_1 correspond to this A.C. potential being at a positive peak, and therefore have maximum slope. The slopes of lines moving from left to right of the diagram correspond to paths of electrons leaving the buncher at successive equal time intervals; their slopes depending on the electrons' velocities in accordance with formula (296) where t increases from zero in steps represented by intervals AB, BC, CD, etc., on the time axis.

If the initial emission of electrons is uniform, then the reciprocal of the distance between lines will be proportional to the current, and where lines such as CC_1 and GG_1 intersect at plane P, corresponding to some definite distance from the buncher, bunching occurs giving a region of high electron density, corresponding to a very large instantaneous current value. It can be deduced from the Applegate diagram that the electron currents at the various planes N, O, P, etc., vary with time in accordance with the graphs of fig. 118*b*. The peaked nature of these wave-forms indicates that the current is rich in harmonics.

Webster* has shown that the instantaneous current i at distance l beyond the buncher is represented mathematically by

$$i = I_0[1 + 2\{J_1(K) \cos \omega t + J_2(2K) \cos 2\omega t + J_3(3K) \cos 3\omega t + \ldots + J_n(nK) \cos n\omega t\}],$$

where $K = \pi l V_0 / LV$, L is the average distance traversed by an electron in one cycle, V and V_0 are defined as in equation (295), and J's are Bessel functions (see p. 320). Thus, knowing K in a particular case, and evaluating J_1 (K), J_2 $(2K)$ etc. from the data provided on p. 320, the current wave-form can be expressed in trigonometrical form, from which the component currents at the various harmonics can be deduced. I_0 is the initial beam current.

Thus $$i = I_0(1 + a \cos \omega t + b \cos 2\omega t + \ldots + n \cos n\omega t), \qquad (297)$$

where a, b, . . . n are computed, knowing K and using the graph of Bessel functions.

The Rhumbatron, or Cavity Resonator. In oscillator practice at frequencies below 100 Mc./s. the use of the conventional inductance-condenser combinations considered in Chapter 4 is satisfactory. At frequencies up to 600 Mc./s., the use of quarter wave lines, etc. (see p. 77), is common. In the Klystron and multi-segment magnetron valves, capable of oscillations at

* D. L. Webster, *Jour. of Appl. Phys.*, 10, 501, 1939.

frequencies of 30,000 Mc./s. ($\lambda = 1$ cm.), such LC combinations become impracticable since the values of inductance and capacitance demanded are so small that the valve electrode capacities and external circuit leads become predominant in deciding the effective LC value. The problem has been solved by the use of resonant circuits in the form of closed, or almost closed, vessels of some specific geometrical shape with walls of copper, or other highly conducting material. Such cavity resonators, or rhumbatrons, form the main part of the actual electrode construction of the multi-segment magnetron and Klystron oscillators: the LC circuit necessary is a part of the oscillator valve itself. This introduces new problems as to how to feed such resonators with U.H.F. energy, how to tune them and how to feed the U.H.F. energy developed to any external circuit such as a wave-guide, aerial system or measuring device.

Hansen* considers a closed surface of conductivity assumed infinite to avoid resistance losses. It is known that there exist solutions of Maxwell's equations for the electromagnetic field which are zero outside the surface, finite inside, have a definite frequency and necessarily a vanishing tangential component of electric field at the conducting boundary. Thus an oscillating electromagnetic field can be supported inside such a conductor, but at the inside surface of the conductor the component of the electric field along the surface must be zero, since otherwise alternating P.D.'s would be set up across the conductor which would produce large H.F. currents, rapidly dissipating the energy in any such tangential electric field component. From this consideration, that the electric field lines of force must, at all points of the inside walls, be perpendicular to the surface,† it follows that only certain discrete frequencies are allowed. In other words, the volume and shape of the cavity resonator used decides the value of the resonant electrical oscillations it can sustain. At any other frequencies the losses will be very large, so that this device becomes comparable with a sharply resonant tuned circuit of very high "Q."

Hansen considers for cavity resonators the effective values of

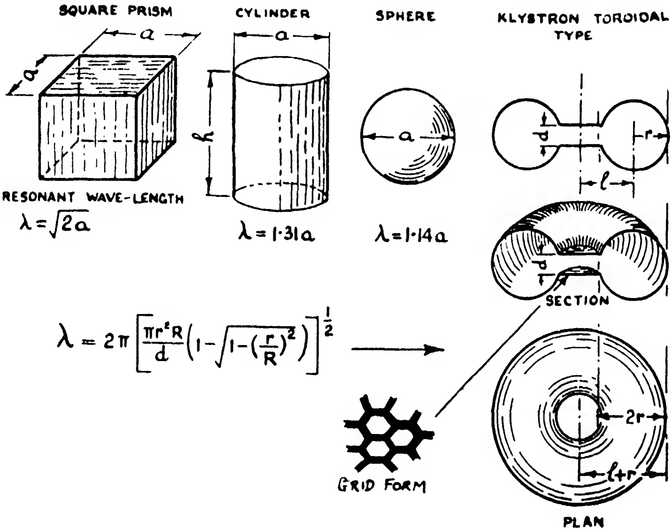
* W. W. Hansen, *Jour. of Appl. Phys.*, **9**, 654, 1938; and W. W. Hansen and R. D. Richtmyer, *Jour. of Appl. Phys.*, **10**, 189, 1939.

† In accordance with electromagnetic theory, it follows that the magnetic field variation will necessarily be parallel to the cavity walls at the inner surface, since the electric and magnetic field vectors must be at right angles to one another.

$\sqrt{LC}$ determining the resonant frequency, $\sqrt{(L/C)}$ which is the characteristic impedance of resistance dimensions,

$$Q=\frac{\omega L}{R}=\frac{1/\sqrt{(LC)}\,.\,L}{R}=\frac{\sqrt{(L/C)}}{R}$$

as a measure of the losses in the resonator, and R_S the shunt impedance, comparable with the dynamic resistance at resonance, L/CR, of a parallel tuned circuit.



$$\lambda = 2\pi\left[\frac{\pi r^2 R}{d}\left(1-\sqrt{1-\left(\frac{r}{R}\right)^2}\right)\right]^{\frac{1}{2}}$$

FIG. 119. Cavity-Resonators, or Rhumbatrons.

$\sqrt{(LC)}=\lambda/2\pi c$ gives λ_m, the maximum wave-length corresponding to the lowest allowed frequency, c being 3×10^{10} cm./sec., the velocity of electromagnetic radiation. In a cavity resonator, all three dimensions are in general comparable with a wave-length, and nodes can exist in three different directions. Thus the highest resonant frequency is not necessarily an integral multiple of the lowest one, as in the case of a long transmission line. Various modes of oscillation in the same shaped cavity are possible: any configuration of electromagnetic field variation which can be produced within the cavity (providing the electric field lines are normal to the inner surface, and of course normal to the magnetic field) being allowable.

For a spherical rhumbatron, $\lambda_m = 1{\cdot}14 \times$ diameter. This factor 1·14 is understandable on considering that the wave-fronts are strongly curved to allow the electric field to be normal to the inside spherical surface, so that the distance between successive nodal surfaces, or surfaces of zero electric field strength, will be greater than the distances between nodes for plane waves of the name frequency.

For a cylindrical cavity (fig. 119) $\lambda_m = 1{\cdot}31a$, where a is the cylinder diameter, and λ_m is independent of the cylinder length.

$Q = \frac{\sqrt{(L/C)}}{R}$ as a measure of losses can be shown to be equal to π times the ratio of the energy stored in the resonator to the energy loss per half-cycle, which is equal to $\frac{\pi(\text{volume of resonator})}{\text{surface area} \times \text{skin depth}}$. In general, a shape which provides an increased ratio of the volume to the surface area increases Q. An electron beam, entering a resonator and producing a space-charge which depends upon the beam velocity and current, causes a reduction of Q. The presence of loops within the resonator which encircle the lines of the magnetic field, and enable the electromagnetic energy within the cavity to be coupled by a coaxial line to another resonator, or to an aerial system, also influences the value of Q, as does the load presented by the external circuit on such loops plus transmission lines. Losses in the conductor material of the resonator are not directly proportional to the resistivity of the conductor, but to the square root of this value, since the skin resistance effect depends on the depth of penetration of the eddy currents within the conductor. Magnetic materials are to be strongly deprecated for the manufacture of resonator cavities since large values of permeability introduce excessive eddy-current loss.

The shunt impedance R_S is reduced by losses to the same degree that such losses reduce Q. The shunt impedance of a rhumbatron is of great importance in deciding how to sustain oscillations within it, since it decides the minimum "make-up" current required to sustain oscillations.

In velocity-modulated tubes, such as the Klystron, the rhumbatrons used are re-entrant cavities (see fig. 119).* Such a design is necessary to allow the electrons to traverse the cavity in a short time compared with the time of one half-cycle of the oscillations

* In Fig. 119, note $R = l + r$, in case of Klystron rhumbatron.

produced. Moreover, the parallel disk geometry of the type of rhumbatron used allows an effective capacity to be obtained across which there is furnished a strong, uniform electric field. These disks are in grid form (fig. 119) to allow a beam of electrons to enter and leave the rhumbatron with a minimum interception of electrons, yet with a robust structure to avoid damage due to over-heating. The volume of these re-entrant rhumbatrons is kept large, to achieve a satisfactory effective inductance value, by the ring-shaped cavity around the parallel-disk geometry. In this way, a satisfactory value of shunt impedance, leading to efficient working, is achieved.

The Klystron. This tube is capable of acting as a generator, amplifier, mixer, multiplier or detector of U.H.F. oscillations of wave-lengths between 1 and 10 cm. with a practical efficiency of about 35%, and a maximum theoretical efficiency of 58%.

An electron gun in a vacuum tube consists of a small, disk-shaped indirectly heated cathode with the usual oxide coating, and a focusing ring-shaped electrode (fig. 120*a*). Sometimes a grid or aperture for controlling the beam current is provided. The cathode heater must be non-inductive if heated by 50 c./s. A.C. to avoid the possibility of a low-frequency alternating magnetic field being produced. Alternatively, D.C. heating may be used. This gun produces an approximately parallel beam of constant velocity electrons which enter the buncher.

The buncher grids necessary to obtain a velocity-modulated electron beam are arranged as two networks of hexagonal apertures at the inlet and outlet of a re-entrant type of parallel-disk geometry rhumbatron (fig. 119). This rhumbatron is arranged to have its parallel sides at such a separation, and yet with sufficient total internal volume, that its shunt impedance and "Q" are high, and yet the transit time of the electrons between the grids is short compared with the period of oscillation.

Beyond this buncher rhumbatron, the electrons traverse a critical length of field-free space in a metal cylinder, or "drift" tube. This length of unaffected flight is necessary to enable the velocity-modulated electron beam to bunch effectively before entering a second rhumbatron, called the "catcher". This catcher is of precisely the same form as the buncher. The velocity modulated electron beam bunches must enter the catcher at that time in the oscillation cycle when the catcher alternating potential is negative, tending to retard electron entry so that the electrons are capable of giving up their energy to the radio-frequency field of the catcher resonator.

The remaining electrode, farthest from the cathode, is a positive collector, beyond the catcher.

Energy from the catcher is transferred to a small conducting loop inside the catcher resonator, which loop is in a plane at right angles to and encircling some of the magnetic lines of force. The rapidly fluctuating magnetic field within the catcher induces a corresponding current in this loop, and the energy is fed via a short coaxial line back to a similar loop within the buncher rhumbatron.

If the H.F. power produced by the entry of the velocity-modulated electron beam into the catcher is greater than the power required at the buncher to produce the beam modulation, then such an arrangement is capable of continuous oscillation at a frequency decided by the cavity dimensions and mode of oscillation, so long as the H.T. supply is maintained.

The energy delivered by the buncher to the electrons is insignificant, being only that required to overcome the buncher resistance losses. When electrons first enter the buncher, free oscillations at the buncher resonant frequency begin, since the rhumbatron capacity becomes charged up, and the entering electrons have superimposed fluctuations due to shot effect.

The number of electrons accelerated during one positive half-cycle of buncher oscillation is equal to the number decelerated to the same, but opposite, extent during the negative half-cycle. Hence the energy given up to the electrons by the buncher is equal to the energy given back to the buncher by the electrons, the net energy transfer being zero over a number of whole cycles, except for the loss due to the buncher rhumbatron skin resistance effect.

At the catcher, on the other hand, the bunched electrons are arranged to have such spacing that they enter its field only during the decelerating half-cycle of catcher oscillations, and therefore deliver considerable energy to it. When the catcher field is an accelerating one, the entering electrons are very small in number, since at such times the sparsely populated spaces between the dense electron bunches are entering the catcher field.

The catcher power is consequently much greater than the buncher power, so the tube can be used as a power amplifier with respect to an alternating potential set up at the buncher, and consequently, with feed-back from catcher to buncher, as an oscillator.

For oscillations to take place in a Klystron, it is essential that

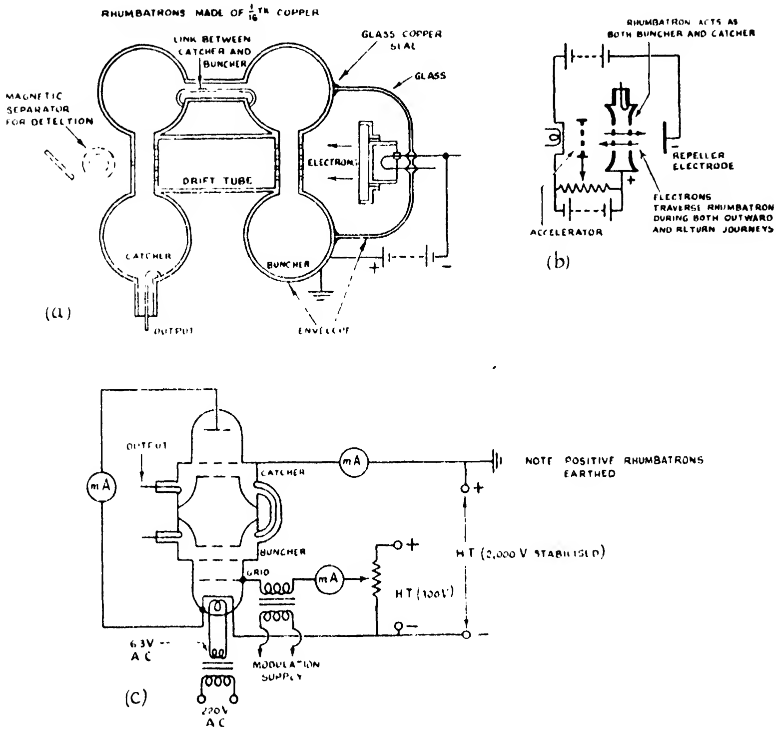


FIG. 120. *a*, The Klystron. *b*, The Reflex Klystron. *c*, Circuit for operating Klystron.

the alternating potential fed back to the buncher from the catcher be in phase with the buncher potential variations. For this to occur, the total phase shift round the electron circuit from buncher to catcher and back to buncher must be $2\pi n$ where n is an integer.

$$\therefore \quad 2\pi n = \frac{2\pi l}{L} + \frac{\pi}{2} + \theta. \qquad (298)$$

where l/L is the electron transit time, l being the length of the drift tube from buncher to catcher, and L is the distance traversed at the mean velocity by the electron in one cycle. θ is the phase angle between the voltage at the buncher and the catcher current. The $\pi/2$ is due to the fact that, to fulfil the condition that the bunched electrons enter the catcher during its retarding half-cycle, the electron at the centre of the bunch at the catcher at maximum retardation must be that which passes the buncher when the alternating potential across it is instantaneously zero.

From (298) the ratio l/L decided by the drift tube length and the mean electron accelerating potential on the buncher, is decided.

Tuning the rhumbatron resonators is usually achieved by either utilising a flexible diaphragm to permit variation of the resonators' grid spacings, or by inserting a plunger into the resonator to change its effective internal volume. Generally, the copper rhumbatrons form part of the actual tube envelope to permit such variations being produced by external screw adjustments.

Since the Klystron is at most 35% efficient in practice, the rest of the electrons' energy is dissipated as heat in the tube structure. The spent electrons from the beam are either reflected to the metal walls of the Klystron by maintaining the collector electrode at cathode potential, or alternatively this electrode is at anode potential so that such electrons are collected by it. In either case, the electrode structure used must be sufficiently robust to withstand the heating produced, this process being facilitated by providing the external electrodes which form part of the Klystron envelope with cooling fins.

To use a Klystron as an amplifier, the signal voltage is fed to the buncher, and the output taken from the catcher. Unfortunately, velocity modulated tubes suffer from very considerable "noise" (see p. 165). A Klystron has been used as a detector by setting up a magnetic field beyond the catcher so that an auxiliary electrode can sort out those electrons of which the velocity has changed in passing through the tube. By tuning the catcher to a harmonic of the buncher frequency, frequency multiplication can be readily achieved in view of the richness in harmonics of the beam current wave-form.

Two other main types of Klystron are the reflex tube which has only one rhumbatron, and the three rhumbatron tubes which have an additional cavity resonator beyond the catcher, the advantage being that variations of power supplied do not affect the oscillation.

The Reflex Klystron. The type of Klystron already described is difficult to adjust because tuning and the spacing of the resonators, deciding the condition for oscillation in accordance with equation (298), are interdependent. The reflex Klystron employing a single rhumbatron which operates as both buncher and catcher is therefore the type generally employed as a generator of oscillations (fig. 120*b*).

After the electrons' first passage through the buncher grids whereby they became velocity modulated, they approach an electrode at a negative potential with respect to the emitter cathode so that they are decelerated to zero velocity in front of this electrode, and then retrace their paths back to the rhumbatron at a mean positive potential. By proper adjustment of the negative potential on the repeller plate, the velocity-modulated electrons may be made to return through the rhumbatron at the correct time to deliver energy to it, i.e. as before, the centre electron of the returning bunch must arrive at the resonator, on the return passage, at the time when the alternating potential across the resonator is at its negative peak value. This tube is then self-oscillating, the back-coupling being provided by the electron beam itself. Spent electrons are removed from the tube by the grids of the resonator, or by the positive accelerator grid. Energy is fed from the cavity to any external circuit by means of a one-turn coupling loop which encircles some of the lines of the magnetic field within the cavity.

By altering the negative potential on the repeller electrode, the frequency of oscillation of this reflex Klystron can be varied over a small range. However, this negative potential also affects the drift time of the electrons, from the rhumbatron as buncher back to the rhumbatron as catcher, since the more negative the electrode, the shorter is the distance traversed by the velocity-modulated beam. Since the time taken to traverse this distance decides the phase relationship between the buncher voltage variations and the feed-back voltage, so the negative potential on the repeller electrode also affects the H.F. power output. The variation of frequency which can be produced by this means is therefore limited unless large variations of power output can be tolerated. Usually the variation of this repeller voltage is employed to achieve only fine frequency adjustments, the main frequency changes being produced by altering the internal volume of the rhumbatron by means of a plunger.

In operating Klystrons at constant frequency the provision of a stabilised H.T. supply is important, since variations up to 30 kc./V. can take place in the frequency of oscillation with change of accelerating voltage (fig. 120*c*).

APPENDIX

A. Physical Constants

Charge of the electron $(e) = 4{\cdot}803 \times 10^{-10}$ E.S.U. (or es. unit).
$= 1{\cdot}601 \times 10^{-20}$ E.M.U. (or em. unit).
Mass of the electron $(m) = 9{\cdot}107 \times 10^{-28}$ gm.
Ratio of electron charge to mass $e/m = 1{\cdot}759 \times 10^{7}$ E.M.U./gm.
Planck's constant $(h) = 6{\cdot}62 \times 10^{-27}$ erg sec.
Boltzmann's constant $(k) = 1{\cdot}379 \times 10^{-16}$ erg/degree.
Gas constant $(R) = 8{\cdot}314 \times 10^{7}$ erg/gram-molecule.
Avogadro's number $(N) = 6{\cdot}027 \times 10^{23}$ per gram-molecule.
Mass of hydrogen atom $= 1{\cdot}662 \times 10^{-24}$ gm.
Velocity of light $(c) = 2{\cdot}99796 \times 10^{10}$ cm./sec.
1 electron-volt $(ev) = 1{\cdot}6 \times 10^{-12}$ erg.

B. Atomic Weights and Atomic Numbers

	At. No.	At. Wt.		At. No.	At. Wt.
Ag Silver	47	107·88	Ne Neon	10	20·183
Al Aluminium	13	26·97	Ni Nickel	28	58·69
Ar Argon	18	39·94	O Oxygen	8	16
Au Gold	79	197·2	P Phosphorus	15	31·02
Ba Barium	56	137·36	Pb Lead	82	207·22
C Carbon	6	12	Pt Platinum	78	195·23
Ca Calcium	20	40·08	Rb Rubidium	37	385·44
Cr Chromium	24	52·01	Rh Rhodium	45	102·9
Cs Caesium	55	132·81	S Sulphur	16	32·06
Cu Copper	29	63·57	Se Selenium	34	79·2
Fe Iron	26	55·84	Sn Tin	50	118·70
H Hydrogen	1	1·008	Sr Strontium	38	87·63
He Helium	2	4·002	Ta Tantalum	73	181·4
Hg Mercury	80	200·61	Th Thorium	90	232·12
K Potassium	19	39·10	U Uranium	92	238·14
Mg Magnesium	12	24·32	W Tungsten	74	184·0
Mo Molybdenum	25	54·93	Zn Zinc	30	65·38
N Nitrogen	7	14·008	Zr Zirconium	40	91·22
Na Sodium	11	22·997			

C. Units in Photometry

The standard candle is the unit of intensity of a light source. This is taken as one-tenth of the intensity of a standard Harcourt lamp burning pentane fuel under carefully controlled conditions. Light sources have their intensities specified in candle-power (C.P.).

Luminous flux, or quantity of light energy is measured in units called *lumens*. This is the quantity of luminous energy received per second per unit area of a sphere of unit radius when a point source of one candle is placed at the sphere centre. Alternatively, it is the light flux through unit solid angle from a point source of 1 candle-power. There are thus 4π lumens radiated in all directions by a point source of one candle.

The *lux* is the unit of illumination, defined as the illumination at a point on the surface of a sphere of 1 m. radius when unit candle exists at the sphere centre. It is a flux of 1 lumen illuminating a surface of 1 sq. m.

The *foot-candle* is the illumination produced when the light from a point source of one candle falls normally on a surface 1 ft. from the source. The *metre-candle* is defined in the same manner.

The rate of emission of light energy can be quoted in the fundamental power unit, the *watt*, being a rate of flow of energy of 10^7 ergs/sec. The human eye has a peak sensitivity (see fig. 89) at a wave-length of 5550 Å. At this wave-length, 1 W. is equal to 621 lumens. At any other wave-length, the ratio of watts/lumens can be determined from the eye sensitivity curve. Thus light energy flow as quoted in watts is in absolute units of energy, independent of the human eye, whereas quoted in lumens, the sensitivity of the eye is of concern.

D. Bessel Functions

A Bessel function of the zero-th order is given by

$$J_0(x)=1-\frac{x^2}{2^2}+\frac{x^4}{2^4(2\,!)^2}-\;.\;.\;.+(-1)^k\frac{x^{2k}}{2^{2k}(k\,!)^2}+\;.\;.\;.,$$

for the first order,

$$J_1(x)=\frac{x}{2}-\frac{x^3}{2^3 2\,!}+\frac{x^5}{2^5 2\,!\,3\,!}-\;.\;.\;.+(-1)^k\frac{x^{2k+1}}{2^{2k+1}k\,!\,(k+1)\,!}$$

and for the *n*th order,

$$J_n(x) = \frac{x^n}{2^n n!} - \frac{x^{n+2}}{2^{n+2}(n+1)!} + \frac{x^{n+4}}{2^{n+4}2!(n+2)!} - \cdots + (-1)^k \frac{x^{n+2k}}{2^{n+2k}k!(n+k)!}.$$

Such Bessel functions of the first kind are shown plotted in fig. 121 for the zero-th, first and second orders. Other orders can be plotted from the series given above for the nth order.*

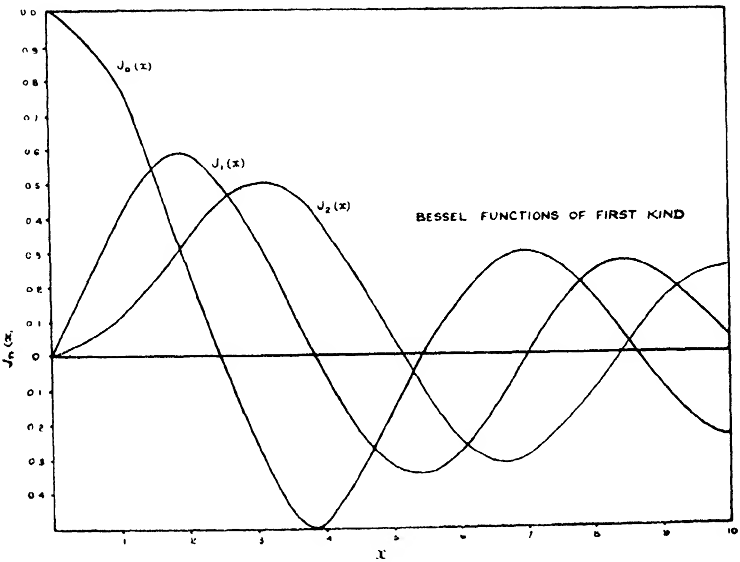


FIG. 121. Bessel Functions.

* See I. S. and E. S. Sokolnikoff, *Higher Mathematics for Engineers and Physicists*, McGraw-Hill, 1941.

BIBLIOGRAPHY

A list of books on electronics and allied subjects is given to act as a guide for wider reading.

Admiralty Handbook of Wireless Telegraphy (2 vols.), His Majesty's Stationery Office, London, 1938.

APPLETON, E. V., *Thermionic Vacuum Tubes*, Methuen, London, 1931.

ARDENNE, M. VON, *Cathode-Ray Tubes*, Pitman, London, 1939.

ARDENNE, M. VON, *Elektronen-Ubermikroskopie*, Springer, Berlin, 1940.

BIJL, H. J. VAN DER, *The Thermionic Vacuum Tube*, McGraw-Hill, New York, 1920.

BLOCH, E., *The Kinetic Theory of Gases*, Methuen, London, 1930.

BOER, J. H. DE, *Electron Emission and Adsorption Phenomena*, Macmillan, New York, 1935.

BRAINERD, J. G. (Editor), *Ultra-High-Frequency Techniques*, Van Nostrand, New York, 1942.

CAMPBELL, N. R., and RITCHIE, D., *Photo-Electric Cells*, Pitman, London, 1934.

CHAFFEE, E. L., *Theory of Thermionic Vacuum Tubes*, McGraw-Hill, New York, 1933.

COSSLETT, V. E., *Introduction to Electron Optics*, Oxford, 1946.

CROWTHER, J. A., *Ions, Electrons and Ionizing Radiations*, Arnold, London, 1938.

DOW, W. G., *Fundamentals of Engineering Electronics*, Wiley, New York, 1937.

EASTMAN, A. V., *Fundamentals of Vacuum Tubes*, McGraw-Hill, New York, 1941.

EMERY, W. L., *Ultra-High-Frequency Radio Engineering*, Macmillan, New York, 1944.

ENGEL, A. VON, and STEENBECK, M., *Elektrische Gasentladungen*, Springer, Berlin, 1932.

EVERITT, W. L., *Communication Engineering*, McGraw-Hill, New York, 1937.

FRANK, N. H., *Introduction to Electricity and Optics*, McGraw-Hill, New York, 1940.

GABOR, D., *The Electron Microscope*, Hulton, London, 1944.

HARVEY, A. F., *High Frequency Thermionic Tubes*, Chapman & Hall, London, 1943.

HENNEY, K. (Editor), *The Radio Engineering Handbook*, McGraw-Hill, New York, 1941.

HENNEY, K., *Electron Tubes in Industry*, McGraw-Hill, New York, 1934.

HIRSCHLAFF, E., *Fluorescence and Phosphorescence*, Methuen, London, 1938.

HUGHES, A. L., and DUBRIDGE, L. A., *Photo-Electric Phenomena*, McGraw-Hill, New York, 1932.

JEANS, J. H., *Dynamical Theory of Gases*, Cambridge, 1925.

JONES, T. J., *Thermionic Emission*, Methuen, London, 1935.

KLEMPERER, O., *Electron Optics*, Cambridge, 1939.

KOLLER, L. R., *The Physics of Electron Tubes*, McGraw-Hill, New York, 1937.

LOVELL, B. (Editor), *Electronics and Their Application in Industry and Research*, Pilot Press, London, 1947.

MACGREGOR-MORRIS, J. T., and HENLEY, J. A., *Cathode-Ray Oscillography*, Chapman & Hall, London, 1936.

MALOFF, I. G., and EPSTEIN, D. W., *Electron Optics in Television*, McGraw-Hill, New York, 1938.

MAXFIELD, F. A., and BENEDICT, R. R., *Theory of Gaseous Conduction and Electronics*, McGraw-Hill, New York, 1941.

MCARTHUR, E. D., *Electronics and Electron Tubes*, McGraw-Hill, New York, 1936.

MILLMAN, J., and SEELY, S , *Electronics*, McGraw-Hill, New York, 1941.

MOULLIN, E. B., *Spontaneous Fluctuations of Voltage*, Oxford, 1938.

MYERS, L. M., *Electron Optics*, Chapman & Hall, London, 1939.

PARR, G., *The Cathode Ray Tube*, Chapman & Hall, London, 1944.

PIERCE, G. W., *Electric Oscillations and Electric Waves*, McGraw-Hill, New York, 1920.

PUCKLE, O. S., *Time Bases*, Chapman & Hall, London, 1943.

RATCLIFFE, J. A., *The Physical Principles of Wireless*, Methuen, London, 1931.

REICH, H. J., *Theory and Applications of Electron Tubes*, McGraw-Hill, New York, 1939.

REIMANN, A. L., *Thermionic Emission*, Wiley, New York, 1934.

RICHARDSON, O. W., *Emission of Electricity from Hot Bodies*, Longmans, London, 1921.

RICHTMYER, F. K., and KENNARD E. H., *Introduction to Modern Physics*, McGraw-Hill, New York, 1942.

RISSIK, H., *The Rectification of Alternating Current*, English Universities, London, 1938.

RUTHERFORD, D. E., *Vector Methods*, Oliver & Boyd, London, 1943.

SCOWEN, F., *Electric Wave Filters*, Chapman & Hall, London, 1945.

SLATER, J. C., *Microwave Transmission*, McGraw-Hill, New York, 1942.

SOKOLNIKOFF, I. S. and E. S., *Higher Mathematics for Engineers and Physicists*, McGraw-Hill, New York, 1941.

SOMMER, A., *Photo-Electric Cells*, Methuen, London, 1946.

STARLING, S. G., *Electricity and Magnetism for Degree Students*, Longmans, London, 1937.

STURLEY, K. R., *Radio Receiver Design* (2 vols.), Chapman & Hall, London, 1945.

TERMAN, F. E., *Radio Engineers' Handbook*, McGraw-Hill, New York, 1943.

TERMAN, F. E., *Radio Engineering*, McGraw-Hill, New York, 1937.

THOMAS, H. A., *Theory and Design of Valve Oscillators*, Chapman & Hall, London, 1939.

TURNER, L. B., *Wireless*, Cambridge, 1931.

WALKER, R. C., *Photo-Electric Cell Applications*, Pitman, London, 1933.

WATSON-WATT, R. A., and Others, *Applications of the Cathode-Ray Oscillograph in Radio Research*, His Majesty's Stationery Office, London, 1933.

WELLER, B. F., *Radio Technology*, Chapman & Hall, London, 1943.

WHEATCROFT, E. L. E., *Gaseous Electrical Conductors*, Oxford, 1937.

WHITE, F. W. G., *Electromagnetic Waves*, Methuen, London, 1934.

WILSON, W., *The Cathode-Ray Oscillograph in Industry*, Chapman & Hall, London, 1943.

ZWORYKIN, V. K., and WILSON, E. D., *Photo-Cells and Their Application*, Wiley, New York, 1932.

ZWORYKIN, V. K., and MORTON, G. A., *Television*, Wiley, New York, 1940.

ZWORYKIN, V. K., and Others, *Electron Optics and The Electron Microscope*, Wiley, New York, 1946.

INDEX